An Intelligent Person Following Shopping Support Robot for the Elderly

(高齢者のための知的追従する買い物支援ロボット)

Saitama University
埼玉大学

Md Matiqul Islam

Graduate School of Science and Engineering

Saitama University, Japan

Supervisor: Professor Yoshinori Kobayashi

A thesis submitted for the degree of

Doctor of Philosophy

September, 2020

This thesis is dedicated to
Md. Tafhim Mahamid, Anika Farha,
Alhaz Md. Mokbul Hossain Sarkar, and *Matina Begum*
for their endless love and support

Acknowledgements

First of all, I thank ***ALLAH*** for giving me the strength and ability to complete this study. There are also many people who have supported and encouraged me through out this doctoral research that I would like to acknowledge.

First and foremost, I would like to thank my supervisor, ***Professor Yoshinori Kobayash*** and my ex supervisor ***Professor Yoshinori Kuno***, Saitama University, Japan, for their trust, continuous support, guidance, patience over the past few years. They have provided me the opportunity to work in the field of targeted person-following intelligent shopping support robot. I feel incredibly lucky and honored to be one of their doctoral students. I am grateful to them for their constant encouragement, and their stimulating ideas.

I am incredibly grateful to my thesis examiners, ***Professor Yoshinori Kobayashi, Professor Tetsuya Shimamura and Professor Takashi Komuro*** for their comments and insightful discussion about my works.

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

I owe a huge debt to my friends ***Kamal Unddin*** and ***Mahmudul Hasan***, Doctoral student of Saitama University, Japan, for their unconditional supports and cooperations on my doctoral study.

The generous financial support from JSPS KAKENHI (Grant number JP26240038), Japan Science and Technology Agency (JST) has made it possible to pursue my research.

Finally, but very certainly not least, I am grateful to my son ***Md Tafhim Mahamid*** , and my wife, ***Anika Farha*** who had to bear my moods and impatience when the research did not work out the way I wanted. Without their love, care, support and existence, I would not be able to complete my doctoral study in Japan. I thank my parents and parents-in-law for their continuing DOA throughout all the phases of my life that have led to this point. I am lucky to have you all, and this dissertation could not have happened without you.

Abstract

The lack of caregivers in an aging society is a major social problem. Without assistance, many of the elderly and disabled are unable to perform daily tasks. One important daily activity is shopping in supermarkets. Carrying heavy weighted goods or pushing a shopping cart and moving it from shelf to shelf is tiring and laborious, especially for customers with certain disabilities or the elderly.

Many researