Human-Robot Interaction System Based on Natural Calling Gestures

(自然な手招きジェスチャによる人間とロボットのインタラクションシステム)

2019年9月

理工学研究科 (博士後期課程)

数理電子情報 (主指導教員 - 久野　義徳)

AYE SU PHYO
(アエスピョ)

Human-Robot Interaction System Based on Natural Calling Gestures

AYE SU PHYO

Graduate School of Science and Engineering
Saitama University, Japan
Supervisor: Professor Yoshinori Kuno

A thesis submitted in partial fulfillment of the requirement for the degree of

Doctor of Philosphy

September, 2019

Acknowledgement

First of all, I thank my ***PARENTS*** for giving me the strength and ability to complete this study. There are also 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, enormous support, perfect guidance, strong encouragement, patience and for being always there since the beginning. I am sincerely grateful to him for his constant encouragement, and his stimulating ideas.

I am incredibly grateful to ***Professor Yoshinori Kobayashi, and Assistant Professors Antony Lam and Hisato Fukuda*** for their comments, insightful discussions, valuable advice, and feedback during evaluating of my thesis work. I am forever thankful to all lab members for their sharing knowledge and their time and help that they have given to me.

For this dissertation I would like to thank my oral defense committee members: ***Professor Testsuya Shimamura and Professor Takashi Komuro*** for their time, interest, insightful questions and helpful comments.

And very certainly not least, I am grateful to my scholarship foundation ***ROTARY*** which has offered not only financial but also volunteer experience in Japan.

Finally, my family, my friends, colleagues, teachers and Saitama University Myanmar Association (SUMA) members for their continuous encouragement in my hectic time of research work. There are many people that have helped me while studying abroad to name personally, but I am sure you know who you are. Thank you, and thank you all for your kind support and encouragement.

Abstract

The steadily expanding population of elderly persons in Japan and other industrialized countries has posed a vexing problem to the health care systems that serve aging citizens. The field of robotics, in particular the development of service robots that can provide assisted-care, has made many gains over the last decade. For actual care tasks, we need to develop communication technology to allow users to easily ask for services to assisted-care robots. Thus, human-robot interaction (HRI) becomes one of the most important aspects of development in service robots. Interacting with service robots via nonverbal cues allows for natural and efficient communication with humans.

Human-Robot Interaction system must be designed and implemented so that age-related challenges in functional ability, such as perceptual, cognitive and motor functions, are taken into account. There is the increasing popularity of service robots communication using interfaces: touch panels, voice control, etc. Compared with these interfaces, gesture interfaces which users use the movements of the hands, fingers, head, face and other parts of the body have the advantage of simplicity; they require less learning time. For older users, who may operate other interfaces with limited speed and accuracy, the gesture interfaces can be attractive and make interactions more flexible. Gesture interfaces may make interaction with robots more attractive and friendly to older users because they are natural and intuitive, they require minimal learning time and they lead to a high degree of user satisfaction.

The main objective for this research is to develop the Human-Robot interaction system by empowering them to take into account gestures performed by the human in a flexible, fast and natural way and by the means of an intentional control architecture that enables the robot to quickly react to the users' stimuli. To get this objective, we proposed natural hand calling gesture recognition algorithm using skeleton features in crowded environments for human-robot interaction. For the gesture interface to communicate with robot, this work mainly focuses on natural calling hand gestures. Hand gesture recognition is a challenging problem in computer vision and is a topic of active research.

In real situations, the user may perform gestures in various positions and the environment may also have many people with hand motions. And, we make the observation that if the person does not have any intention to call the robot, he/she may not move his/her arm against the gravity. When a person calls someone, it is natural to direct his/her hand with an open towards the target person. However, there are still challenges in vision-based hand gesture recognition such as illumination changes and the background-foreground problem, where objects in the scene might even contain skin-like colors. Another issue is the presence of crowds moving around with many hand motions. In crowded environments, conventional methods might erroneously recognize hand movements as calling gestures.

Based on these findings from observations of people's daily activities and challenges, a service robot was developed and used to interact with elderly people for helping their daily activities in a natural way. We developed a hand calling gesture recognition method that can recognize in real time natural gestures not specified in advance. Following this research program, the following challenges have been identified:(1) illumination changes and the background-foreground problem, where objects in the scene might even contain skin-like colors, (2) the presence of crowds moving around with many hand motions and randomly moving objects, (3) the caller's position that may be varied and not in front of the camera but in the view of the camera, (4) the natural gestures that is no need to remember the defined gesture, and (5) less learning time and more satisfaction for elderly people because of the gestures used in childhood.

In our approach, only the people who gaze towards the robot with defined wrist positions are checked from the scene. This comes from our observation that people typically gesture to call others while gazing towards the target person. Then based on the overall body poses of some people, we determine candidate people that might be calling the robot. We then zoom into each candidate person's hand-wrist part to extract finer details of the person's hand pose. Our approach then uses the key-points of the fingertips to make final decision on whether something is a calling gesture or not. So essentially, we process in stages, the combination of overall body pose and local hand pose to determine whether someone is calling the robot or not. This cascade of calling gesture detection stages allows for efficient recognition in crowded settings.

The major goal of this research is to recognize natural calling gestures from people in an interaction scenario where the robot continuously observes the behavior of a humans. In our approach, firstly, the robot moves amongst people. At that time, when the person calls the robot by a hand gesture, the robot detects the person who is calling the robot from among the crowd. While approaching to the potential caller, the robot observes whether the person is actual calling the robot or not. We tested the proposed system at a real elderly care center. We validate our findings using our experimental setup, which is composed of a humanoid robot (Aldebaran's NAO) and an i-Cart mini (T-frog) that carries the NAO humanoid and a webcam.

This thesis proposes a service robot system that provides assisted-care to the elderly. This system recognizes natural calling gestures in an interaction scenario where the robot visually observes the behavior of humans. Therefore, an algorithm for natural calling gesture recognition in crowded environments, for human-robot interaction is introduced. To detect users, this study uses the key-points from the OpenPose real-time detector. Using these key-points, gaze detection and finding the hand-wrist positions are performed. If the algorithm finds the gaze and defined hand-wrist position, it zooms into the hand-wrist part. After that, it finds the key-points of the hand's fingertips. From these key-points, this algorithm recognizes whether the user is calling or not by a simple but effective rule-based classification, developed based on basic observations about how people perform calling gestures in real settings. After detecting the calling gesture, the robot moves to the caller. While approaching, the robot observes whether the user is actually calling or not. From this result, the interaction between humans and robot more effective.

Contents

List of Figures

List of Tables

List of Equations

Chapter 1

Introduction

With improved living conditions due to the continuous development of society, people are living longer. Therefore, many countries in the world are faced with the aging of its population, which needs help and care [1]. It is widely accepted that more research is needed to address this issue. One solution, robots, have been introduced to solve this problem as shown in Figure 1.1. Autonomous robots are making their way into human inhabited environments such as homes and workplaces: for entertainment, helping uses in their domestic activities of daily living, or helping disabled people care or basic activities which would improve their autonomy and quality of life [2]. According to the International Federation of Robotics (IFR), a service robot is a robot that is semi-autonomously or fully-autonomously operated to perform useful tasks for humans. Specifically, a personal service robot is used to benefit humans or enhance human productivity [3]. There is a great diversity of service robots and several fields where they can be used, for example in health care, assistance to disabled, transportation, safety, and security, among others [4; 5].

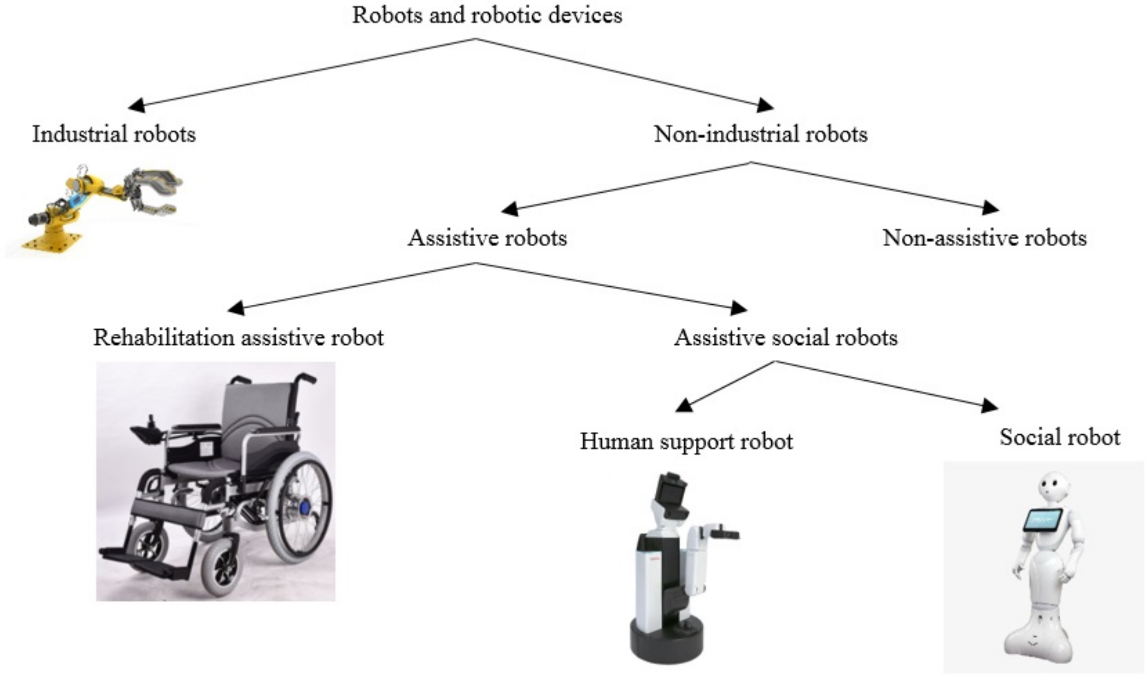
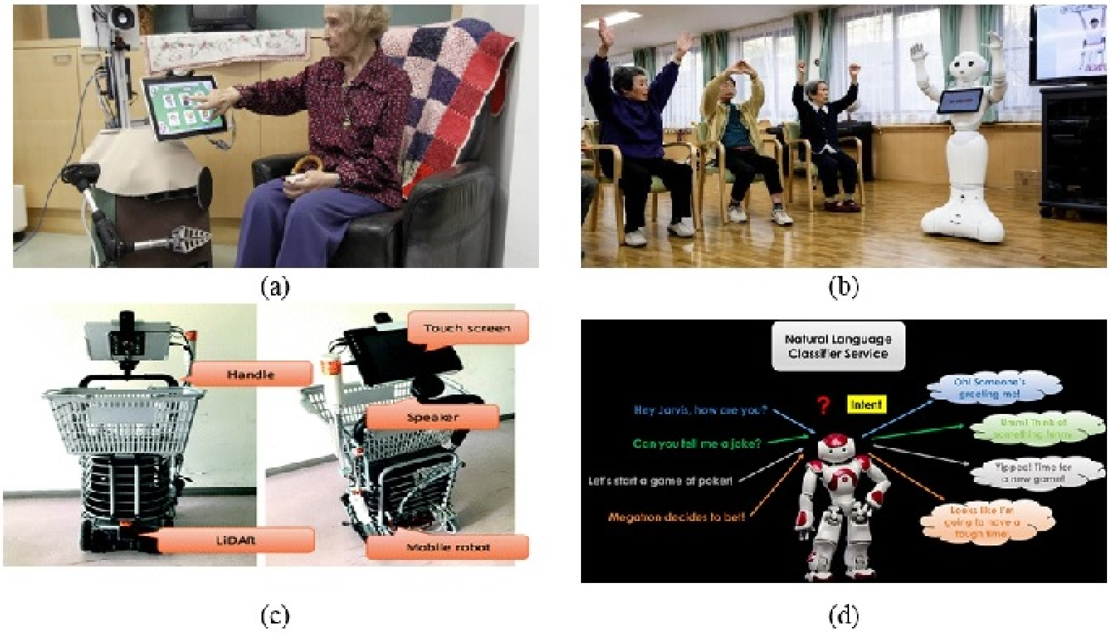


Figure 1.1: Categorization of robots

Service robots, may even replaced home care staff (e.g a caregiver) to take care of elderly. In addition, more and more service robots are being designed for elderly care in order to support health care staff [6]. The goal of these robots is to communicate with humans in a human-like manner and perform different tasks as instructed by human users. For service robots to perform tasks like humans, service robots should have the abilities people have. The service robots should be able to recognize humans, their verbal communication and gestures in order to realize natural communication.

1.1 Motivation

Now that robots are able to complete thousands of interesting tasks for us, it is time to make them understand humans as well as communicate with. This is the main aim of Human Robot Interaction (HRI), to design interfaces and situations to better operate and interact with robots. Social behavior in robots generally depend upon efficient human-robot interaction (HRI). Becoming information technologies commonplace in society, service robots have also been developed using modern technologies of devices. The common modes of human-robot interaction (HRI) are either via interfaces or by speech and gestures. Speech and textual interfaces are widely used in this field, but as many psychologists claim, approximately more than the 60% of human communication is performed through non-verbal cues. According to [7], 65% of our communication consists of human gestures and only 35% consists of verbal content. This two-thirds of our mode of communication shows the significance of gestures. For this purpose, recognition of nonverbal content becomes essential for HRI. Human gestures are an important form of nonverbal content, which is used with or without verbal communication in expressing the intended meaning of the speech. Such gestures may include hand, arm, or body gestures and it may also include use of the eyes, face, head and more. So, humans tend to interact with themselves via gestures as an important element of communication. We usually wave to our acquaintances, or point at an object to refer to them instead of describing all the scene. In many cases, gestures are more efficient to be performed and be understood, hence it is really interesting to include these abilities into robotics systems.


(a) (b) (c) (d)

Handle
Touch screen
Speaker
LiDAR
Mobile robot
Natural Language Classifier Service
?
Intent
Hey Jarvis, how are you?
Can you tell me a joke?
Let's start a game of poker!
Megatron decides to bet!
Oh! Someone's greeting me!
Yippee! Time for a new game!
Looks like I'm going to have a tough time.

Figure 1.2: Examples of service robots using interfaces

Robots are classified into several types based on their functionality (service and utility robots or those designed to communicate with humans) and appearance (humanoid robots or mechanical robots) like in Figure 1.1. With information technologies becoming commonplace in society, the opportunity and necessity for elderly people to access these technologies in their everyday activities have been increasing. Human-robot interaction must be designed and implemented so that age-related challenges in functional ability, such as perceptual, cognitive and motor functions, are taken into account. Recent years have seen the increasing popularity of service robots using interfaces: touch panels, voice control, etc. as shown in Figure 1.2. Compared with these interfaces, gesture interfaces which users use the movements of the hands, fingers, head, face and other parts of the body have the advantage of simplicity; they require less learning time. For older users, who may operate other interfaces with limited speed and accuracy, the gesture interfaces can be attractive and make interactions more flexible.

In the last decades, much attention has been devoted to understanding and accommodating the needs of the elderly with respect to interaction with robots through touch-screen. Recent years have seen the increasing popularity of gesture-based communications, where users use the movements of the hands, fingers, head, face and other parts of the body to interact with robots. Furthermore, studies have been carried out to investigate how older users use gesture inputs in their interactions with information technologies [8]. Gesture interfaces may make interaction with robots more attractive and friendly to older users because they are natural and intuitive, they require minimal learning time and they lead to a high degree of user satisfaction. The touch screen has been suggested as a suitable input device for elderly users because it is easy to learn and operate. There is a large body of work on single finger touch screen applications for older users, but relatively little work has been done on the use of multi-touch, hand, face and body gestures for interacting with robots.

The main motivation for this research is the Human-Robot interaction by empowering them to take into account gestures performed by the human in a flexible, fast and natural way and by the means of an intentional control architecture that enables the robot to quickly react to the users' stimuli.

1.2 Contributions

This thesis proposes a human-robot interaction system using gestures to address the challenging problem of providing for an engaging and effective interaction as natural as possible for helping elderly people. A service robot was developed and used to interact with elderly people for helping their daily activities in a natural way. We developed a hand calling gesture recognition method that can recognize in real time natural gestures not specified in advance. To evaluate the performance of human-robot interaction system based on gesture recognition, an indoor test scenario with several specific situations was performed at a real elderly care center.

The main contributions of this thesis are:

- Development of a natural hand calling gesture recognition algorithm, which uses a rule-based heuristic for classifying the recognition calling gestures using skeleton features in crowded environments with randomly moving objects (Chapter 3).
- Development of an efficient and real-time human-robot interaction system based on natural hand calling gestures for elderly care (Chapter 4).
- Evaluation of the performance of human-robot interaction system based on gesture recognition at a real elderly care center (Chapter 5).

Following this research program, the following challenges have been identified:

1. illumination changes and the background-foreground problem, where objects in the scene might even contain skin-like colors
2. the presence of crowds moving around with many hand motions and randomly moving objects
3. the caller's position that may be varied and not in front of the camera but in the view of the camera
4. the natural gestures that is no need to remember the defined gesture
5. less learning time and more satisfaction for elderly people because of the gestures used in childhood.

1.3 Organization

The thesis is organized as follows:

- **Chapter 2** presents the background and literature reviews of the hand gesture recognition and human-robot interaction system based on hand gestures. The literature review focused on vision-based hand gesture recognition and service robot system for elderly care. The various approaches of hand gesture recognition system employed for service robots are discussed, highlighting the strengths and weaknesses for each method.

  To begin with, we have studied key concepts, state of the art solutions and tools available to accomplish this goal. And also, this chapter thoroughly discusses about the humanoid robot and hand gesture recognition with the help of computer vision and machine learning algorithms. In details, this chapter explains the concepts service robots, human-robot interaction for elderly care, gesture and hand gesture recognition, conventional hand gestures recognition approaches and literature review on human-robot interaction system based on gestures.

- **Chapter 3** presents detailed process of hand calling gesture recognition method which uses a rule-based heuristic for classifying the recognition calling gestures using skeleton features in crowded environments with randomly moving objects. Chapter 3 thoroughly discusses about a natural calling gesture recognition approach using skeleton features in crowded environments, processed in two-stages. In details, this chapter explains the step by step processes of the method - body key-points feature acquisition, gaze detection and finding hand-wrist position, zooming in the hand-wrist part and fingertip key-points features acquisition and recognizing of calling gestures using rule-based classification method.

- **Chapter 4** describes human-robot interaction system based on natural hand calling gestures for elderly care. This chapter discusses about the system architecture for human-robot interaction service robot system, the process model of the proposed service robot with gesture recognition, algorithm procedure for human-robot interaction system for this research, and measurement of distance and angle between human and robot in details.

- **Chapter 5** discusses the evaluation of the performance of human-robot interaction system based on gesture recognition at a real elderly care center. This chapter gives the experimental analysis of natural hand calling gesture recognition method with two different experiment situations and the analysis of human-robot interaction system.

- **Chapter 6** presents conclusions and future work plans. It is concluded as the summary of the natural hand calling gesture recognition method and human-robot interaction system for service robot which helps elderly care area efficiently and effectively by the potential future work and application.

1.4 Concluding Remarks

In this chapter we presented an introduction covering some important aspects, of human-robot interaction system of service robots for elderly care. Then, the motivation for this thesis and our contributions were discussed. Organization of thesis is described in finally.

Chapter 2

Background and Literature Review

2.1 Background

Nowadays, with robots entering in our daily lives, it is becoming important to provide the users a simple an intuitive way to interact with them. Human Robot Interaction (HRI) has already proved an active research field from many different points of view: from making humans understand the robot states through verbal and non verbal communication to doing it the other way around, making the robot understand humans. Besides, given that a great part of the human communication is carried out by means of non-verbal channels, skills like gesture recognition and human behavior analysis reveal to be very useful for this kind of robotic systems, which would include viewing and understanding their surroundings and the humans that inhabit them [10].

Body postures are a powerful means for humans to convey information. Gestures are thought to be one of the natural forms of communication between even people in different societies and speaking different languages, particularly in noisy environments or where speech is not possible. Gesture recognition is a growing research area due to its wide range of applications, ranging from medical systems and assistive technologies, entertainment, crisis management, disaster relief, human-robot interaction and many more [11]. Among these areas, the most common application is human robot interaction (HRI). It aims to replace the traditional interfaces with a more natural interface.

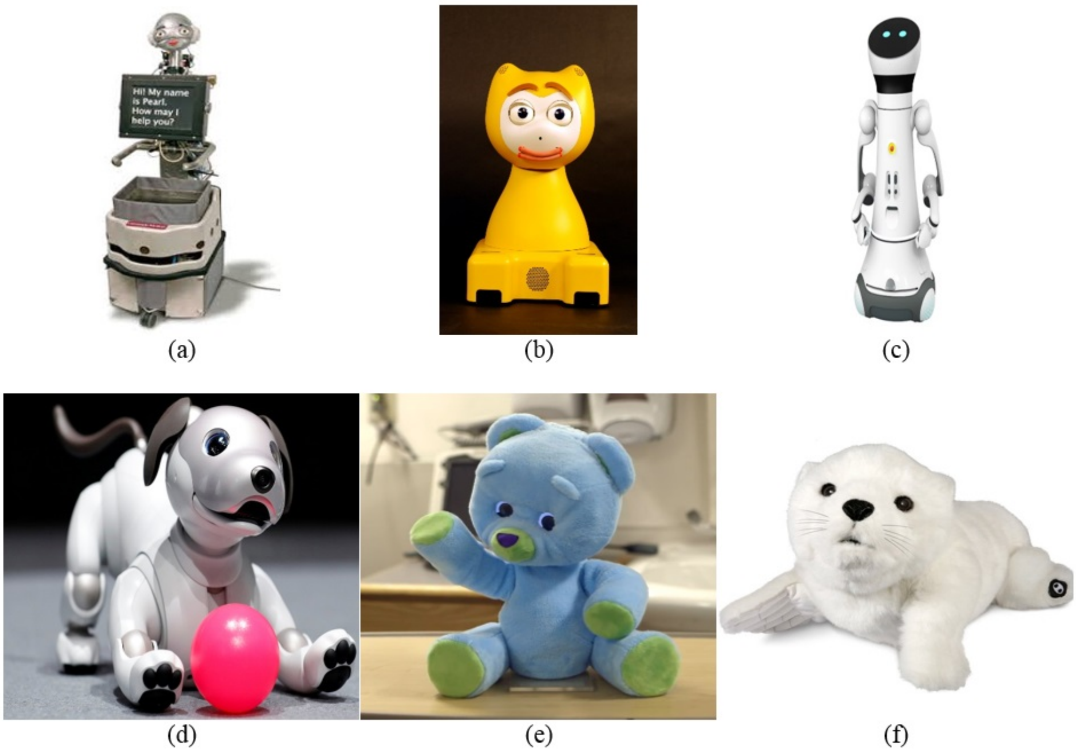
Hand gesture recognition is a challenging problem in computer vision and represents an active area of research. Hand gesture communication involves hand and arm motion information and two approaches are commonly used to interpret them: sensor-based and vision-based [12]. At present, although the study of hand-gesture recognition has made great process and achieved high recognition rate in different areas, it is still facing many challenges.

In this thesis, we will focus on building hand gesture recognition system for service robots in elderly care. As this work is focused on interaction with robots based on gestures, this section will present a review of the available work in this field.

2.1.1 Service Robots

Because of the growing of our world population, there is a growing necessity for new technologies that can assist elderly in their daily living. One area of digital-physical information systems is service robotics. Service robots are technical devices that perform tasks useful to the well-being of humans in a semi or fully autonomous way (International Federation of Robotics 2015a) [13]. The differentiation between industrial and service robots is based on their area of application and closeness to end-users [14]. Since service robots have to operate and communicate in an unconstrained, human-centered environment, a high degree of autonomy is an inherent characteristic of them [15]. For example, a service robot responsible for the cleaning of floors in a hospital has to autonomously navigate its way through the building. While navigating, the service robot has to react to its environment in a real time manner: It has to stop for patients, doctors, and other human beings that might block the robot's route as well as to independently circumnavigate any obstacles like patients' beds or wheelchairs.

As service robots have become more versatile and diffuse to different areas of human life, they also have found their way into research. While extant research rather focuses on technical aspects of the robot itself, there is still a lack of context specific research on service robots [16]. Since service robots join the human environment, they must meet different requirements. To ensure a smooth human-robot interaction, technologies have to be used to help the robot blend in with the human-centered context. Advanced machine learning capabilities, which help the service robot to better understand and react to the environment and context in which it is acting, as well as fine motor skills, which aid the robot in mimicking human appearance and behavior, are just two examples where further research is needed.



(a) (b) (c)

(d) (e) (f)

Figure 2.1: Types of Service Robots: (a) Pearl Nursebot, (b) Dutch iCat, (c) Care-o-bot , (d) Abio, (e) Huggable Teddy Bear, (f) Paro

Studies on robots in eldercare feature different robot types. First, there are robots that are used as assistive devices which we will refer to as service type robots. Functionalities are related to the support of independent living by supporting basic activities (eating, bathing, toileting and getting dressed) and mobility (including navigation), providing household maintenance, monitoring of those who need continuous attention and maintaining safety [17]. Examples of these robots shown in Figure 2.1 are 'nursebot' Pearl [18], the Dutch iCat (although not especially developed for eldercare) and the German Care-o-bot [19]. Also categorized as such could be the Italian Robocare project, in which a robot is developed as part of an intelligent assistive environment for elderly people [20]. The social functions of such service type robots exist primarily to facilitate interfacing with the robot. Studies typically investigate what different social functions can bring to the acceptance of the device in the living environment of the elder, as well as how social functions can facilitate actual usage of the device. Second, there are studies that focus on the pet-like companionship a robot might provide.

The main function of these robots is to enhance health and psychological well-being of elderly users by providing companionship. We will refer to these robots as companion type robots. Examples are the Japanese seal shaped robot Paro [21], the Huggable teddy bear [22] (both especially developed for experiments in eldercare) and Aibo (a robot dog by Sony). Social functions implemented in companion robots are primarily aimed at increasing health and psychological well-being. For example, studies investigate whether companion robots can increase positive mood in elderly living in nursery homes. However, not all robots can be categorized strictly in either one of these two groups. For example, Aibo is usually applied as a companion type robot, but can also be programmed to perform assistive activities [23] and both Pearl and iCat can provide companionship.

2.1.2 Human-Robot Interaction for Elderly Care

Varying in the objectives, some robots are developed for helping humans in industrial environment and some are designed to function in indoor environment (as in Figure 2.2. As the technology gets sophisticated and more advanced, the focus has been shifted to social service robots. The goal of these robots is to communicate with human in a human-like way and perform different tasks as instructed by human. This leads us to social behavior in robots. These social robots should recognize humans, their verbal communication and gestures in order to realize natural communication. Furthermore, they should also recognize human emotions in order to predict the internal state of human for better communication. Social behaviour in robots generally depend upon efficient human-robot interaction (HRI) [25]. The most common way of human interaction is either by vocal communication or by body gestures. Other medium includes newspapers, notes and other writing material, however, this type of communication is not applicable in face-to-face communication. For this purpose, recognition of nonverbal content becomes essential task for HRI. This has given rise to the field of HRI, the study of how to design and implement robotic systems that can interact with a human environment in a safe and efficient manner.

HRI is a challenging field because the system needs to be able to perceive, understand and react to human activity in real time. These challenges should be taken into consideration when designing a HRI system. Challenges include:

- There is a limitation on the size of the gesture recognition system, it must be able to fit on the robot.
- As both the robot and the human are mobile, static backgrounds cannot be used for segmentation and a a fixed camera location cannot be assumed.
- The robot mobility could lead to drastic changes in environmental conditions, such as lighting.

Figure 2.2: Some Examples of Service Robots for Elderly Care

- The system must be able to work in real-time. Ideally, there must not be a perceivable lag between the user performing a gesture and the robot response.

Service robots can help people to do some homework, and reduce the workload of caregivers, in some home-care system, it is useful to improve the life quality of the old and the disabled. To make service robot be widely used in practical, it is a very necessary job to achieve natural communication between human and robots. Unlike in industrial robots field, only expert operators can communicate and operate industrial robots, the operators are ordinary people in service robots field, such as the old, the disabled and autistic patients, etc [25],and many of them have some cognitive impairment or a certain weakening ability. So it is valuable to study the interaction mechanism between human and robots.

People and the service robots are generally working in a relatively large collaborative space (e.g., in home environment), so the human-robot interaction and collaboration takes place except in short range, the mobile robot also needs to communicate with people in remote to perform tasks. Thus, the full communication between people and service robots through natural interface for interactive access is the basis for interactive collaboration. Many researchers in the world have carried out a lot of work in terms of interaction between human and nursing robot, such as the development of voice, eye gaze, brain machine interaction, body gestures, facial expressions and other modalities [26].

2.1.3 Gestures and Hand Gestures Recognition

It is essential that the communication between robots and humans, or human-robot interaction (HRI), be as natural as possible. To this end, there has been much research focused on imbuing robots and computers with the ability to understand human gestures. Human gestures are nonverbal content, which are used with or without verbal communication in expressing the intended meaning of the speech.

2.1.3.1 Gestures Recognition

Gestures, one of the most natural forms of communication, can be performed using any part of the body, from the arms, the head, as in a nod, or even the face as shown in Figure 2.3. Gestures can be divided into two types according to their movement along time: static or dynamic. Static gestures do not change with time, they are described by the pose/posture in a single instant. Dynamic gestures change the posture across time and the gestures are described by its movement [31]. Gesture Recognition is the process of identifying the gesture performed by a user and usually has the aim of interpret certain commands [32]. We divided the gesture recognition process in four important parts, Data Acquisition where the information from the ambient is acquired, Segmentation where the features necessary to perform the gesture recognition are extracted, Tracking where the hand is tracked across time and Classification where the gestures are modelled and recognized.

Figure 2.3: Some Gestures for Communication

According to Wachs et al. [33], the basic requirements for any gesture recognition system are:

1. Responsiveness: the system must be able to recognise gestures almost instantaneously (a maximum delay of 45 ms) as a slow system is impractical.
2. User adaptability and feedback: the majority of gesture recognition systems have a defined number of gestures which the system is able to identify. These gestures are programmed through an offline classifier algorithm. The challenge is to provide a classifier that is able to generalise a gesture from minimal training samples.
3. Learnability: the gestures used to control the system should be easy to remember and execute.
4. Accuracy: the system must be able to first detect if the hand or body is within the view, track the hand from frame to frame, and match the gesture to learnt templates (recognition).
5. Intuitiveness: gestures used in the system should be intuitive in order to mimic communication between humans. For example a closed fist with the thumb up could represent "OK". This is strongly dependent on cultural background and experience.
6. Lexicon size: a lexicon is a dictionary of the gestures used in the system. Ideally increasing the number of signs in the system should affect the performance and accuracy of the system as little as possible.
7. Garment and environment requirements: the system should not require the user to wear additional aids or to be wired to a device, and in terms of background and illumination the environment should not need to be fixed.
8. Reconfigurability: hands are different in size, shape and skin colour, thus the gesture recognition system should be invariant to these variations.
9. Mobility: many systems are dependent on the assumption that the user stands in a fixed position. For many applications this is not a valid assumption.
10. Unintended gestures: the system must be able to distinguish between intentional and unintentional gestures.

These ten requirements are integral to the development of a robust, reliable and accurate gesture recognition system. Frequently, Gestures can be dialect particular or culture-specific. More specifically gestures can be categorized as mentioned below [34]:

1. Hand and arm gestures: utilizing hand and arm signals individuals can associate with virtual condition. These sorts of signals are required close by postures acknowledgment, in communications via gestures, and stimulation applications [34].
2. Face head gestures: Some of the illustrations are provided here:
   - head rotation,
   - head moving up and down,
   - eye rotation,
   - eyebrows raising,
   - winking the eye,
   - To talk by opening the mouth,
   - flaring the nostrils, and
   - human emotions like amazement, illness, fear, outrage, misery, and so on [34].
3. Body gestures: Full body movement is involved in it, as in [34]:
   - Tracking interaction between people,
   - Dancer movement analysis, and
   - Human gaits recognition for medical treatment and athletic training.

2.1.3.2 Hand Gestures Recognition

Human gesture recognition has been a popular topic and a challenging research in computer vision field. The topic has been studied numerous times because of its important applications in surveillance systems, elderly care, in the field of medicine (e.g. gait analysis, surgical navigation), in the field of sports, augmented reality, sign language for hearing impaired people and human behavior analysis. Among hand gestures, this thesis focuses on hand calling natural gestures to communicate human and robot. There are many forms of calling gestures depending on the person and regions. Figure 2.4 shows some forms of calling gestures to someone come here.

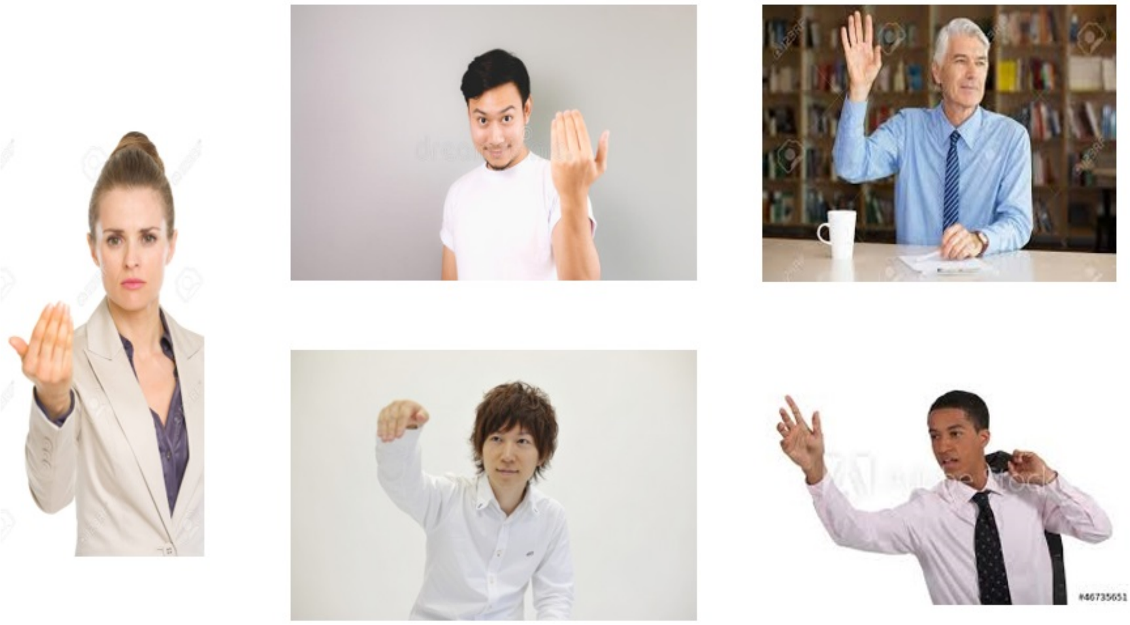


Figure 2.4: Some Calling Hand Gestures

Hand gestures are critical in face-to-face communication scenarios. To complete any task, first step would be to collect sufficient data. Recently, for recognition system for hand gestures, various technologies are used to acquire input data. All these technologies are categorized as following types shown in Figure 2.5 [35]:

- Vision based approaches
- Depth based approaches
- Instrumented(data) Glove approaches and
- Colored Markers approaches.

Vision Based approaches : Pictures are taken to interface between human and PCs. Camera is the main gadget which is used to capture pictures in vision-based approach. There is no necessity for other gadgets. To capture pictures, computerized cameras are used. Acquired pictures are additionally prepared and broken down by utilizing the vision-based procedures. Cameras such as fish eye, infrared, monocular and time of flight are various divers sort ones [35]. Through vision-based techniques, recognition of the represented alphabets and numbers are becoming easier.

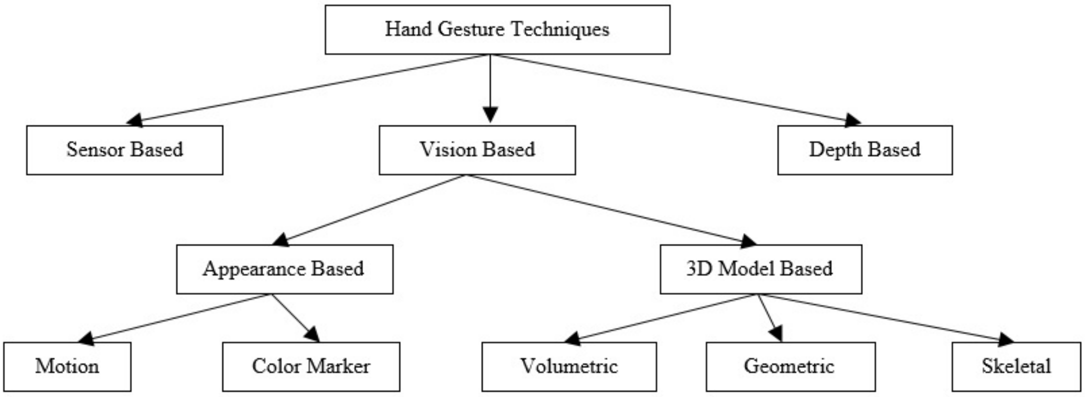


Figure 2.5: General Classification of Hand Gesture Recognition Techniques

This approach is simple and common to utilize so, this approach is extremely famous [36]. In this approach, there is a direct connection and interaction to human and computer devices. Conversely, numerous challenges for gestures need to be considered. Varieties of lights, different objects which are of skin color, brightening changes, and complex foundation are such difficulties in this approach. Other than that, recognition time, strength, speed and computational effectiveness are additionally challenging issues of this approach.

- Appearance Based Recognition : In this approach highlights are separated from visual appearance of information hand pictures. Examination is done between effectively characterized layout and this extricated picture. The primary advantage is its compelling execution of continuous and effective examination using 3D displays. This approach can additionally recognize distinctive skin hues. The issue of impediment is overwhelmed by this approach [37]. Fundamentally two classes are there as 2D static based model and movement-based model.
- Motion based Recognition : For object recognition model based approach is used. Object motion is calculated by sequence of the images. Adboost framework is used for the learning action model. Additionally, histogram of local motion is framed for object detection. Moreover, object description that is modelling of motion and motif recognition is the necessity for the gesture recognition and that has very high complexity [37].

Instrumented Glove Approaches : For capturing hand position and motion, sensor devices are used in the instrumented data glove approach. Sensors such as vision based, multi touch screen and mount bases are utilized. Using this approach, the fingers and palm location, their orientation, configuration are calculated precisely in this approach [38]. Reaction speed is fast, and it has high accuracy rate. The main requirement of this approach [38] the user must be physically connected with the computer. This limitation is a barrier to interaction between users and computers. These devices are very costly [[38] and not very efficient for virtual reality.

Colored Markers Approaches : To locate palm and fingers and to direct tracking process, human hand is worn with some colors in colored marker approach [39]. By forming hand shapes, necessary geometric features are extracted using this functionality. The color glove shape has small regions with different colors. A wool groove is used when three different colors are to indicate the finger palms in [40] where a wool grove is used. Its very easy to use and quite cheap compared to instrumented data glove. Still human computer interaction is not that much natural in this technology [39]. That is the principle impediment of this method.

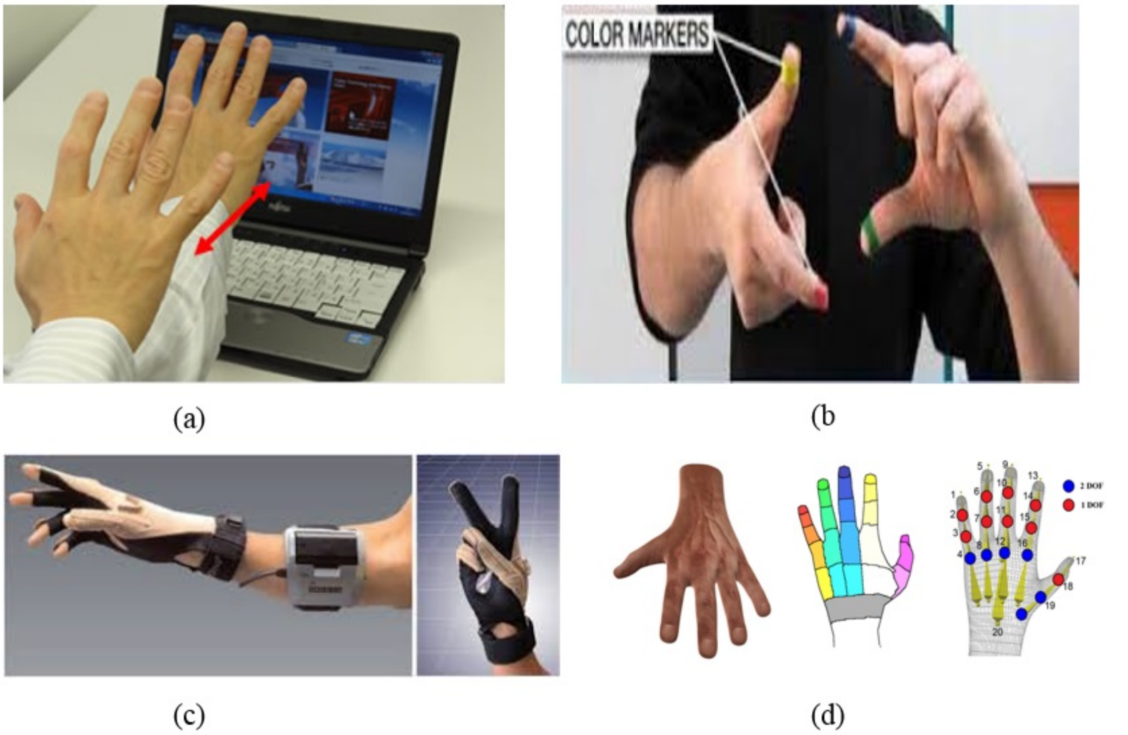


Figure 2.6: Examples of Hand Gesture Recognition Approaches: (a) Vision-based, (b) Color Marker based, (c) Instrument based, (d) Skeleton-based

Recognition using skeletal based Approaches : To overcome the problem of volumetric model Skeleton model is used. Using sparse coding this model provides higher efficiency. Complex features are optimized here. Compressive sensing is used for sparse signal recovery which reduces resource consumption [37].The study of vision-based methods, which uses only a camera without the use of any extra devices, has made great progress in different areas. However, there are still challenges such as illumination changes and the background-foreground problem, where objects in the scene might even contain skin.

Various methods have been proposed to locate and track body parts (e.g., hands and arms) including markers, colors, and gloves. We propose a method that does not require color information or extra device to be worn by the user (as in Figure 2.6. Such a gesture recognition module would have many applications including human robot interaction, intelligent rooms, and interactive games.

2.2 Conventional Hand Gestures Recognition Approaches

Gesture recognition is a growing research area due to its wide range of applications, ranging from medical systems and assistive technologies, entertainment, crisis management, disaster relief, human-robot interaction and many more. Among these areas, the most common application is human computer interaction (HCI). It aims to replace the traditional keyboard and mouse interfaces with a more natural interface. A large amount of the research in gesture recognition is dedicated to sign language translation, including recognition of both static alphabets and dynamic word gestures.

2.2.1 Sensors used for Gesture Interfaces

Numerous hand gesture recognition systems have been reported in the literature. The earlier systems require use of glove sensor for storing hand and finger motion and then use this data to recognize the action. Huang et al. [41] used gloves to record the hand and fingers flex data and then use machine learning algorithms to classify 5 dimensional finger flex data. Although, this type of systems may provide a 3D representation of hand however, wearing a heavy and expensive glove is not suitable for natural human interaction. On the other side, vision based systems take the information of the hand itself as an input using a camera to collect hand movements for gesture recognition without the use of any wearable sensor. Vision based approaches can be divided into two categories, i.e., 3D hand model based method and appearance based methods. The 3D hand model can provide ample information of hand that allows to realize wide class of hand gestures but the main disadvantage lies in extraction of features in case of ambiguous poses, unclear views and high computational complexity, which makes the overall system unrealistic for real time interaction.

In glove-based systems the user is required to wear a glove fitted with sensors such as accelerometers and flex sensors [42]. These devices directly measure the hand or arm joint angles and spatial positions . The gloves may be wired or wireless. The wired system requires the user to b e in close proximity to the computer. Glove-based systems restrict natural gestures as users are required to wear additional aids that may hamper natural movements. These systems also require the glove to b e calibrated for each new user, limiting the reconfigurability of the interface. However, they are typically more accurate.

Whilst there are various application specific uses for the glove-based systems, they are not a "natural" method of interacting with a robot. This shortfall has driven a large proportion of gesture recognition research toward vision-based systems. These approaches use single or multiple cameras and more recently depth sensors such as the Microsoft Kinect to capture and interpret gestures. The advantages of vision based sensors are numerous: they are simple, low cost and do not need to b e adjusted for each user. However, they suffer from the limitation that the gesture must be performed in the camera's field of view in order to b e detected. In addition, depending on the distance of the user from the camera the hand may only b e represented by a few pixels, making it challenging to extract useful features for hand gesture recognition.

In appearance based approaches, images with hands are considered only for feature extraction and gesture recognition task. The simplest technique is to look for skin color regions. Marilly et al. [48] extracts hand region using skin color and foreground information. For feature extraction, they use statistical and geometric features and then classify the gestures using principle component analysis. However, this method has some serious shortcomings. The major drawback of colorbased techniques is the variability of the skin color in different lighting conditions. This frequently results in undetected skin regions or falsely detected non-skin textures. The problem can be somewhat alleviated by considering only the regions of a certain size (scale filtering) or at certain spatial position (positional filtering).

Another appearance based approach presented in the literature [49], that uses Gabor filters for extraction of hand gesture features. Gabor filters can capture the most important visual properties such as spatial locality, orientation selectivity and spatial frequency. Due to the high dimensionality of features, principle component analysis is used for feature reduction. The drawback of this and other similar approaches is that these methods are not invariant to translation, rotation and scaling. Moreover, these approaches are also effected by illumination variation. In [50], cascade of classifier approach is used. Each cascade is capable of detecting hands with certain angle of rotation. The drawback of this approach is that, it can not classify the same gesture with different viewpoints and is not rotation invariant. The authors of [51] extract a distinct and unified hand contour to recognize hand gestures, and then compute the curvature of each point on the contour. Due to noise and unstable lighting in the cluttered background, it has difficulties in obtaining segmentation of integrated hand contour. The eigen space is another technique, which presents a robust representation of a huge feature set of high-dimensional points using a small set of basis vectors. However, eigen space methods are not invariant to translation, scaling, and rotation. The most common and serious shortcoming of all of these methods discussed so far is that they only work with uniform background. These approaches lack in detecting hands in cluttered and dynamic environment.

Local invariant features are used for object recognition task. In the paper [52], to perform reliable matching between different views of an object or scene, a method is presented for extracting distinctive invariant features, as known as scale invariant feature transform (SIFT) features, that can be used for object recognition. This method for image feature extraction transforms an image into a large collection of feature vectors, each of which is invariant to image translation, scaling, and rotation, partially invariant to illumination changes and robust to local geometric distortion. Hartanto et al. [53] use SIFT features along with skin detection method for background subtraction and contours for localization of hand. Their matching stage is relatively simpler and hence, reports less than 70% accuracy for Indonesian sign language database. Their approach is computationally extensive and is not applicable for real time recognition. Dardas and Georganas [54] used bagof-features approach using SIFT features as keypoints and then used support vector machines to recognize the hand gesture. They segment the hand based on the skin color, discard the face using Viola-Jones face detector and then extract features. The shortcoming of this approach lies in segmentation of hand. It depends highly on illumination variation and the subject wearing half or full sleeves. Vision-based sensors use single or multiple cameras and depth sensors such as the Kinect to capture the gesture. Vision-based sensors suffer from the limitation that the gesture must b e performed in the camera's field of view in order to be detected.

Prior to the release of the Microsoft Kinect, vision-based gesture recognition used ordinary colour or grey scale images. These typically focused on static and dynamic hand gestures. Skin colour is used as a cue to segment the hand from the image [55]. However, as there is no depth information available, the majority of these systems require simple, solid backgrounds [56] or that the user wear a coloured or textured wrist band to aid in segmentation [57]. Another method of gesture recognition using colour images, presented by Drake [58] and Davis and Shah. [59], requires that the user wear a glove where each fingertip or hand joint is a different colour, making identification easier.

One of the challenges of using colour as a feature is its sensitivity to illumination changes. This makes robustness difficult to achieve, motivating the use of depth sensors together with the colour images.Since the release of the Microsoft Kinect in 2010, much of the gesture recognition work has b een based on depth sensors. These sensors provide depth information in addition to the RGB image captured by a traditional camera. The depth information aids in hand segmentation [60] and the Op enNI and NiTE SDKs provide skeleton information [61]. This has made depth sensors ideal for gesture recognition. A thorough review of gesture recognition is provided by Suarez and Murphy [62]. A sensor similar to the Kinect is pro duced by Asus. This sensor is known as the Asus Xtion Live Pro and has the same operating principle as the Kinect. However it do es not require an external power supply.

In 2013, Intel released a depth sensor for HCI known as the Creative Senz3D camera. This sensor is intended for closer range, from 0.01 m to 1 m. It uses time-of-flight to recover depth information from the scene rather than structured light employed by the Asus and first generation Kinect sensors. Given the novelty of the sensor there are few works that use it [63; 64] . The second generation of Kinect sensors that are shipped with the Xbox One, also now use time-of-flight for depth recovery. This sensor was made available to develop ers in late 2014; therefore there is not much literature available [65]. However, it rep ortedly has a higher depth resolution and range compared to the first generation Kinect. Another relatively new gesture device is the Leap Motion sensor. This sensor is also for close-range HCI, and is capable of tracking all fingers with an accuracy of up to a 100th of a millimetre over a range of 1 m. The operating principle is similar to that of the Asus, projecting a distinct IR pattern and using this to recover depth. Several works that use Leap Motion for HCI are presented in the literature [66; 67].

Whilst the earlier generation of depth sensors had several flaws, including low sensor depth resolution and reduced close-range depth recovery for applications where the user is further than 1 m, the Kinect is b est for HCI. The ideal system would therefore combine two sensors, one for close interaction and another for far interaction.

2.2.2 Gesture Recognition Methodology

Krueger [71] was the first who proposed Gesture recognition as a new form of interaction between human and computer in the mid-seventies. The author designed an interactive environment called computer-controlled responsive environment, a space within which everything the user saw or heard was in response to what he/she did. Rather than sitting down and moving only the users fingers, he/she interacted with his/her body. In one of his applications, the projection screen becomes the wind-shield of a vehicle the participant uses to navigate a graphic world. By standing in front of the screen and holding out the users hands and leaning in the direction in which he/she want to go, the user can fly through a graphic landscape. However, this research cannot be considered strictly as a hand gesture recognition system since the potential user does not only use the hand to interact with the system but also his/her body and fingers, we choose to cite this [71] due to its importance and impact in the field of gesture recognition system for interaction purposes. Gesture recognition has been adapted for various other research applications from facial gestures to complete bodily human action [72]. Thus, several applications have emerged and created a stronger need for this type of recognition system [72]. In their study, Dong [72] described an approach of vision-based gesture recognition for human–vehicle interaction.

The models of hand gestures were built by considering gesture differentiation and human tendency, and human skin colors were used for hand segmentation. A hand tracking mechanism was suggested to locate the hand based on rotation and zooming models. The method of hand-forearm separation was able to improve the quality of hand gesture recognition. The gesture recognition was implemented by template matching of multiple features. The main research was focused on the analysis of interaction modes between human and vehicle under various scenarios such as: calling-up vehicle, stopping the vehicle, and directing vehicle, etc. Some preliminary results were shown in order to demonstrate the possibility of making the vehicle detect and understand the humans intention and gestures. The limitation of this study was the use of the skin colors method for hand segmentation which may dramatically affect the performance of the recognition system in the presence of skin-colored objects in the background. Hand gesture recognition studies started as early as 1992 when the first frame grabbers for colored video input became available, which enabled researchers to grab colored images in real time. This study signified the start of the development of gesture recognition because color information improves segmentation and real-time performance is a prerequisite for HCI [73].

Hand gesture analysis can be divided into two main approaches, namely, glove-based analysis, vision-based analysis [74]. The glove-based approach employs sensors (mechanical or optical) attached to a glove that acts as transducer of finger flexion into electrical signals to determine hand posture. The relative position of the hand is determined by an additional sensor. This sensor is normally a magnetic or an acoustic sensor attached to the glove. Look-up table software tool-kits are provided with the glove for some data-glove applications for hand posture recognition. This approach [75] was applied to recognize the ASL signs. The recognition rate was (75%). The limitation of this approach is that the user is required to wear a cumbersome device and generally carry a load of cables that connect the device to a computer [76]. Another hand gesture recognition system was proposed in [77] to recognize the numbers from 0 to 10 where each number was represented by a specific hand gesture. This system has three main steps, namely, image capture, threshold application, and number recognition. It achieved a recognition rate of 89%. The second approach, vision-based analysis, is based on how humans perceive information about their surroundings [74]. In this approach, several feature extraction techniques have been used to extract the features of the gesture images. These techniques include Orientation Histogram [79; 80], Wavelet Transform [81], Fourier Coefficients of Shape [82], Zernic Moment [83], Gabor filter [84], Vector Quantization [85], Edge Codes [86], Hu Moment [87], Geometric feature [88] and Finger-Earth Movers Distance (FEMD) [89]. Most of these feature extraction methods have some limitations.

In orientation histogram for example, which was developed by McConnell [90], the algorithm employs the histogram of local orientation. This simple method works well if examples of the same gesture map to similar orientation histograms, and different gestures map to substantially different histograms [91]. Although this method is simple and offers robustness to scene illumination changes, its problem is that the same gestures might have different orientation histograms and different gestures could have similar orientation histograms which affects its effectiveness . This method was used by Freeman and Roth [91] to extract the features of 10 different hand gesture and used nearest neighbor for gesture recognition. The same feature extraction method was applied in another study for the problem of recognizing a subset of American Sign Language (ASL). In the classification phase, the author used a single-layer perceptron to recognize the gesture images. Using the same feature method, namely, orientation histogram, Ionescu et al. [74] proposed a gesture recognition method using both static signatures and an original dynamic signature. The static signature uses the local orientation histograms in order to classify the hand gestures. Despite the limitations of orientation histogram, the system is fast due to the ease of the computing orientation histograms, which works in real time on a workstation and is also relatively robust to illumination changes. However, it suffers from the same fate associated with different gestures having the same histograms and the same gestures having different histograms as discussed earlier.

In [92], the authors used Gabor filter with PCA to extract the features and then fuzzy-c-means to perform the recognition of the 26 gestures of the ASL alphabets. Although the system achieved a fairly good recognition accuracy (93.32%), it was criticized for being computationally costly which may limit its deployment in real-world applications [92]. Another method extracted the features from color images where they presented a real-time static isolated gesture recognition application using a hidden Markov model approach. The features of this application were extracted from gesture silhouettes. Nine different hand poses with various degrees of rotation were considered. This simple and effective system used colored images of the hands. The recognition phase was performed in real-time using a camera video. The recognition system can process 23 frames per second on a Quad Core Intel Processor. This work presents a fast and easy-to implement solution to the static one hand gesture recognition problem. The proposed system achieved (96.2%) recognition rate. However, the authors postulated that the presence of skin-colored objects in the background may dramatically affect the performance of the system because the system relied on a skin-based segmentation method. Thus, one of the main weaknesses of gesture recognition from color images is the low reliability of the segmentation process, if the background has color properties similar to the skin [[93]. The feature extraction step is usually followed by the classification method, which use the extracted feature vector to classify the gesture image into its respective class. Among the classification methods employed are: Nearest Neighbor , Artificial Neural Networks, Support Vector Machines (SVMs) , Hidden Markov Models (HMMs)[94] .

As an example of classification methods, Nearest Neighbor classifier is used as hand recognition method combined with modified Fourier descriptors (MFD) to extract features of the hand shape. The system involved two phases, namely, training and testing. The user in the training phase showed the system using one or more examples of hand gestures. The system stored the carrier coefficients of the hand shape, and in the running phase, the computer compared the current hand shape with each of the stored shapes through the coefficients. The best matched gesture was selected by the nearest-neighbor method using the MED distance metric. An interactive method was also employed to increase the efficiency of the system by providing feedback from the user during the recognition phase, which allowed the system to adjust its parameters in order to improve accuracy. This strategy successfully increased the recognition rate from (86%) to (95%). Nearest neighbor classifier was criticized for being weak in generalization and also for being sensitive to noisy data and the selection of distance measure [**?**].

To conclude the related works, we can say that hand gesture recognition systems are generally divided into two main approaches, namely, glove-based analysis and vision-based analysis. The first approach, which uses a special gloves in order to interact with the system, and was criticized because the user is required to wear a cumbersome device with cables that connect the device to the computer. In the second approach, namely, the vision-based approach, several methods have been employed to extract the features from the gesture images. Some of these methods were criticized because of their poor performance in some circumstances. For example, orientation histograms performance is badly affected when different gestures have similar orientation histograms. Other methods such as Gabor filter with PCA suffer from the high computational cost which may limit their use in real-life applications. In addition, the efficiency of some methods that use skin-based segmentation is dramatically affected in the presence of skin-colored objects in the background. Furthermore, hand gesture recognition systems that use feature extraction methods suffer from working under different lighting conditions as well as the scaling, translation, and rotation problems.

2.3 Literature Review in Human-Robot Interaction System based on Gestures

Liu et al. [96] proposed a modular model ofan HGR system for human-robot interaction. This model consists of five steps:

- sensor data collection,
- gesture identification,
- gesture tracking,
- gesture classification, and
- gesture mapping.

The first four steps represent hand gesture recognition, while during the gesture mapping step, the recognized gesture label is translated into a set of robot hand control commands.

Many publications focus on HRI applications of gesture recognition. One of the first works in this topic can be found in [78], in which both, person and arm tracking in color images was performed. Two recognition methods were compared, one template-based approach and an artificial neural network, both combined with a Viterbi algorithm. An approach to moving gesture's recognition is presented in [92], where a Kinect sensor is used to recognize gestures while the robot is moving. The method tracks the face of the person in order to perform background subtraction and then joint positions are estimated by means of a Voronoi diagram. A generated motion context is used to train a Multi-Layer Perceptron (MLP) classifier in order to recognize similar gestures to the ones proposed here. A low cost RGB-D sensor is used in [68] to perform dynamic gesture recognition by skeleton tracking. The recognition method uses a Finite State Machine which encode the temporal signature of the gesture. An adaptive method was developed for identifying the person which is performing the gestures. The goal was to learn from gestures and therefore adapt the system to the specific person, being the same gesture performed by two different persons understood in different ways, even having the opposite meaning. Another Kinect application to gesture recognition with HMM (Hidden Markov Models) and skeletal data is presented in [69], in which the user performs gestures to control the robot and it responds with voice or a message in the display.

Deep neural networks have also been used to recognize gestures, as done in [70], aiming to recognize gestures in real time with minimal pre-processing in RGB images. They show high classification rates working online, the application being a robot that gives speech feedback. User defined gestures can be added in a semi supervised way to the system from [97], which contributes a non-parametric stochastic segmentation algorithm, the Change Point Model. This procedure does not need to be supplied with the gesture's starting and ending points, making the user able to create its own gestures to control a robot and thus being highly customizable without the need of explicit user learning or adaptation.

Elderly assistance is another interesting field in which service robotics is applied, and gesture interaction may be really useful in such case. A Kinect based approach to recognize calling gestures is proposed in [98]. This approach use a skeleton based recognition system to detect when the user is standing up, and an octree one when the skeleton is not properly tracked. Erroneous skeletons are filtered by face detection in order to determine whether the data is actually a person or a false positive. An application to object handling to the user is implemented and tested with different elder users. Besides, some contributions are only concerned with hand gestures. The hand gesture recognition system introduced in [99] performs gesture classification in each arm independently, using two artificial neural networks, as well as HMMs, to perform arm tracking. Their trajectories are used as the input to the classifier. Another hand gesture decomposition application to HRI is proposed, in which a color segmentation algorithm is used to find skin regions and a cascade of Adaboost classifiers is used for the hand posture. The method was validated in a museum robot guide. An RGB-D camera is used in [100] for hand gesture recognition. Human segmentation is performed by background subtraction and hand tracking is then calculated from both color and depth information.

Some static gestures are employed to indicate the start and end of the gesture to the system, such as opening and closing the hand. The trajectory followed in the meantime is then used to recognize the gesture by applying an HMM, as it was similarly done in [101]. A tour-guide robot able to understand gestures and speech feedback is introduced in [102], which tracks the user using depth information and performs the recognition with a Finite State Machine gesture modeling. Hand tracking approaches as well as the related work being developed in Microsoft Research with a Kinect v2 sensor [9], which shows impressive results, may imply a great improvement in the recognition of hand gestures, with applications to sign language recognition – which is another application of gesture recognition systems –. Cooperation tasks is another research topic, as in the case of the current work, when the user cooperates with the robot to achieve a given task, for instance approaching the desired object. The system proposed detects a person with a color camera to recognize the face and laser range finders to find his legs. Then, the person perform gestures to make the robot guide him or to carry a load together. A similar study is conducted in [103], in which they evaluate the effect of the robot utterances when they are accompanied by gestures such as the robot looking to the person when he speaks or pointing in the direction of an object.

In [104], an image based human-robot collaboration system is proposed using a Kinect mounted on a wifibot which carries a NAO robot. The robot is able to navigate towards an object pointed on the floor. The proposed facial tracker fails to detect gestures quite often, as untrained users make those gestures subtly. Besides, no physical interaction between the human and user is present in this work. A similar – but hardware demanding – scenario is proposed in [27] where three Kinects are mounted around the workspace of a mobile robot. The authors detect dynamic gestures using a FastFourier Transform which is used to segment a gesture, through the estimation of its period. This requires continuous repetition of each gesture in a loop several times in preoperational/training phases. Training neural networks only on pre-processed images can prevent them from extracting and learning diverse features and may compromise the detector performance during recognition phase. The authors train 10 different neural networks for each gesture and this requires substantial resources. Moreover, this scenario involves no physical interaction between robot and human. In [28], Kinect-based object recognition through 3D gestures is proposed. The OpenNI and NITE middleware are used to extract the skeleton information of the human standing in front of the camera. The object location is fixed and a rigid object segmentation procedure is used with predefined constraints (e.g., the table color is white). Such conditions may not always be present in a real human-robot interaction task. Also, the removal of background using depth information may fail in some conditions (e.g., when the human operator stands near a wall). The objects chosen in the demonstration appear to be only of rectangular/box shape thus are detected using corner detectors. The histogram matching algorithm is used to recognize the objects. This has been outperformed by modern deep learning techniques like Convolutional Neural Networks (CNN).

Recently, [29] makes use of CNN for hand gesture recognition for a Human-Computer Interface. The author proposes a color independent classifier by feeding a pre-processed binary images into a LeNet network [30]. This makes classification accuracy dependent on the pre-processing step although, if provided with sufficient data, CNN are inherently robust enough to learn color features. In [105], the authors propose a HRI system for navigation of a mobile robot using a Kinect. For body and hand skeleton detection, a skeleton topology with multiple nodes is fit on the point cloud acquired from the sensor. This technique is not reliable, as both skeleton and hands have several non-linear anatomical constraints, that make the task of accurate pose detection difficult. A system targeting pHRI is proposed , where a human-user gives commands to a robotic arm to follow, grasp, move and place an object. The arm gestures are used to control the robot, so that the gestures are distinguished with respect to predefined elbow angle ranges. The method does not incorporate hand gestures detection. This also makes the interaction system less intuitive for comfortable human-robot interaction tasks, as the human operator will have to learn the required elbow angles. Besides, a color-coordinate based algorithm is proposed for object detection, limiting the detection of multicolor objects.

In [43], the author uses the skin color for hand segmentation assuming a planar background. Although the skeleton of the hand is extracted using distance transform, the approach only works with open hand gestures and mostly when the palm is facing the camera. The authors of [44] propose a technique to navigate a mobile robot with a Kinect. The OpenNI middleware is used to extract hand position and no physical interaction is present. The localization of human body and of its subparts (e.g., hands or face) depends on the choice of sensor used and on its output. [45] use Microsoft SDK; object searching is done on the basis of color and shape of the object point cloud. Tracking of hands and fingers can also be done by optical and infrared based sensors like Leap Motion. This has a hand model built-in, which is combined with the raw sensor data to track the positions and motion of the hand precisely. However, the effective range of this sensor is only 25 to 600 millimeters approximately which is not always suitable for the distant interactive applications between humans and robots.

Applied to elderly care scenarios, a Kinect based approach to recognize calling gestures [46] is proposed which used a skeleton-based recognition system to detect when the user is standing up, and an octree when the skeleton is not properly tracked. [106] applied gesture recognition methods for human-multirobot interaction using the Euclidean distance between the hand and the neck joints, the angle in the elbow joint, the distance between the hip and the hand joints, and the position of the hand joint. [107] recognized ten gestures from visual signals used by the army to control a mobile robot performed by Neural Network classifiers using quaternions and angles. [108] used the Euler angle to recognize left arm gestures to control a Pioneer robot with four gestures (come, go, rise, and sit down) using joint angles (left elbow yaw and roll left shoulder yaw and pitch).

Some methods [108] are based on detecting and counting which of the 5 fingers are extended. [110] proposed an algorithm that extracts fingers from salient hand edges and high-level geometry features from hand parts. More recently, [111] proposed a skeleton-based approach for a dynamic hand gesture recognition system that used the shape of connected joints, histogram of hand directions and wrist rotations. [112] have collaborated in the development of a robot that can look towards and then approach a person who waves a hand at the robot in a laboratory setting, to our knowledge, there has not yet been a robot designed that can respond to a person's request through hand waving or other embodied means in naturally occurring multiparty settings. [113] have reported the results of interaction between caregivers and the elderly and have discussed the implications for developing such robotic system. Based on these findings, [114] have developed a system that can initiate interaction with a human through nonverbal behaviors- such as head gestures (e.g. head turning) and gaze. Most of the above methods focus on one person performing the gesture in an empty workspace [115].

However, in our approach, we propose a system that works in cluttered and noisy environments that include moving people and multiple natural hand motions. In addition, most of the aforementioned works used pre-defined gestures under controlled environments, fixed the position of the person to perform the gesture, and defined start and end-points in of the gesture. In our work, the person who performs the calling gesture can be in various positions and locations may be varied.

2.4 Concluding Remarks

This chapter covers the main topics related to Human-robot interaction using gestures as stated in the literature. The first part of this chapter provides an overview of service robots, human-robot interaction for elderly and gestures and hand gesture recognition. The second part describes the conventional hand gesture recognition approaches. Then, the last part of the chapter concentrates on literature review in human-robot interaction system based on gestures.

Chapter 3

Natural Calling Gesture Recognition

Based on the previous works, we propose a natural calling gesture recognition approach using skeleton features in crowded environments, processed in two-stages. First, the skeleton key-points of individual people are obtained from the OpenPose real-time detector [116] by using the camera. And then, to remove the false alarms, according to the hand-wrist position, the gaze is calculated, and fingertip positions are extracted by zooming into the hand-wrist. Finally, we classify using the key-points of the fingertips whether something is a calling gesture or not. Through observations and experiments on a realistic environment, we devised robust heuristics that are effective and computationally inexpensive.

3.1 Algorithm Overview

In our process of hand calling gesture recognition, suitable techniques are employed to implement and classify the hand gesture in the system. We observe that in real world, to call someone, firstly, we must tend to face the one who we want to call. If the person does not have any intention of calling, he/she may not move his/her arm against gravity. And, when the person calls someone, it is natural to direct his/her hand open at the one being called. Based on these findings, we propose our hand calling gesture recognition system. Our approach can be divided into some general steps: body key-point acquisition, detection of gaze and hand wrist position, zooming into the hand wrist part, and hand fingertip key-point acquisition and calling gesture recognition as in Figure 3.1. First, we extract the body and hand key-points of individuals in the scene as in Figure 3.2 and Figure 3.3. Our system first detects the gaze of people that are looking at the robot to find candidate people that are more likely to be calling the robot. This removes the need to tracking all the people in the scene since we only look at some candidate people. Then, we localize the hand-wrist positions. We not that only people with clearly defined hand, wrist, and arm positions are considered because someone calling the robot would likely try to make his/her hand clearly visible to the robot.

Specifically, we look for positions of the wrist that are higher than the position of the elbow and at the same time, the position of the elbow is either higher or lower than the position of the

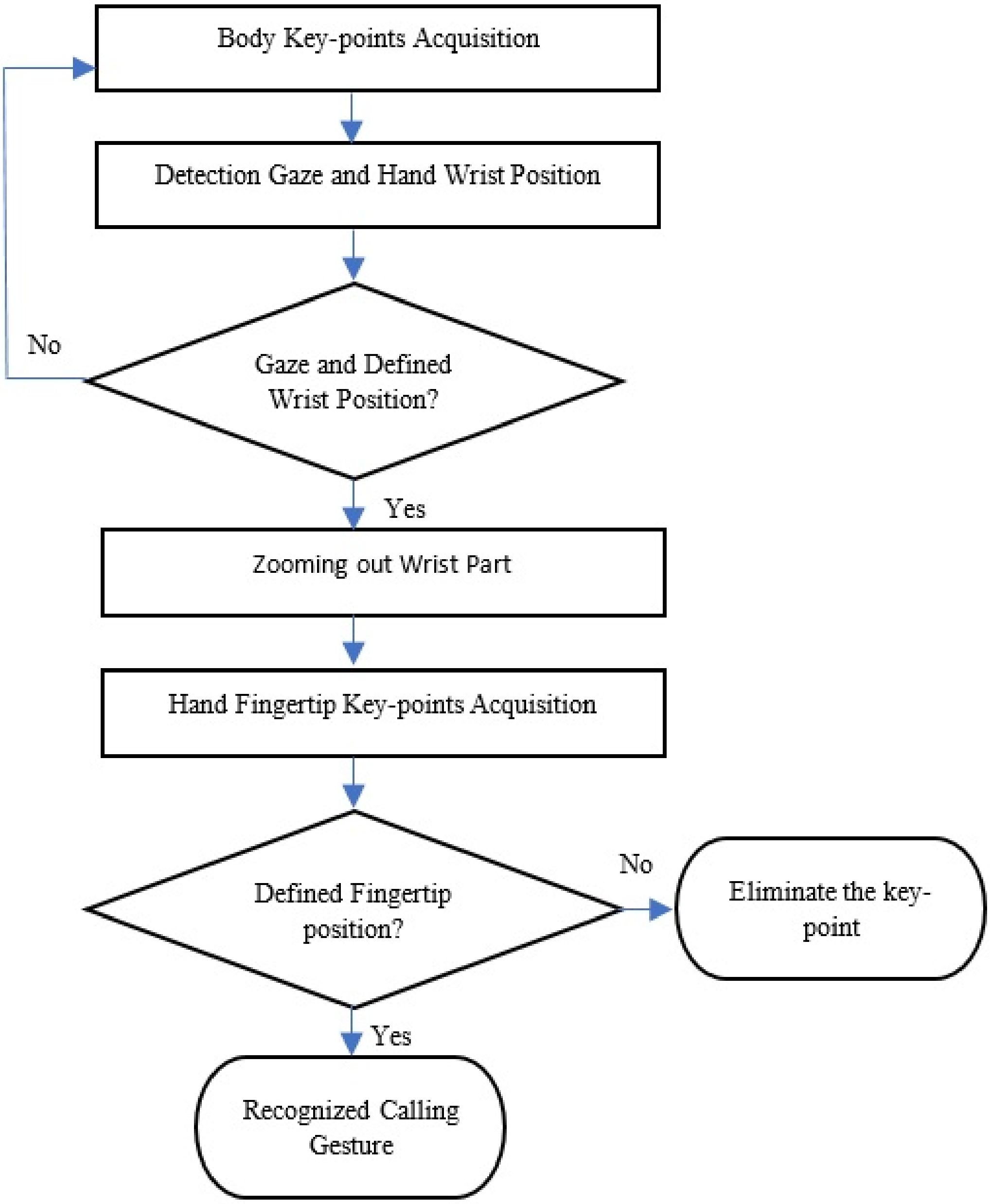

Body Key-points Acquisition
Detection Gaze and Hand Wrist Position
No
Gaze and Defined Wrist Position?
Yes
Zooming out Wrist Part
Hand Fingertip Key-points Acquisition
Defined Fingertip position?
No
Eliminate the key-point
Yes
Recognized Calling Gesture

Figure 3.1: Overall Process of Natural Calling Gesture Recognition

shoulder. By detecting key gaze and hand-wrist positions, we can reduce the need to track other hand motions. In the final step, we look at the configuration of the fingertips for classification. Note that there are times when the tracker misses detailed of the hand in crowded scenes because the hand area is too small. To solve this, we zoom into the hand part of the candidate person based on the locations of body key-points. After zooming in the hand-wrist part and extracting detailed key-points of the fingertips, the system calculates positions of the fingertips. Finally, depending on the positions of the fingertips, our approach recognizes calling gestures.

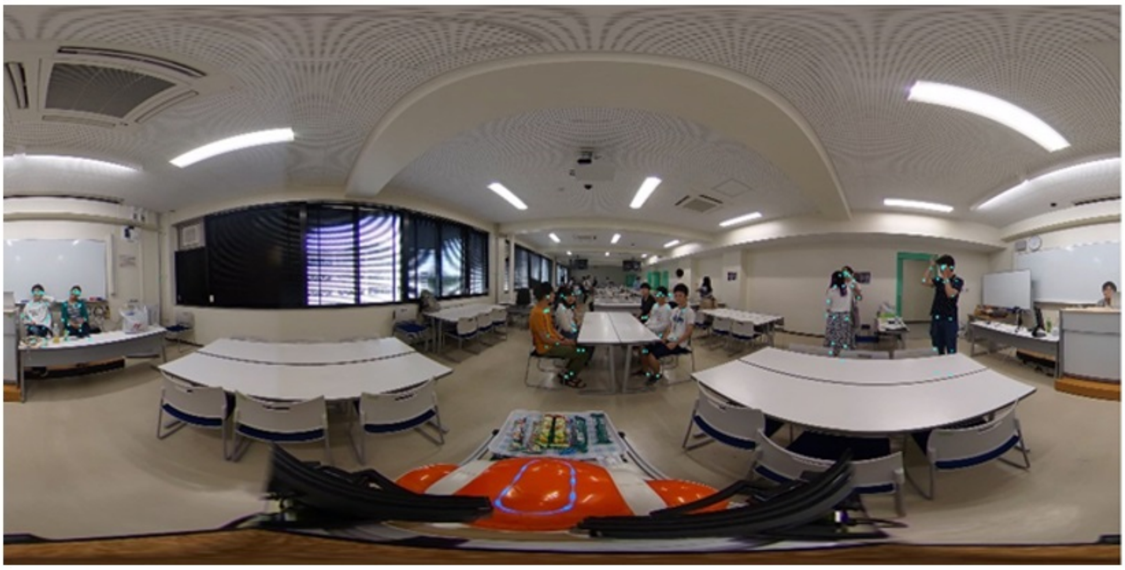


Figure 3.2: Example of Key-point of Individual Person in the Scene using Omnidirectional Camera

3.2 Body Key-points Feature Acquisition

This is the first step in our calling gesture recognition system that utilizes video from an omni-directional camera. First, we extract body key-points information from the OpenPose detector using the omni-directional camera. By using the omni-directional camera, the system can also detect callers that are behind the robot. Using Openpose, we can obtain both key-points of the body and hand of the individual person in the scene as shown in as Figure 3.4.

3.3 Detect Gaze and Find Hand-Wrist Position

This is a key step in our hand gesture recognition system. In the real world, to call someone, firstly, we tend to face the one who we want to call. And then, we perform some hand calling gestures. Based on this observation, our work first detects gaze directed towards the robot and finds the hand-wrist positions that could be indicative of a calling gesture. This removes need to track all the people in the scene.

For gaze detection, we use the location of the two eyes, two ears and nose. When the user is looking straight ahead, the angles between two eyes and two ears are as shown in Figure 3.5(a). When looking at the left, the eye positions move to the left and the angles changes as in Figure

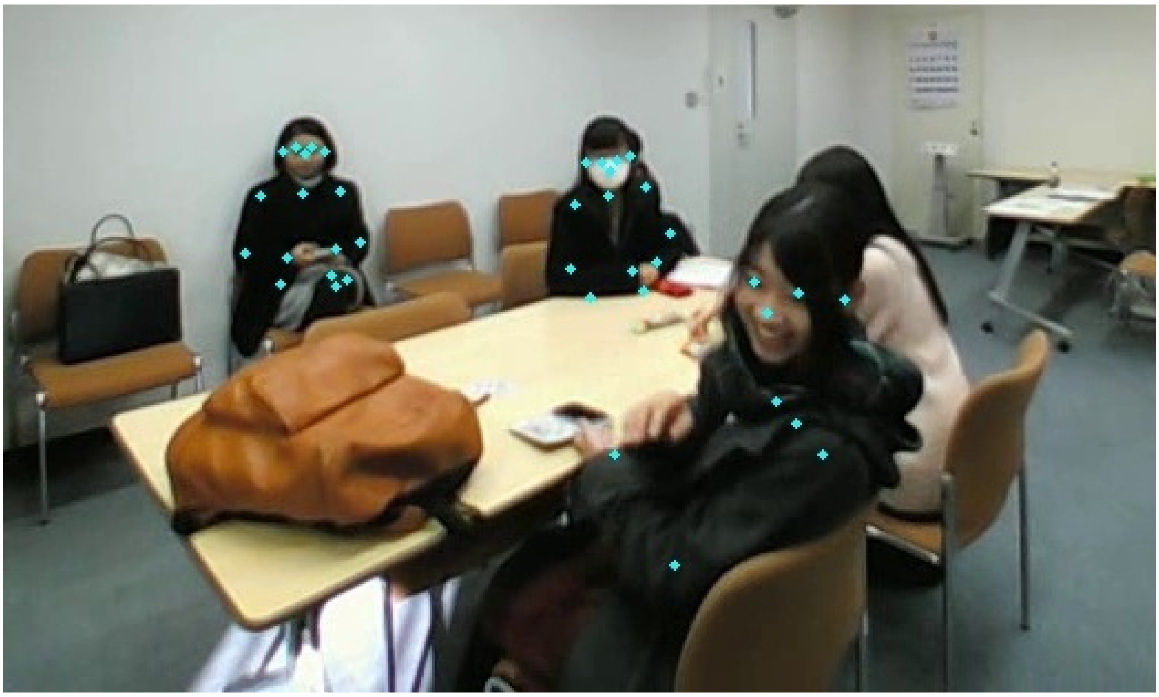


Figure 3.3: Example of Key-point of Individual Person in the Scene using USB Web-camera

3.5(b). When looking at the right, the opposite situation occurs in Figure 3.5(c). By detecting the angles between the two eyes and the positions of the two ears, it is, therefore, possible to measure the gaze direction.

If a person has no intention of calling, he/she may not raise his/her arm against gravity. So, if the hand is raised up, this action could be an intentional one. Such cases are then considered candidates of calling gestures. The hand-wrist position may take on two types of patterns. For the first one, the position of the hand-wrist is higher than the position of the elbow. And, the position of the elbow is higher than the position of the shoulder. For the second one, although the position of the wrist is higher than the position of the elbow, sometimes, the position of the elbow is lower than the position of the shoulder. Figure 3.6 shows examples of the hand-wrist position that are calling gestures.

3.4 Zoom into the Wrist Part and Hand Fingertip Key-points Features Acquisition

Sometimes, OpenPose misses small details of the hand in crowded scenes because the hand area is small, and the key-points are not tracked. We find that in practice, it is possible to get the fingertip position if the hand is near the camera (e.g. ˜1.5m). Otherwise, we find that detailed positions of the fingertips are generally missed. For that reason, after detecting the gaze and candidate calling hand-wrist position, our system zooms into the hand-wrist part to perform key-point detection of the fingertips.

Although the defined hand-wrist position of Figure 3.6 are indicative of potential hand calling gestures, without looking closely at the fingertip configurations, we cannot confirm whether

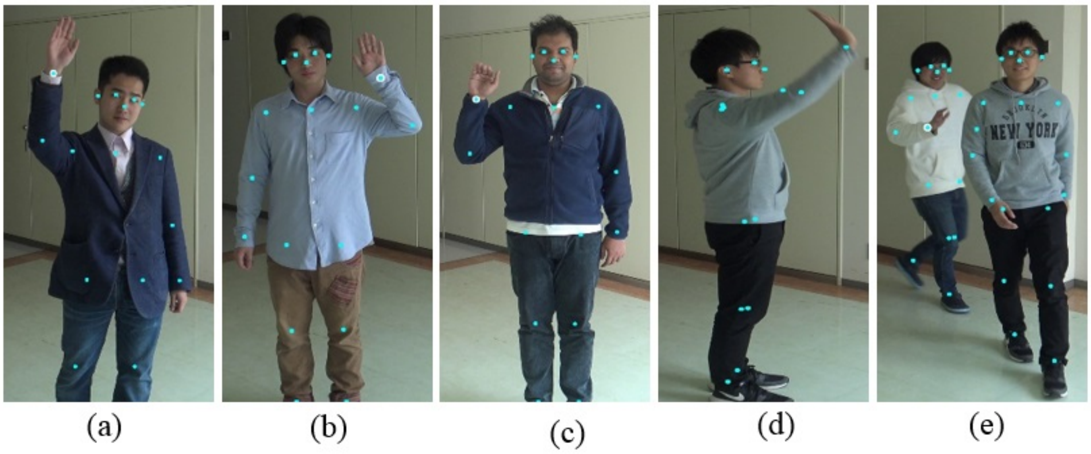


Figure 3.4: Examples of Calling Gestures and Similar Hand Motions with Key-points

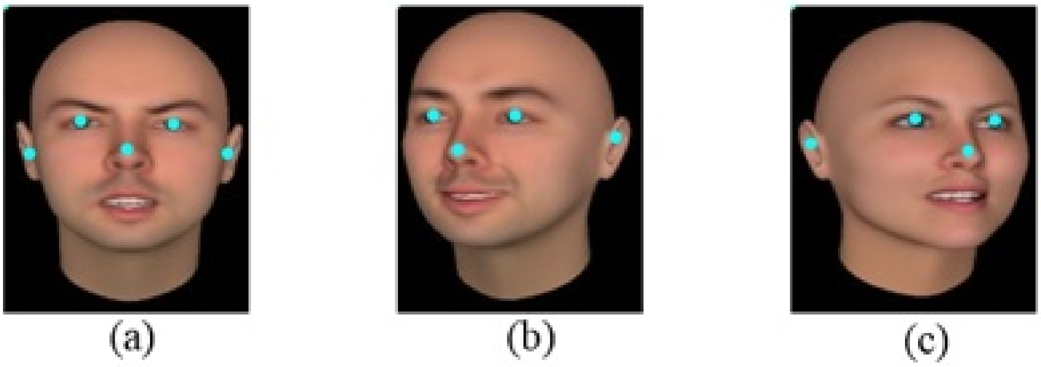


Figure 3.5: Example of Positions of Fingertips

they really are calling gestures. For instance, in that position, a person could be making a phone calling gesture or any number of other gestures. The fingertips position in such cases are not the same as a calling gesture. We observed that the typical poses of fingertips of calling gestures are open (e.g. Figure 3.7 (a), (b)). For that reason, we apply OpenPose a second time to get the detailed key-points of the fingertips.

After detecting the key-points of the fingertips, the system calculates the coordinates of the fingertips relative to the base of the palm. However, we find that even if the same gestures are captured, the coordinates of the fingertips can be different. Likewise, even between different gestures, the coordinates can be similar. As a result, our algorithm takes the relative coordinates and calculates the direction and degrees in the captured fingertips for accurate gesture recognition. (See section 3.5 for details.) Our idea of what kinds of open gestures is illustrated as follows: The pose of the fingertips of calling gestures are typically open like in Figure 3.7 (a) and (b). Although Figure 3.7 (c) and (d) are similar to the open-type poses, we do not consider these to be calling gestures. Based on this, Figures 3.8(a-c) show the fingertip positions of the calling gesture and Figure 6d is not a calling gesture.

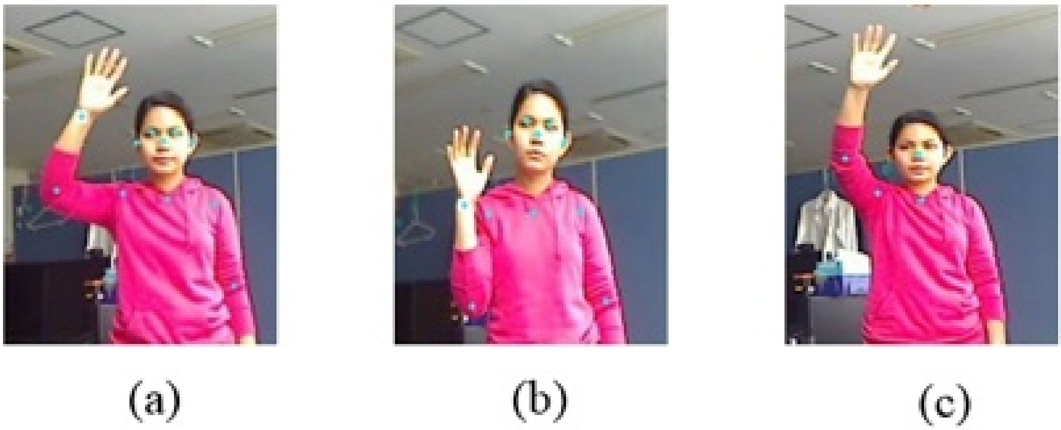


Figure 3.6: Example of Hand-wrist position

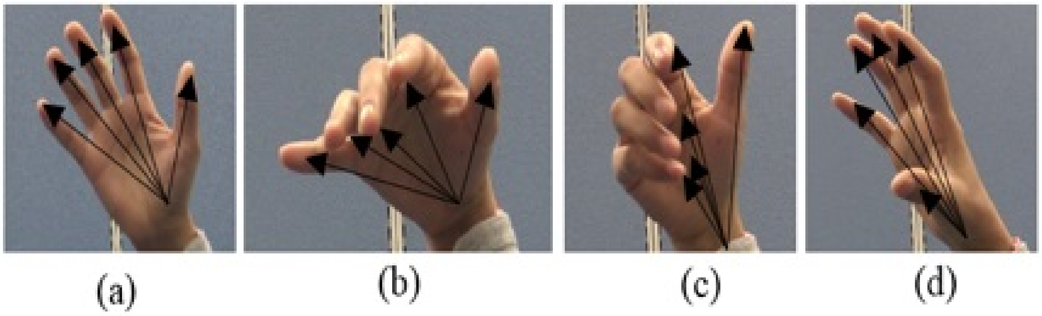


Figure 3.7: Example of Positions of Fingertips

3.5 Recognized Calling Gestures Using Rule-Based Classification

Rule-based classification provides a set of encoded rules extracted from input gestures to compare the feature inputs. Firstly, our rules consider whether the person's gaze is towards the camera. And then, whether the hand is in a position that would potentially be of a calling gesture or not. If the gaze is detected, it will classify the calling gesture according to the rules shown in Figure 3.9.

To classify calling gestures, our approach uses the x-coordinates of each fingertip and the angles from the base of the palm to the thumb and to the little finger. Firstly, we consider whether the order of the x-coordinate values of each fingertip (from the thumb to the little finger) are in order or not. In the 2D image plane, we want the left to right order of the fingertips to start at the thumb and end on the little finger, or in the reverse order. Figure 3.7(d) shows and example of "out-of-order" fingers. If the order is either ascending or descending, the angle of the vector from the base of the palm to the thumb is between 30 and 70 degrees, and the angle of the vector from the base of the palm to the little finger is between 0 to 45 degrees, we consider that a calling gesture.

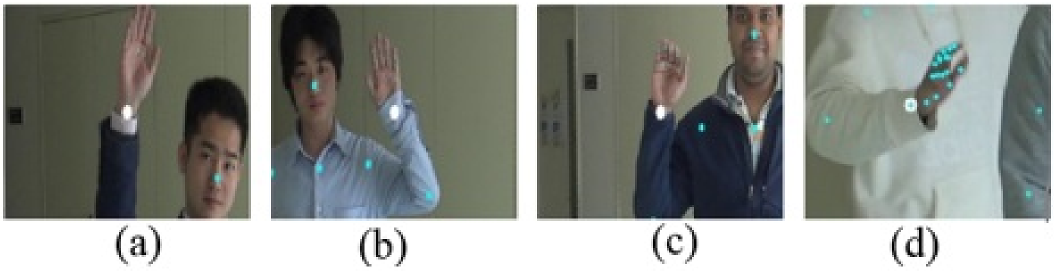


Figure 3.8: Hand Fingertips Key-points Features Acquisition

3.6 Concluding Remarks

In this chapter we presented natural hand calling gesture recognition algorithm in crowded environment. In details, this chapter explains the step by step processes of the method - body key-points feature acquisition, gaze detection and finding hand-wrist position, zooming in the hand-wrist part and fingertip key-points features acquisition and recognizing of calling gestures using rule-based classification method.For each step, the detail process is described with illustrations.

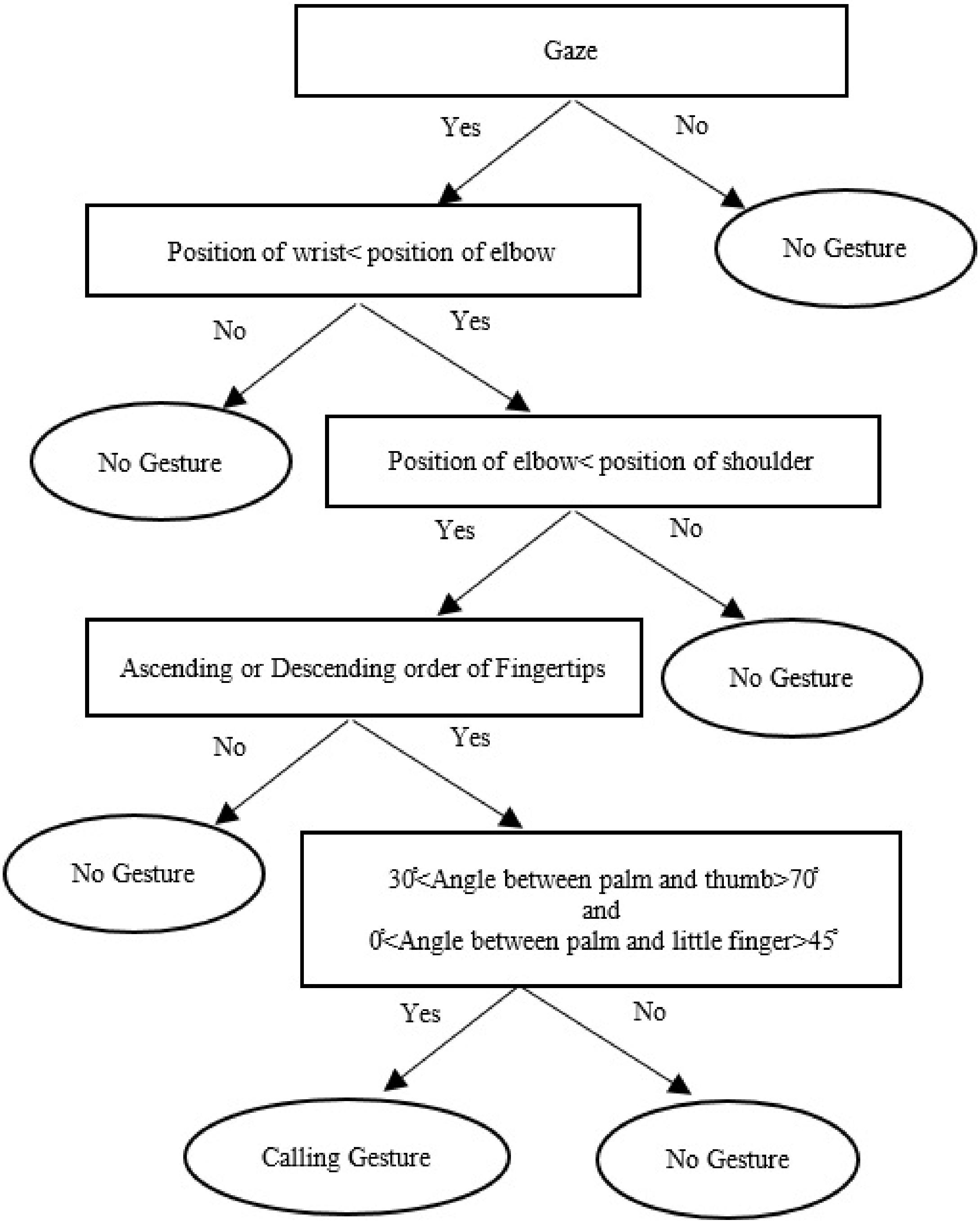

Gaze
Yes
No
Position of wrist< position of elbow
No Gesture
No
Yes
No Gesture
Position of elbow< position of shoulder
Yes
No
Ascending or Descending order of Fingertips
No Gesture
No
Yes
No Gesture
30°<Angle between palm and thumb>70°
and
0°<Angle between palm and little finger>45°
Yes
No
Calling Gesture
No Gesture

Figure 3.9: Rules of Calling Gesture

Chapter 4

Human-Robot Interaction System

Human-robot interaction (HRI) is the study of communication between robots and humans. It exists at the intersection of the fields of artificial intelligence, computer vision, robotic design, social science, and the humanities. Robots are increasingly becoming involved in more complex tasks and activities, sometimes requiring interaction with people to complete these tasks.

4.1 Service Robot Setup

The current prototype of our service robot consisting of an i-Cart mini (T-frog), USB webcam, a humanoid robot (Aldebaran's NAO) and a mobile workstation, is shown in Figure 4.1. The i-Cart mini (T-frog) is a small, light weighted robot that can easily move. To control the robot's hardware, we used the Robot Operating System (ROS), which is a framework with software libraries and tools designed specifically to develop robot applications. The USB wide-angle webcam allows 120°wide coverage to cap-ture the whole body of each person in the scene. Hokuyo Laser sensor is used to measure the distance between the caller and robot. The NAO robot will be used for our future work to interact with the caller once the robot has arrived the caller's location.

The system is configured as follows: The USB wide camera, mounted on top of the robot, obtains a subject's full-body position data and Hokuyo laser sensor's data is used to calculate the distance between the caller and robot. Such data becomes available to all devices connected throughout the Robot Operating System (ROS) network and subsequently, it is processed to recognize the body gestures that will indicate the robot's actions.

4.2 Process Model of Human-Robot Interaction Based on Gesture

To recognize gestures in HRI, it is beneficial to investigate into a generic and simplified information processing model. As shown in Figure 4.2, the generalized human information processing is broken

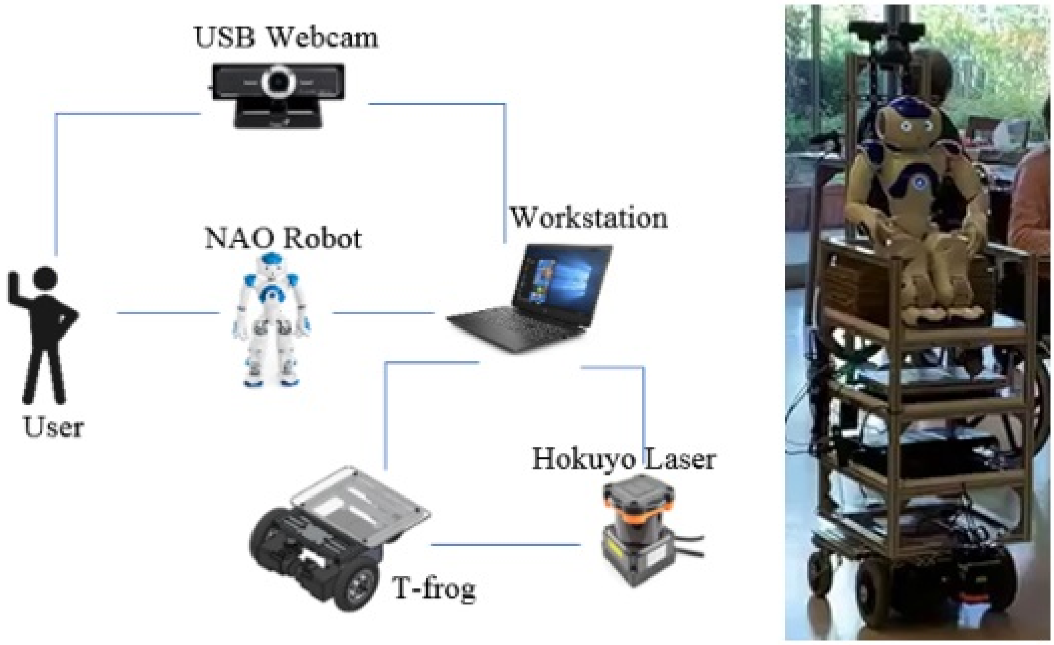


Figure 4.1: System Architecture Showing Interactions between Components

into a four-stage model. Based on this model, we propose a specific model for gesture recognition in HRI.

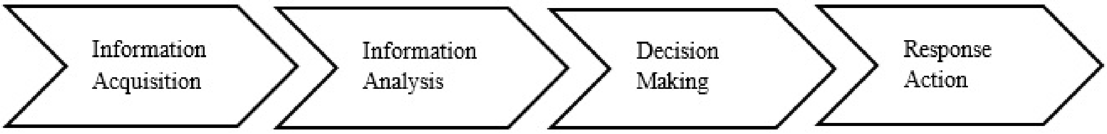


Figure 4.2: A Four-stage Model of Information Processing

Figure 4.3 shows the process model of the proposed service robot with gesture recognition system. Firstly, let us say there is a natural calling gesture within the camera's field of view. Then it is detected in real time and the key-points information of the whole body are obtained, i.e., positions of nodes in the skeleton model detected by OpenPose [116]. And, gaze and hand detections are performed based on the key-points information of individual people. After detecting the gaze and hand, the detailed key-points of the fingertips are extracted by zooming into the hand-wrist part. After detecting potential calling gestures, the robot faces and approaches to the candidates of calling gestures by transmitting data to the motion control system. At that time, the robot approaches the potential caller and further verifies whether they the caller was really calling to reduce false positives. After determining whether there was an actual call or not from the hand gesture's command, the robot responds to the action by moving to the ones who are calling the robot. We then map the recognized gesture to the appropriate command – so that if the caller continues the calling hand gesture while approaching, the robot will keep coming to the caller. And, if the caller still holds gaze to the robot while approaching, the robot will continue coming action. Like this action, service robots can help the elderly.

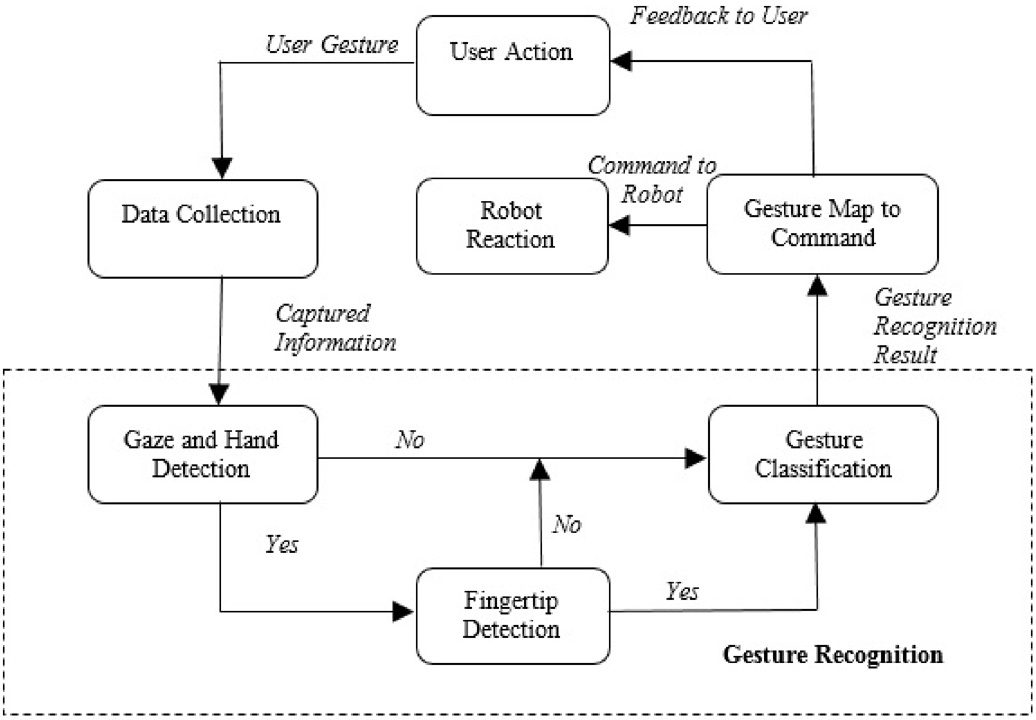


Figure 4.3: A Process Model of the Proposed Service Robot with Gesture Recognition

Algorithm 1 Procedure for Human-Robot Interaction

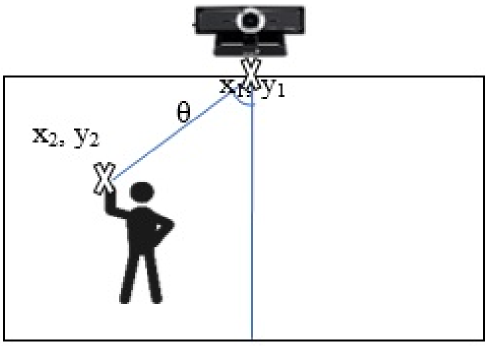
```
open human-robot interaction switch;
loop:
capture skeleton data and feed data into system, and move around the people;
if no calling gesture then
    goto loop;
else
    turn toward direction of caller and move towards caller;
    if user keeps calling or holding gaze then
        continue moving towards caller;
    else
        goto loop;
    end if
end if
```

4.3 Gesture Recognition

In this step, the system uses the skeleton key-points of individuals among the crowd using OpenPose that as in our work (in Chapter 3). Firstly, we extract the body and hand fingertip key-points of individuals in the scenes. The gaze of people that are looking at the robot is detected to find candidate people that are more likely to be calling the robot. We then localize the hand-wrist positions. Specifically, the system looks for positions of the wrist that are higher than the position of the elbow and at the same time, the position of the elbow is either higher or lower than the position of the shoulder. In the final step, the system looks at the configuration of the fingertips for classification. So as to not to miss fine details of the hand in crowded scenes, we also zoom into the hand part of the candidate person based on the locations of body key-points. After zooming into the hand-wrist part and extracting detailed key-points of the fingertips, the system calculates positions of the fingertips. Finally, depending on the positions of the fingertips, our approach recognizes calling gestures.

4.4 Measuring Distance and Angle

Firstly, the angle between the caller and robot, , is calculated by using the hand-wrist coordinates. And, using this angle and Hokuyo laser's data, the system calculates the distance between the caller and robot.

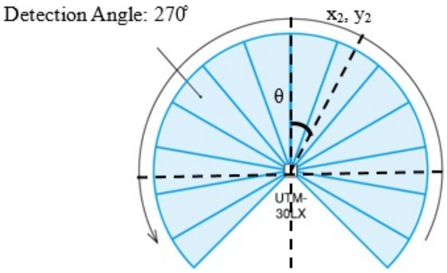


$$\theta = \tan^{-1} \frac{|(x_1 - x_2)|}{(y_2 - y_1)} \tag{4.1}$$

where:

θ = the angle between the caller and robot
x_1 = half of image horizontal dimension (e.g. if 640 $\times 480, x_1 = 320$)
$y_1 = 0$
x_2 = x coordinate of detected calling hand-wrist
y_2 = y coordinate of detected calling hand-wrist

In Equation (1), x_1 is defined as half of image horizontal dimension and y_1 is 0 because the camera is mounted on the center of the i-Cart mini. The direction of the angle,, depends on the sign of (x_1-x_2). If the sign is '+', the direction is right, otherwise, the direction is left. The system calculates the distance between the caller and robot by using the angle, , x, y coordinates of detected calling hand-wrist and Hokuyo laser's data. Hokuyo laser sensor can detect range from 0.1m to 30m within 270 wide angle. When the system implements on ROS, the obstacles for navigation can be avoided.



The system calculates the distance between the caller and robot by using the angle, , x, y coordinates of detected calling hand-wrist and Hokuyo laser's data. Hokuyo laser sensor can detect range from 0.1m to 30m within 270°wide angle. When the system implements on ROS, the obstacles for navigation can be avoided.

4.5 Concluding Remarks

This chapter explains human-robot interaction system based on natural hand calling gestures for elderly care. This chapter discusses about the system architecture for human-robot interaction service robot system, the process model of the proposed service robot with gesture recognition, algorithm procedure for human-robot interaction system for this research, and measurement of distance and angle between human and robot in details.

Chapter 5

Experimental Results and Analysis

5.1 Experiment Analysis of Natural Hand Calling Gesture Recognition Method

In order to test the robustness of our natural calling gesture recognition algorithm, we set hundreds of videos in several specific situations and evaluated tests on them shown in Figure 5.1. In addition, we determined how well the gesture recognizer could differentiate the normal actions from calling gestures. The proposed algorithm is based on spotting-less gesture recognition. That is, the algorithm does not have to detect the start and end position of gesture. This fact may be a great advantage to the recognizer, but the probability of unrecognized calling gesture will increase at the same time.

5.1.1 Experiment 1

The first experiment presents a situation in which the camera is located at a fixed position. Table5.1 shows the characteristics of each situation. In Table 1, S and W means sitting and walking respectively. The resolutions of all video are 1920×1080.

Table 5.1: Characteristics of Each Video

Video ID	V1-V10	V11-V20	V21-V30	V31-V40	V41-V50	V51-V60	V61-V70	V71-V80	V81-V100
No. of person	1	2	3	4	1	2	3	4	¿4
Situation	S	S	S	S	W	W	W	W	S, W
Distance	2m	2m	2m	2m	2m	2m	2m	2m	3m
FPS	30	30	30	30	30	30	30	30	30

Figure 5.2 shows sample results of the accuracy of our algorithm in three different scenes. The quantitative performance of our algorithm is shown in Table 5.2. The reason we recorded the videos at a high resolution 1920 × 1080, was to get the hand part clearly. We find that a body is

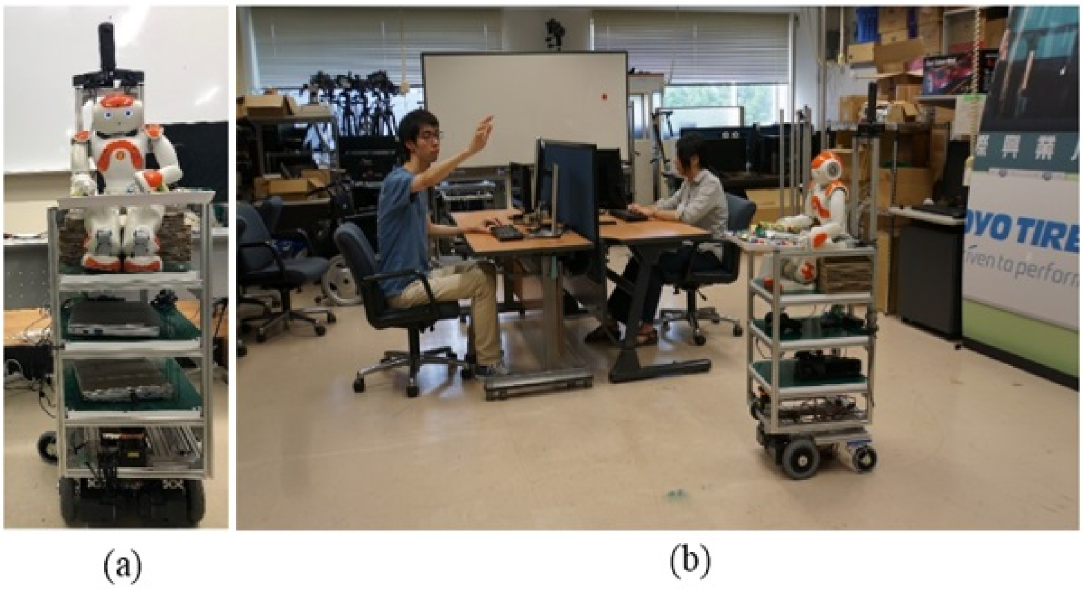


Figure 5.1: The Robotic System Designed for This Work, (b) Example of Recording Video

usually correctly detected if it is within 2m.

With a gesture recognition accuracy of 70% in the case of sitting and walking with more than four people, recognition performance is notably worse. This can be explained by the varied motion of the people. If people walk fast, we cannot correctly detect the body key-points. If we cannot get the correct key-points, our algorithm reduces in performance. In addition, we may sometimes miss the hand fingertip key-points if the calling gesture is dynamic. Also, if gaze detection fails, we miss the calling gesture.

Table 5.2: Performance of Gesture Recognition of Experiment 1

Situation	Total no. of gestures	Accuracy Rate	Number of Correctly Identified Gestures
Sitting one person	50	96%	48
Sitting two people	50	96%	48
Sitting three people	50	86%	43
Sitting four people	50	82%	41
Walking one person	50	96%	48
Walking two people	50	90%	45
Walking three people	50	82%	41
Walking four people	50	78%	39
Sitting and Walking more than four people	50	70%	35

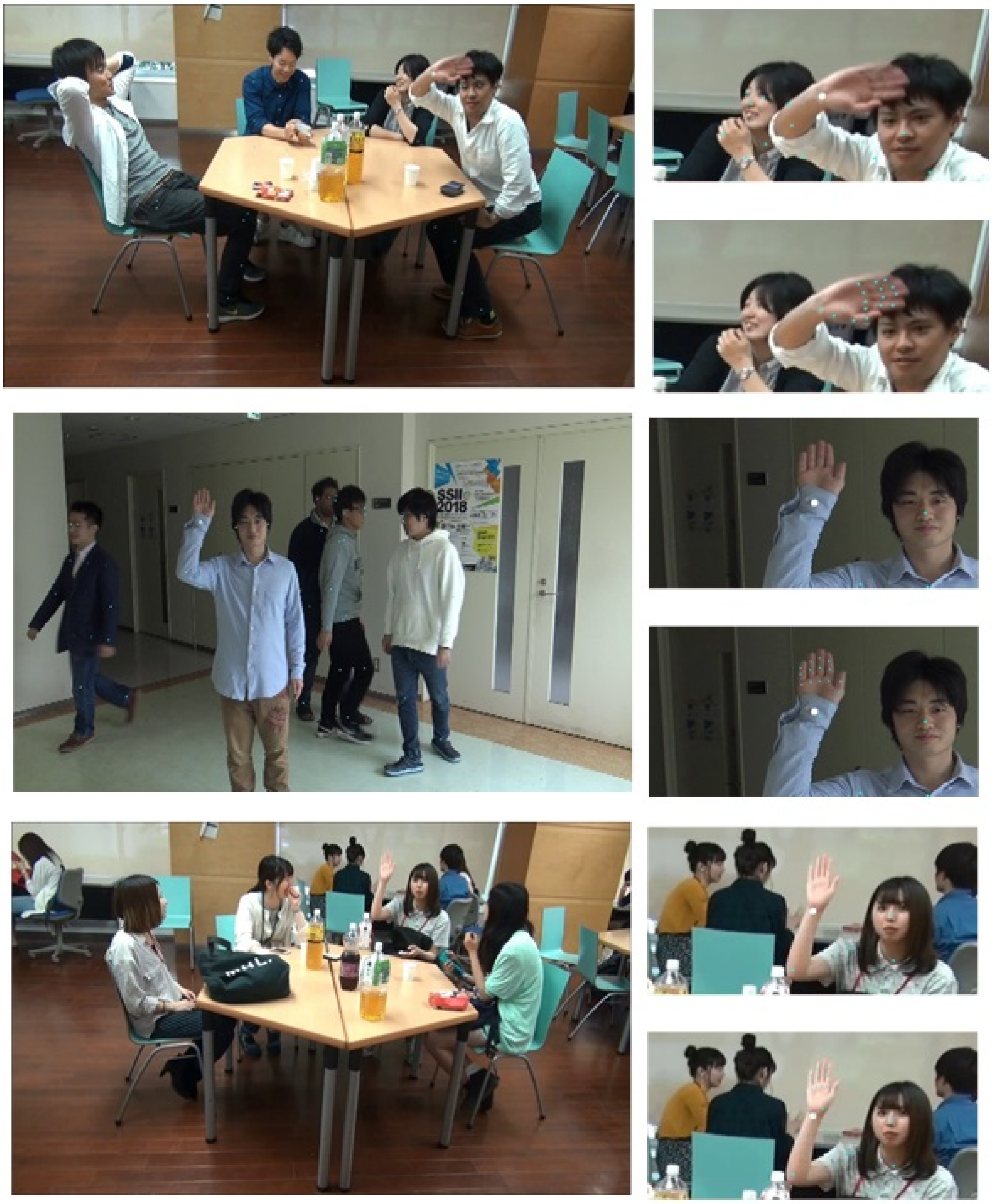


Figure 5.2: Example Results Showing Detected Calling Gestures from Our Algorithm in Three Different Scenes

Table 5.3: Performance of Gesture Recognition of Experiment 2

Situation	Total no. of gestures	Accuracy Rate	Number of Correctly Identified Gestures
Sitting one person	100	92%	92
Sitting two people	100	90%	90
Sitting three people	100	90%	90
Sitting four people	100	87%	87
Walking one person	100	87%	87
Walking two people	100	87%	87
Walking three people	100	80%	80
Walking four people	100	80%	80
Sitting and Walking more than four people	100	75%	75

5.1.2 Experiment 2

The second experiment presents a situation in which the omnidirectional camera is mounted at the top of the mini-cart shown in Figure 5.1(a). We use the recorded video in which the mini-cart is moving around the people shown in Figure 5.1(b). For this second experiment, we captured the videos as in Table 5.3.

Figure 5.3 shows the example results of the accuracy of our algorithm in the different scenes. The performance of the algorithm in Experiment 2 is shown in Table 3. Here, to deal with callers that are behind the robot or would otherwise be out of view with conventional camera, we recorded the video from an omnidirectional camera. We found in the case of the omnidirectional camera, the body of a given person is correctly detected up to within 2m. The performance the system in Experiment 2 is worse than in Experiment 1. This can be explained due to the motion of the camera which moved around the people and resulted in higher gaze detection errors. We also observed that the failure cases with seated people were also mostly due to gaze detection errors.

To analyze the recognition performance, we utilized the metrics Precision and Recall defined as follows:

$$Precision = \frac{True\ Positive}{TruePositive + FalsePositive} \tag{5.1}$$

$$Recall = \frac{True\ Positive}{TruePositive + FalseNegative} \tag{5.2}$$

Figure 5.4 presents the precision-recall comparison for both experiments. Figure 5.5 shows the failure cases of our algorithm. In Figure 5.5(a), although the subject gazes at the robot, performs the calling gesture, and has his hand in the defined position, the system shows he is not calling. The reason for the failure is likely due to the hand is being blurred we could not get the detail fingertips key-points. In Figure 5.5(b), although the subject makes the calling gesture, his gaze is in

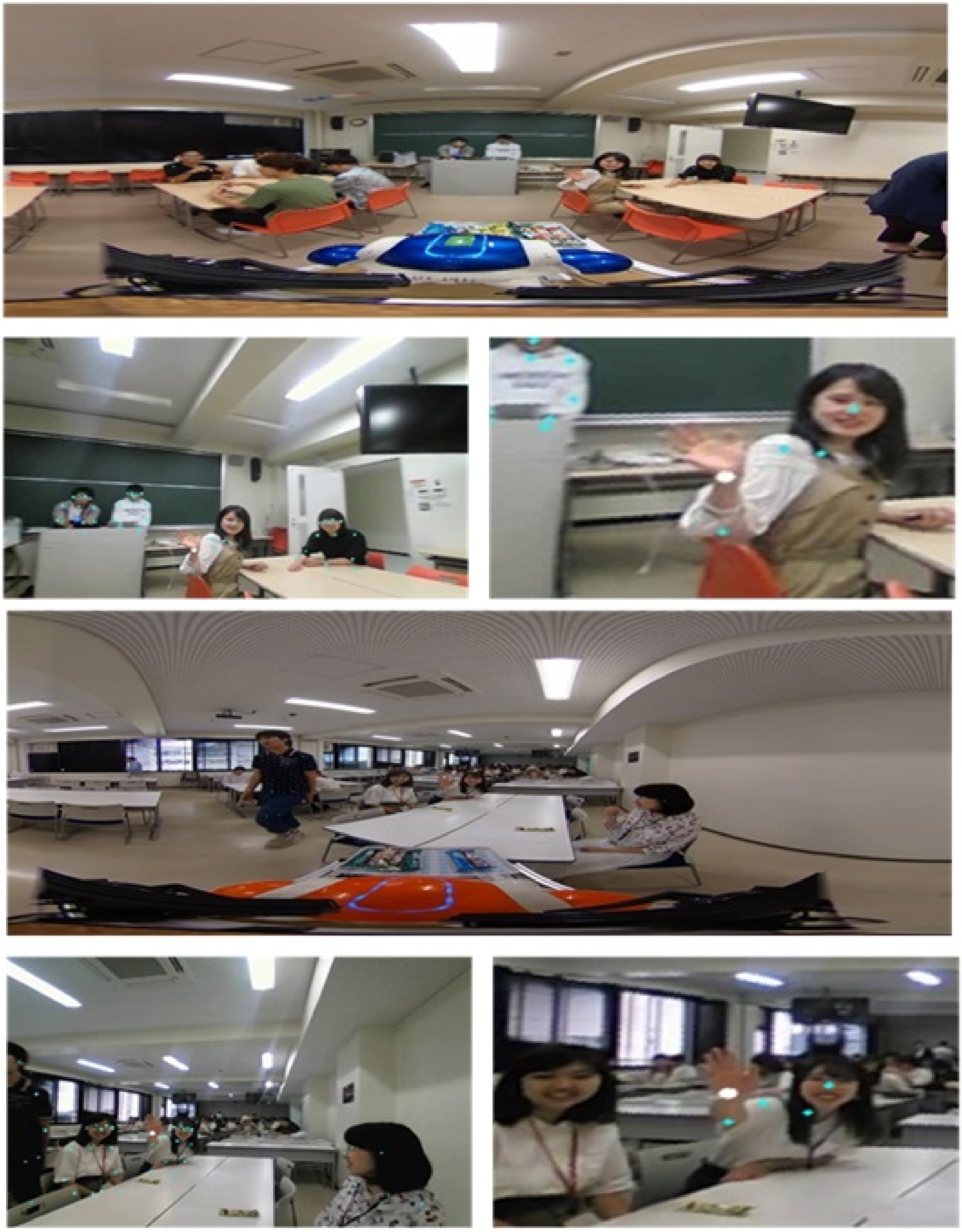


Figure 5.3: Example Results Showing Detected Calling Gestures from Our Algorithm in Three Different Scenes

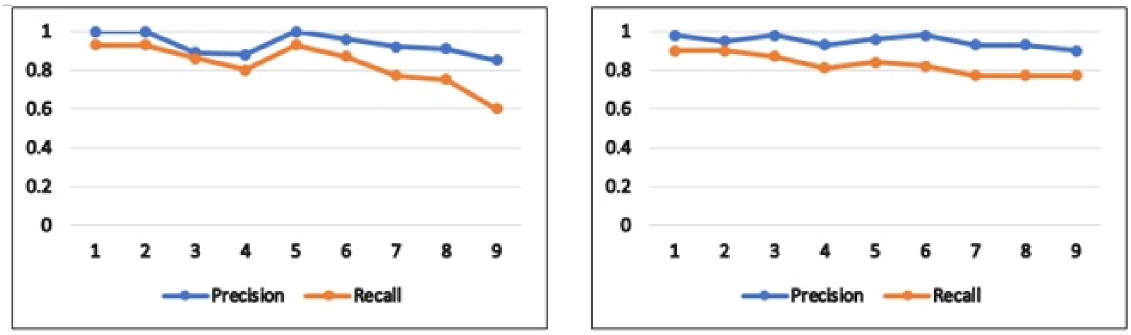


Figure 5.4: Precision-recall Comparison (a) for the experiment 1, (b) for the experiment 2

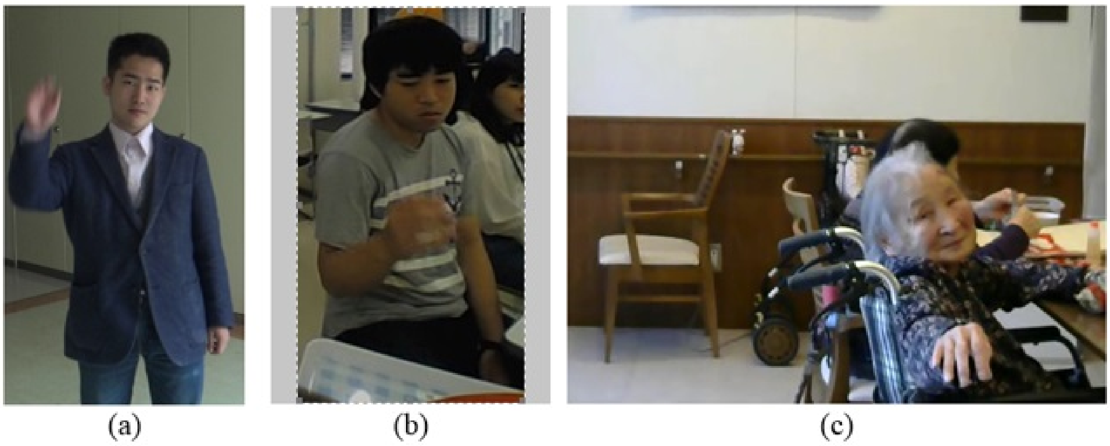


Figure 5.5: Example Failure Cases of Our Algorithm

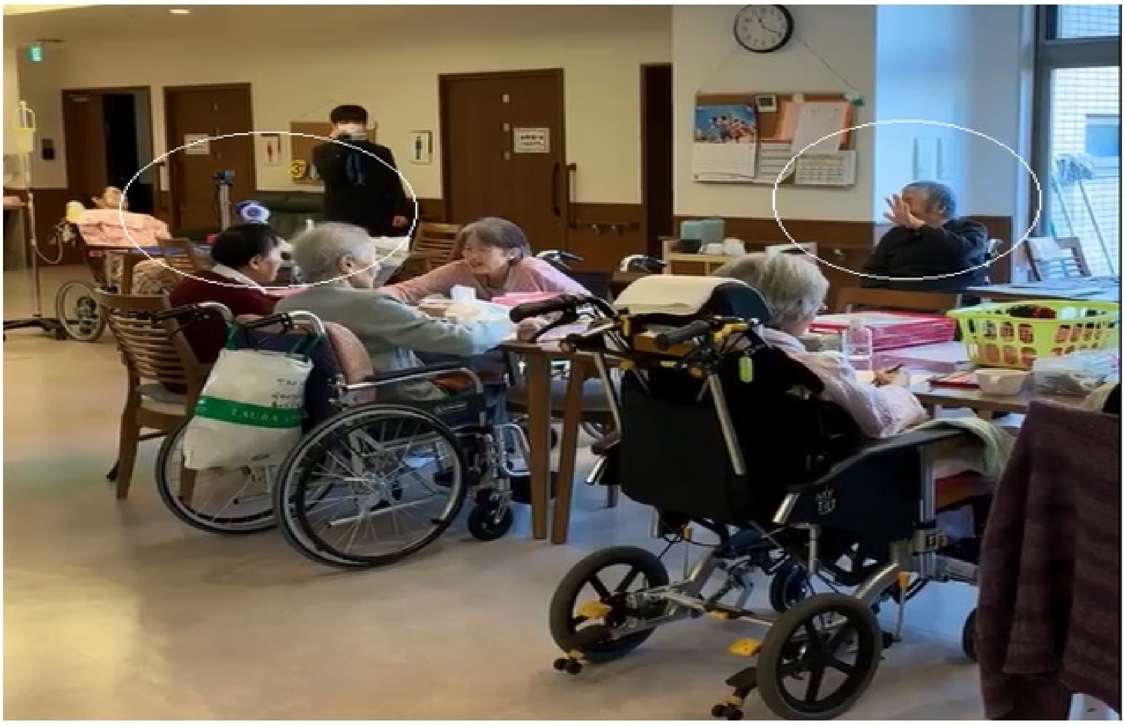


Figure 5.6: Experiment of Our Natural Calling Hand Gesture Recognition algorithm in an Elderly Home Care Center

a different position and the hand is also blurred by shaking. As a result, the system shows the wrong result. As in Figure 5.5(c), the subject gazes at the robot and performs the calling gesture. But, her hand-wrist position is not in one of the defined positions (Figure 3.6). Therefore, the system indicates she is not calling. Figure 5.6 shows an example of how our algorithm can work in a real elderly care center.

5.2 Experiment Analysis of Human-Robot Interaction System

In order to evaluate the performance of human-robot interaction based on gesture recognition, we prepared an indoor test scenario at a real elderly care center in several specific situations and evaluated our system on them. Human test subjects were asked to call the robot by hand. In our approach, firstly, the robot moves around the people as shown in Figure 5.7. At that time, if the person within the camera's field of view calls the robot, the robot moves to that person. If two or more different person within the camera views start to call the robot at a time, the robot will move first detect first move.

Figure 5.8 shows example cases of our experiment. In Figure 5.8(a) and (b), when the robot recognizes the hand calling gesture, the robot moves to that person. The robot keeps moving as the caller continues the hand calling gesture. In Figure 5.8(c) and (d), when the robot detects the calling gesture, the robot approaches the caller. Moving towards the caller, the robot observes whether there is an actual call or not. The robot approaches the caller because of the caller maintaining consistent

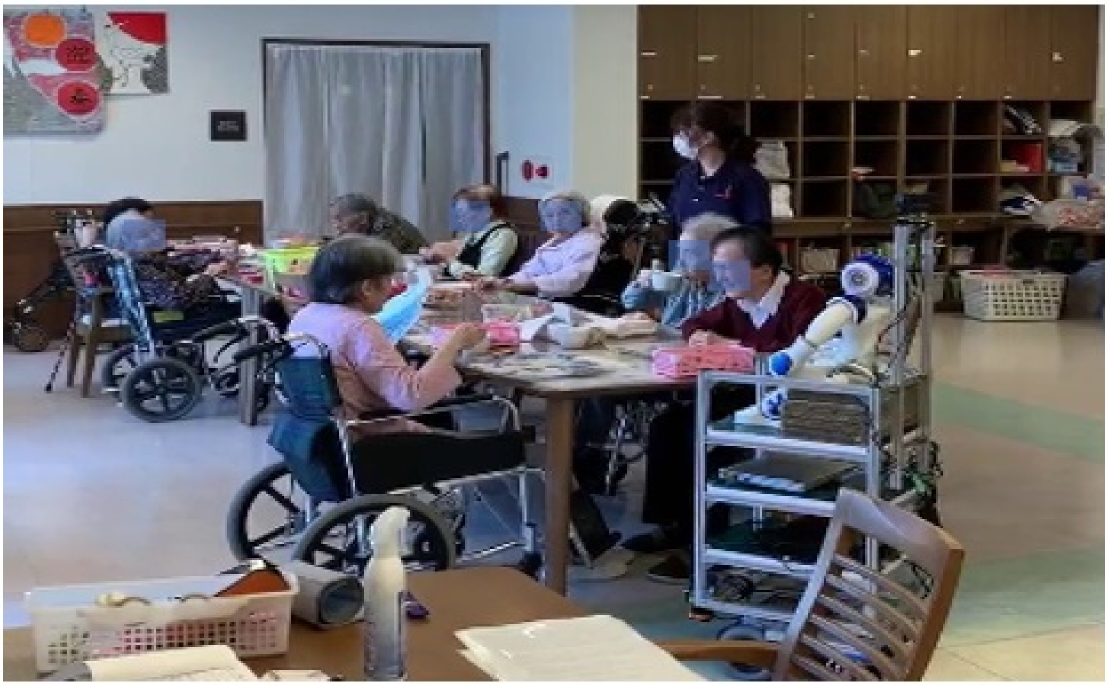


Figure 5.7: The Robotic System Designed for Human-Robot Interaction System in Example Scene

gaze towards the robot. In Figure 5.8(e) and (f), the robot first detects a similar hand calling gesture and the robot moves to that person. But while approaching, the caller changes hand direction and gaze. Therefore, the robot does not keep moving towards that person.

In our experiments, the robot's main computer is a tablet PC with Intel(R) Core (TM) i7-7700HQ CPU @ 2.8G Hz*4, RAM of 32.00 GB, GPU Geforce GTX 1070. In the process of experiments, the system needs a few seconds to execute the behaviors after users signal with hand gestures. If another signal is sent in the same period of time, it takes some time to perform. And, sometimes, gestures are missed as the robot moves around. Based on these cases, the program is set to sleep a short period after one signal is sent. This proves efficient in solving the problem. Another way to over-come our limitation is by using multiple or powerful GPUs to enhance the temporal performance of our system.

5.3 Concluding Remarks

In this chapter discusses the evaluation of the performance of human-robot interaction system based on gesture recognition at a real elderly care center. This chapter gives the experimental analysis of natural hand calling gesture recognition method with two different experiment situations and the analysis of human-robot interaction system.

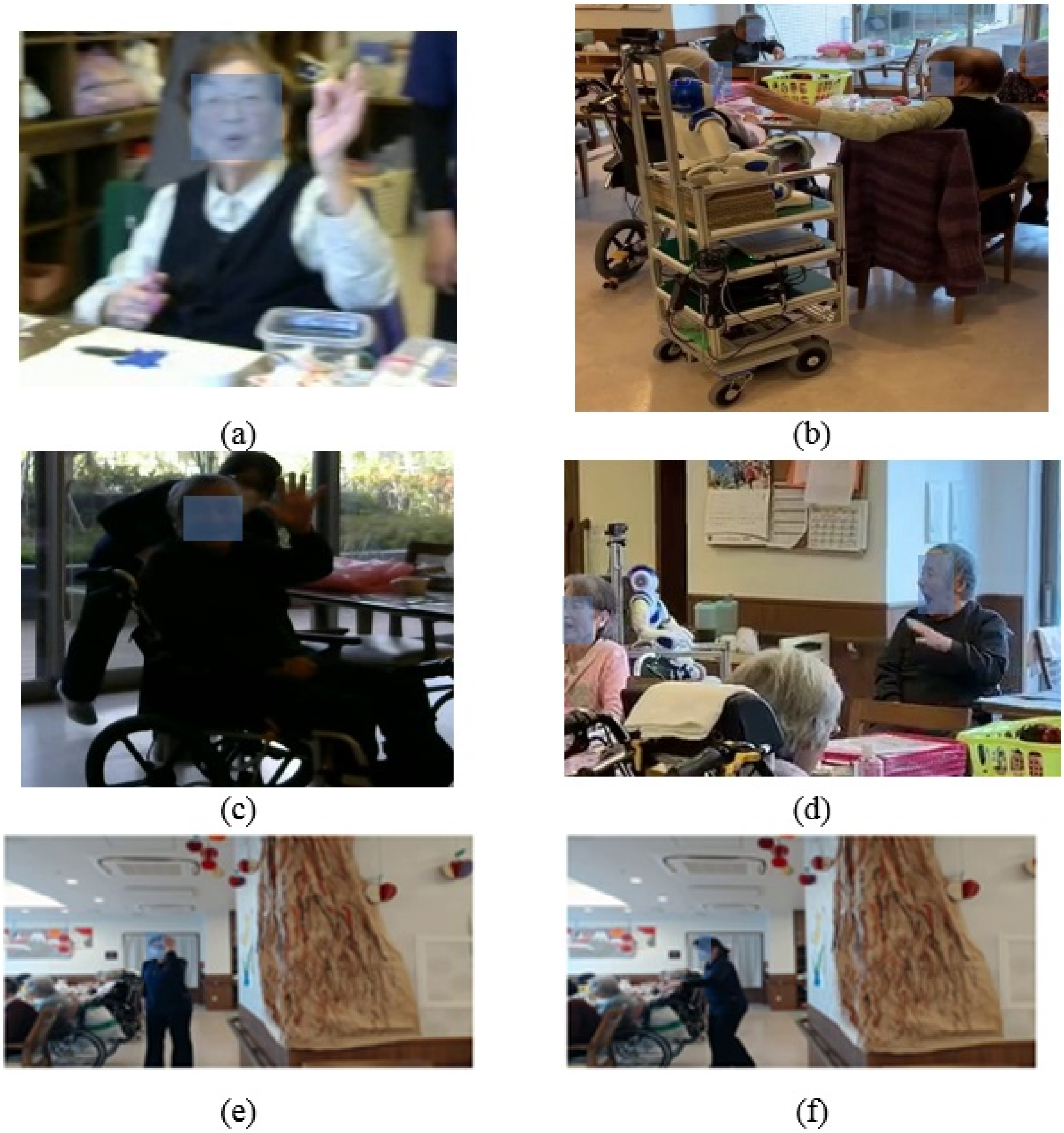


Figure 5.8: Example Results of Our Human-Robot Interaction System

Chapter 6

Conclusions and Future Work

6.1 Conclusions

This thesis proposes a service robot system that provides assisted-care to the elderly. This system recognizes natural calling gestures in an interaction scenario where the robot visually observes the behavior of humans. Therefore, an algorithm for natural calling gesture recognition in crowded environments, for human-robot interaction is introduced. For simple environment, calling using speech is commonly used. However, in a noisy and crowded environment, calling with gesture can be useful in human-robot interaction. To detect users, this study uses the key-points from the OpenPose real-time detector. Using these key-points, gaze detection and finding the hand-wrist positions are performed. If the algorithm finds the gaze and defined hand-wrist position, it zooms into the hand-wrist part. After that, it finds the key-points of the hand's fingertips. From these key-points, this algorithm recognizes whether the user is calling or not by a simple but effective rule-based classification, developed based on basic observations about how people perform calling gestures in real settings. After detecting the calling gesture, the robot moves to the caller. While approaching, the robot observes whether the user is actually calling or not.From this result, the interaction between humans and robot more effective.

We made two experiments to test our natural hand calling gesture recognition algorithm. In the first one, calling gestures can be detected for people that are in front of the camera. For that case, it is possible that if the calling people are behind the camera or in other unseen places, their gestures may be missed. To solve these issues, an omnidirectional camera is used in our second experiment. We tested the proposed approach in video with different conditions from one person to over four people that sit and walk around, and we obtained average classification accuracies of around 86.2% for the first experiment and 85.3% for the second experiment. We validate our findings using our experimental setup, which is composed of a humanoid robot (Aldebaran's NAO) and an i-Cart mini (T-frog) that carries the NAO humanoid and a webcam.

There are also some weakness of our current method now, the calling gesture detection process is not real-time because detecting key-points is still computationally intensive, and the system is affected by gaze detection accuracy. In multiple callers' case at the same time, the robot will move first detect first move.

6.2 Future Work

We will improve the speed of key-point detection and the accuracy of gaze detection in the future. We will extend our work to real-time processing in more complicated situations with a service robot. Furthermore, we will add in the capability to use sound information when it is available in conjunction with vision to enhance human-robot interaction. We expect that many service robot systems will make effective use of the proposed system for human robot interaction.

The use of multiple GPUs for OpenPose can enhance the temporal performance of our system. It can be extended for a more natural and cooperative experience be-tween humans and robots. Multiple object detection, their localization in the scene and handling objects depending on the elderly's actions will be added in future work.

Publications

Journals:

[1] **A.S. Phyo**, H. Fukuda, A. Lam, Y. Kobayashi, Y. Kuno, " Natural Calling Gesture Recognition in Crowded Environment", Book chapter Intelligent Computing Methodologies of the series Springer Lecture Notes in Computer Science, vol. 10954, pp 8-14, Springer, 2018.

[2] **A.S. Phyo**, H. Fukuda, A. Lam, Y. Kobayashi, Y. Kuno, " A Human-Robot Interaction System based on Calling Hand Gestures", International Conference on Intelligent Computing (ICIC 2019) (Accepted and will be published in Springer Lecture Notes in Artificial Intelligence)

Bibliography

[1] S. Lemaignan, M. Warnier, E.A. Sisbot, A. Clodic, R. Alami: Artificial cognition for social human–robot interaction: An Implementation, Artificial Intelligent. 247 , 45–69 (2017)

[2] G. Canal, C. Angulo, and S. Escalera: Gesture based human multirobot interaction. In 2015 International Joint Conference on Neural Networks (IJCNN), 1 - 8, July 2015.

[3] K. G. Engelhardt, R. A. Edwards: "Human robot integration for service robotics," in Human-Robot Interaction, Mansour Rahimi, Waldemar Karwowki. Eds. London: Taylor Francis Ltd., pp. 315–346, (1992)

[4] I. Olaronke, O. Oluwaseun, and I. Rhoda, "State Of The Art: A Study of Human-Robot Interaction in Healthcare", International Journal of Information Engineering and Electronic Business (IJIEEB), vol. 9, num. 3, pp. 43–55, (2017)

[5] A. Singh, J. Buonassisi, and S. Jain, " Autonomous Multiple Gesture Recognition System for Disabled People", International Journal of Image, Graphics and Signal Processing (IJIGSP), vol. 6, num. 2, pp. 39–45, (2014)

[6] G. Canal, C. Angulo, and S. Escalera: (2016). A Real-Time Human-Robot Interaction System based on Gestures for Assistive Scenarios. Computer Vision and Image Understanding, 149: 65–77, (2016). doi:10.1016/j.cviu.2016.03.004

[7] Z. Zafar and K. Berns: "Recognizing Hand Gestures for Human-Robot Interaction," 9th International Conference on Advances in Computer-Human Interactions, pp. 333–338, (2016)

[8] W. Chen (2013): Gesture-Based Applications for Elderly People. In: Kurosu M. (eds) Human-Computer Interaction. Interaction Modalities and Techniques. HCI 2013. LNCS, vol 8007. Springer, Berlin, Heidelberg.

[9] T. Sharp, C. Keskin, D. Robertson, J. Taylor, J. Shotton, D. Kim, C. Rhemann, I. Leichter, A. Vinnikov, Y. Wei, D. Freedman, P. Kohli, E. Krupka, and S. Fitzgibbon, A. amd Izadi. Accurate, robust, and flexible real-time hand tracking. In Proceedings of the ACM CHI '15, April 2015.

[10] C. Breazeal, C. Kidd, A. Thomaz, G. Hoffman, M. Berlin Effects of nonverbal communication on efficiency and robustness in human-robot teamwork Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 708-713,(2005). doi:10.1109/IROS.2005.1545011

[11] S. S. Rautaray, A. Agrawal: "Vision based hand gesture recognition for human computer interaction: a survey", Artif Intelligent Review, 1-54, (2015). doi:10.1007/s10462-012-9356-9

[12] A. Choudhury, A. K. Talukdar, K. K. Sarma : A Review on Vision based Hand Gesture Recognition and Applications. In: Intelligent Applications for Heterogeneous System Modeling and Design. IGI Global, pp 256–281, (2015). doi:10.4018/978-1-4666-8493-5.ch011

[13] International Federation of Robotics (2015a) Definition of Service Robots. http://www.ifr.org/service-robots/ . Accessed 17 Feb 2015

[14] M. Sprenger , T. Mettler : Service robots. Business Information Systems Engineering 57(4):271–274 (2015). doi:10.1007/s12599-015-0389-x

[15] T. Haidegger, M. Barreto, P. Goncalves , M. K. Habib, S. K.V. Ragavan , H. Li, et al: Applied Ontologies and Standards for Service Robots. Robotics and Autonomous Systems 61(11):1215-1223 (2013). doi:10.1016/j.robot.2013.05.008

[16] N. F. Garmann-Johnsen ,T. Mettler , M. Sprenger : Service Robotics in Healthcare: A Perspective for Information Systems Researchers? Thirty Fifth International Conference on Information Systems (ICIS), Auckland (2014).doi:10.13140/2.1.4973.9203

[17] A. Sharkey, N. Sharkey Granny and the Robots: Ethical Issues in Robot Care for the Elderly. Ethics and Information Technology 14(1):27–40 (2012) . doi:10.1007/s10676-010-9234-6

[18] M. E. Pollack, S. Engberg, J. T. Matthews, S. Thrun, L. Brown, D. Colbry, C. Orosz, B. Peintner, S. Ramakrishnan, and J. Dunbar-Jacob, Pearl: A Mobile Robotic Assistant for the Elderly. AAAI Workshop on Automation as Eldercare, (2002)

[19] B. Graf, M. Hans, and R.D. Schraft, Care-O-bot II—Development of a NextGeneration Robotic Home Assistant. Autonomous Robots. 16(2): p. 193-205 (2014)

[20] S. Bahadori, A. Cesta, G. Grisetti, L. Iocchi, R. Leone, D. Nardi, A. Oddi, F. Pecora, and R. Rasconi, RoboCare: an Integrated Robotic System for the Domestic Care of the Elderly. Proceedings of workshop on Ambient Intelligence AI* IA-03, Pisa, Italy, (2003)

[21] K. Wada, T. Shibata, T. Saito, and K. Tanie, Effects of robot assisted activity to elderly people who stay at a health service facility for the aged, (IROS 2003). Proceedings. IEEE/RSJ International Conference on, 2003a. (2003)

[22] W. D. Stiehl, J. Lieberman, C. Breazeal, L. Basel, R. Cooper, H. Knight, L. Lalla, A. Maymin, and S. Purchase, The Huggable: a Therapeutic Robotic Companion for Relational, Affective Touch. Consumer Communications and Networking Conference, 3rd IEEE, (2006)

[23] M. Fujita, AIBO: Toward the Era of Digital Creatures. The International Journal of Robotics Research, 20(10): p. 781, (2001)

[24] K. Dautenhahn : Socially intelligent robots: dimensions of human-robot interaction. Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 362(1480), 679–704, (2007). doi:10.1098/rstb.2006.2004

[25] G. S. Yumakulov, Social Robots: Views of Staff of a Disability Service Organization. International Journal of Social Robotics, 2014, 6(3): pp.457-468

[26] Yan, H., et al. A Survey on Perception Methods for Human–Robot Interaction in Social Robots. International Journal of Social Robotics, 2013, 6(1): pp.85-119

[27] C. Grazia, Carmela Attolico, Cataldo Guaragnella, and Tiziana D'Orazio. A kinect-based gesture recognition approach for a natural human robot interface. International Journal of Advanced Robotic Systems, 12(3):22, 2015

[28] L. R. Jagdish, Mona Chandra, and Ankit Chaudhary. 3d gesture based real-time object selection and recognition. Pattern Recognition Letters, 2017.

[29] X. Pei . A real-time hand gesture recognition and human-computer interaction system. CoRR, abs/1704.07296, 2017

[30] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11):2278–2324, Nov 1998.

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

[32] J. S. Sonkusare, N. B. Chopade, R. Sor, and S. L. Tade. A review on hand gesture recognition system. In 2015 International Conference on Computing Communication Control and Automation, pages 790–794, Feb 2015. doi:doi:10.1109/ICCUBEA.2015.158.

[33] Sanbot. Sanbot s1.http://en.sanbot.com/newsPro/design.html, retrieved May 2017

[34] António Neves Tiago Esteves Patrick de Sousa, Luís Texeira. Human-robot interaction based on gestures for service robots. VipIMAGE, 2017.

[35] I. A. Noor, K. RafiqulZaman: "Survey on Various Gesture Recognition Technologies and Techniques", International Journal of Computer Applications, 0975–8887, July 2012 Volume 50–7.

[36] P. Orasa, N. Chakarida, W. Bunthit : "Human Gesture Recognitions Using Kinetic Camera", 2012, 9th International Join Conference on Computer Science and Software Engineering.

[37] K. Harpreet, R. Jyoti: "A Review: Study of Various Techniques of Hand Gesture Recognition", IEEE International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES), 2016

[38] J. Joseph, J. LaViola : "A Survey of Hand Posture and Gesture Recognition Techniques and Technology", Master Thesis, Brown University, NSF Science and Technology Center for Computer Graphics and Scientific Visualization, USA

[39] M. H. Mokhtar, K. M. Pramod, "Hand Gesture Modeling and Recognition using Geometric Features: A Review", Canadian Journal on Image Processing and Computer Vision,2012, Vol3-1.

[40] L. Luigi, C. Francesco: "RealTime Hand Gesture Recognition Using a Color Glove", Springer 16th international conference on Image analysis and processing: Part I , 2011, pages 365-373

[41] Y. Huang, D. Monekosso, H. Wang, and J. Augusto, "A concept grounding approach for glove-based gesture recognition," in Intelligent Environments (IE), 7th International Conference on, July 2011, pp. 358–361

[42] I. Skrypnyk and D. Lowe, "Scene modelling, recognition and tracking with invariant image features," in Mixed and Augmented Reality, Third IEEE and ACM International Symposium on, Nov 2004, pp. 110–119.

[43] R. C. Luo and Y. C. Wu. Hand gesture recognition for human-robot interaction for service robot. IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems, pages 318–323, Sept 2012

[44] M. Van den Bergh, D. Carton, R. De Nijs, N. Mitsou, C. Landsiedel, K. Kuehnlenz, D. Wollherr, L. Van Gool, and M. Buss. Real-time 3d hand gesture interaction with a robot for understanding directions from humans. RO-MAN, pages 357–362, July 2011

[45] Y. Yang, H. Yan, M. Dehghan, and M. H. Ang. Real-time human-robot interaction in complex environment using kinect v2 image recognition. IEEE 7th International Conference on Cybernetics and Intelligent Systems and IEEE Conference on Robotics, Automation and Mechatronics, pages 112–117, July 2015.

[46] X. Zhao, Naguib, A. M., Lee, S: "Kinect based calling gesture recognition for taking order service of elderly care robot", The 23rd IEEE International Symposium on Robot and Human Interactive Communication 2014 RO-MAN, pp. 525- 530, Aug 2014

[47] "New Humanoid Robot tu kaiserslautern," http://agrosy.informatik.unikl.de/aktuelles/details/news/neuerhumanoiderroboter.html, accessed: 2016-01-25.

[48] E. Marilly, A. Gonguet, O. Martinot, and F. Pain, "Gesture interactions with video: From algorithms to user evaluation," Bell Labs Technical Journal, vol. 17, no. 4, March 2013, pp. 103–118.

[49] D.-Y. Huang, W.-C. Hu, and S.-H. Chang, "Vision-based hand gesture recognition using pca+gabor filters and svm," in Intelligent Information Hiding and Multimedia Signal Processing, Fifth International Conference on, September 2009, pp. 1–4.

[50] A. L. Barczak and F. Dadgostar, "Real-time hand tracking using a set of cooperative classifiers based on haar-like features," in Research Letters in the Information and Mathematical Sciences, Vol. 7. Massey University, 2005, pp. 29–42.

[51] A. A. Argyros and M. I. A. Lourakis, "Vision-based interpretation of hand gestures for remote control of a computer mouse," in In Computer Vision in Human-Computer Interaction. Springer-Verlag, 2006, pp. 40–51.

[52] I. Skrypnyk and D. Lowe, "Scene modelling, recognition and tracking with invariant image features," in Mixed and Augmented Reality, 2004. ISMAR 2004. Third IEEE and ACM International Symposium on, Nov 2004, pp. 110–119.

[53] R. Hartanto, A. Susanto, and P. Santosa, "Preliminary design of static indonesian sign language recognition system," in Information Technology and Electrical Engineering (ICITEE), 2013 International Conference on, Oct 2013, pp. 187–192.

[54] N. Dardas and N. D. Georganas, "Real-time hand gesture detection and recognition using bag-of-features and support vector machine techniques," Instrumentation and Measurement, IEEE Transactions on, vol. 60, no. 11, Nov 2011, pp. 3592–3607.

[55] G. Amit , K. S. Vijay , K. Mamta : FPGA based real time human hand gesture recognition system. Procedia Technology, 6:98–107, 2012.

[56] A. Tasnuva: A neural network based real time hand gesture recognition system. International Journal of Computer Applications, 59(4):17–22, 2012.

[57] L. Raymond, Andrew W. Fitzgibbon. Real-time gesture recognition using deterministic boosting. In BMVC, volume 2002, pages 1–10, 2002.

[58] D. Alex : Kinect hand recognition and tracking. Washington University in St. Louis, Kinect Hand Recognition and Tracking, Project Report, 2012.

[59] D. James, Mubarak Shah. Visual gesture recognition. In Vision, Image and Signal Processing, IEE Proceedings, volume 141, pages 101–106. IET, 1994.

[60] R. Zhou , Jingjing Meng, and Junsong Yuan. Depth camera based hand gesture recognition and its applications in human-computer-interaction. In 2011 8th International Conference on Information, Communications and Signal Processing (ICICS), pages 1–5, Dec 2011.

[61] C. Sait, Ali S.A.S. Aydin, T.T. Talha T. Temiz, and Tarik Arici. Gesture recognition using skeleton data with weighted dynamic time warping. Computer Vision Theory and and Applications. VisApp, 2013.

[62] J. Suarez and R.R. Murphy. Hand gesture recognition with depth images: A review. In RO-MAN, 2012 IEEE, pages 411–417, Sept 2012.

[63] K. Barry , Sven Kratz, and Anthony Dunnigan. Exploring gestural interaction in smart spaces using head mounted devices with ego-centric sensing. In Proceedings of the 2nd ACM Symposium on Spatial User Interaction, SUI'14, pages 40–49, New York, NY, USA, 2014. ACM.

[64] K. Sven, and M.D.T. I. Aumi. AirAuth: A Biometric Authentication System Using In-air Hand Gestures. In CHI '14 Extended Abstracts on Human Factors in Computing Systems, CHI EA '14, pages 499–502, New York, NY, USA, 2014. ACM.

[65] M. Tomás Mantecón, Carlos R. del Blanco, Fernando Jaureguizar, and Narciso García. New generation of human machine interfaces for controlling UAV through depth-based gesture recognition. In SPIE Defense+ Security, pages 90840C–90840C. International Society for Optics and Photonics, 2014.

[66] K. Ondrej ,František Jakab. Approach to hand tracking and gesture recognition based on depth-sensing cameras and EMG monitoring. Acta Informatica Pragensia, 3(1):104–112, 2014.

[67] W. Frank, Daniel Bachmann, Bartholomäus Rudak, and Denis Fisseler. Analysis of the accuracy and robustness of the Leap motion controller. Sensors, 13(5):6380–6393, 2013.

[68] F.-S. Chen, C.-M. Fu, and C.-L. Huang, "Hand gesture recognition using a real-time tracking method and hidden Markov models," Image and Vision Computing, vol. 21, no. 8, pp. 745–758, 2003.

[69] T. Fujii, J. Hoon Lee, and S. Okamoto. Gesture recognition system for human-robot interaction and its application to robotic service task. In Proceedings of The International MultiConference of Engineers and Computer Scientists (IMECS 2014), volume I, pages 63–68. International Association of Engineers, Newswood Limited, 2014.

[70] P. Barros, G. I. Parisi, D. Jirak, and S. Wermter. Real-time gesture recognition using a humanoid robot with a deepneural architecture. In Proceedings of the IEEE-RAS International Conference on Humanoid Robots (Humanoids '14), pages 83–88. IEEE, 2014.

[71] M. W. Krueger: Artificial Reality II, vol. 10. Addison-Wesley, Reading (1991)

[72] G. Dong, Yan, Y., Xie, M.: Vision-based hand gesture recognition for human–vehicle interaction. Paper Presented at the Proceedings of the International conference on Control, Automation and Computer Vision (1998)

[73] V. D. Shet, Shiv, V., Prasad, N., Elgammal, A., Yacoob, Y., Davis, L.S.: Multi-cue exemplar-based nonparametric model for gesture recognition. Paper presented at the ICVGIP (2004)

[74] B. Ionescu, Coquin, D., Lambert, P., Buzuloiu, V.: Dynamic hand gesture recognition using the skeleton of the hand. EURASIP J. Appl. Signal Process. 2005, 2101–2109 (2005)

[75] F. Parvini, Shahabi, C.: An algorithmic approach for static and dynamic gesture recognition utilising mechanical and biomechanical characteristics. Int. J. Bioinform. Res. Appl. 3(1), 4–23 (2007)

[76] V. I. Pavlovic, Sharma, R., Huang, T.S.: Visual interpretation of hand gestures for human–computer interaction: a review. IEEE Trans. Pattern Anal. Mach. Intell. 19(7), 677–695 (1997). doi:10.1109/34.598226

[77] B. Swapna, Pravin, F., Rajiv, V.D.: Hand gesture recognition system for numbers using thresholding. Comput. Intell. Inf. Technol. 250, 782–786 (2011)

[78] H. Il Suk, B. K. Sin, and S. W. Lee, "Hand gesture recognition based on dynamic Bayesian network framework," Pattern Recognition, vol. 43, no. 9, pp. 3059–3072, 2010

[79] W. T. Freeman, Roth, M.: Orientation histograms for hand gesture recognition. Paper Presented at the International Workshop on Automatic Face and Gesture Recognition (1995)

[80] M. W. Krueger: Artificial Reality II, vol. 10. Addison-Wesley, Reading (1991)

[81] J. Triesch, von der Malsburg, C.: Robust classification of hand postures against complex backgrounds. Paper Presented at the Automatic Face and Gesture Recognition, 1996. Proceedings of the Second International Conference on (1996)

[82] A. Licsr, Szirnyi, T.: Hand-gesture based film restoration. Paper presented at the PRIS (2002)

[83] C. C. Chang, Chen, J.J., Tai, W.K., Han, C.C.: New approach for static gesture recognition. J. Inf. Sci. Eng. 22(5), 1047–1057 (2006)

[84] D. Y. Huang, Hu, W.C., Chang, S.H.: Vision-based hand gesture recognition using PCA+gabor filters and SVM. Paper Presented at the Intelligent Information Hiding and Multimedia Signal Processing, 2009. IIH-MSP'09. Fifth International Conference on (2009)

[85] H. Meng, Furao, S., Jinxi, Z.: Hidden Markov models based dynamic hand gesture recognition with incremental learning method. Paper Presented at the Neural Networks (IJCNN), 2014 International Joint Conference on (2014)

[86] H. Chao, Meng, M.Q., Liu, P.X., Xiang, W.: Visual gesture recognition for human-machine interface of robot teleoperation. Paper Presented at the Intelligent Robots and Systems, 2003. (IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on (2003)

[87] Y. Liu, Zhang, P.: An Automatic hand gesture recognition system based on Viola–Jones method and SVMs. Paper Presented at the Computer Science and Engineering, 2009. WCSE'09. Second International Workshop on (2009)

[88] C. Bekir : Hand Gesture Recognition. (Master thesis), Dokuz Eyll

[89] R. Zhou, Junsong, Y., Jingjing, M., Zhengyou, Z.: Robust partbased hand gesture recognition using kinect sensor. IEEE Trans. Multimed. 15(5), 1110–1120 (2013). doi:10.1109/TMM.2013.2246148

[90] R. K. McConnell: US Patent, No. 4567610 (1986) University (2012)

[91] W. T. Freeman, Roth, M.: Orientation histograms for hand gesture recognition. Paper Presented at the International Workshop on Automatic Face and Gesture Recognition (1995)

[92] J. J. . Triesch, C. Von Der Malsburg, and C. von der Malsburg, "A system for person independent hand posture recognition against complex backgrounds," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 12, pp. 1449–1453, 2001

[93] S. Oprisescu, Rasche, C., Bochao, S.: Automatic static hand gesture recognition using ToF cameras. Paper Presented at the Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European (2012)

[94] C. C. Chang, Chen, J.J., Tai, W.K., Han, C.C.: New approach for static gesture recognition. J. Inf. Sci. Eng. 22(5), 1047–1057 (2006)

[95] V. Athitsos, Sclaroff, S.: Boosting nearest neighbor classifiers for multiclass recognition. Paper Presented at the Computer Vision and Pattern Recognition-Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on (2005)

[96] H. Liu and L. Wang, "Gesture recognition for human-robot collaboration: A review," International Journal of Industrial Ergonomics, pp. 1–13, 2016.

[97] E. Bernier, R. Chellali, and I. M. Thouvenin. Human gesture segmentation based on change point model for efficient gesture interface. In Proceedings of the 2013 IEEE RO-MAN, pages 258–263, Aug 2013.

[98] X. Zhao, A. M. Naguib, and S. Lee. Kinect based calling gesture recognition for taking order service of elderly care robot. In The 23rd IEEE International Symposium on Robot and Human Interactive Communication, 2014 RO-MAN, pages 525–530, Aug 2014

[99] M. Sigalas, H. Baltzakis, and Trahanias. P. Temporal gesture recognition for human-robot interaction. In Proceedings of Mutimodal HumanRobot Interfaces Workshop. IEEE International Conference on Robotics and Automation (ICRA), Anchorage, Alaska, USA, May 2010

[100] D. Xu, X. Wu, Y. Chen, and Y. Xu. Online dynamic gesture recognition for human robot interaction. Journal of Intelligent Robotic Systems, pages 1–14, 2014

[101] S. Iengo, S. Rossi, M. Staffa, and A. Finzi. Continuous gesture recognition for flexible human-robot interaction. In Proceedings of the 2014 IEEE International Conference on Robotics and Automation, ICRA 2014, pages 4863–4868, 2014

[102] V. Alvarez-Santos, Iglesias. R., X. M. Pardo, C. V. Regueiro, and A. Canedo-Rodriguez. Gesture-based interaction with voice feedback for a tour-guide robot. Journal of Visual Communication and Image Representation, 25(2):499 – 509, 2014

[103] K. O'Brien, J. Sutherland, C. Rich, and C. L. Sidner. Collaboration with an autonomous humanoid robot: A little gesture goes a long way. In 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pages 215–216, March 2011

[104] G. Canal, S. Escalera, and C. Angulo. A real-time human-robot interaction system based on gestures for assistive scenarios. Computer Vision and Image Understanding, 149:65–77, 2016

[105] K. Ehlers and K. Brama. A human-robot interaction interface for mobile and stationary robots based on real-time 3d human body and hand-finger pose estimation. IEEE 21st International Conference on Emerging Technologies and Factory Automation (ETFA), pages 1–6, Sept 2016

[106] G. Canal, Angulo, C., Escalera, S., (2015). Gesture based human multirobot interaction. In 2015 International Joint Conference on Neural Networks (IJCNN), pages 1 –8, July 2015

[107] C.G.T.D.C. Grazia, C. Attolico, A kinect-based gesture recognition approach for a natural human robot interface. International Journal of Advanced Robotic Systems, 12, March 19, 2015.

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

[109] X. Duan, H. Liu, "Detection of hands-raising gestures based on body silhouette analysis," in Proceedings of IEEE International Conference on Robotics and Biomimetics, 2008, pp. 1756–1761

[110] X. Duan, , Liu, H., Zou, Y., Gao, D., (2009). "Detection of handsraising gestures using shape and edge features," in Proceedings of the 2009 IEEE International Conference on Robotics and Biomimetics, 2009, pp. 1480–1483

[111] Q. D. Smedt, Wannous, H., Vandeborre, J.-P., Guerry, J., Saux, B. L., Filliat, D., (2017). "Shrec'17 track: 3d hand gesture recognition using a depth and skeletal dataset", 10th Eurographics Workshop on 3D Object Retrieval, 2017.

[112] D. Miyauchi, Sakurai, A., Nakamurai, A., Kuno, Y., (2005). Bidirectional eye contact for human-robot communication. IEICE Transactions on Information and Systems E88-D, 11, 2509-2516.

[113] K. Yamazaki, Kawashima, M., Kuno, Y., Akyiya, N., Burdelski, M., Yamazaki, A., Kusuoka. H., Prior-to-request and request behaviors within elderly day care: Implications for developing service robots for use in multiparty settings, European Conference on Computer-Supported Cooperative Work (ECSCW2007), pp. 661 -78

[114] W. Quan, Niwa, H., Ishikawa, N., Kobayashi, Y., Kuno, Y., (2009). "Assisted-care robot based on sociological interaction analysis", Proc. ICIC2009, 2009.

[115] C. Y. Kao, C.S. Fahn: A human-machine interaction technique: hand gesture recognition based on hidden Markov models with trajectory of hand motion Procedia Eng., 15 (2011), pp. 3739-3743

[116] Z. C. Hidalgo, T. Simon, S.-E. Wei, H.J., Sheik., Y., Openpose. https://github.com/CMU-Perceptual-ComputingLab/openpose.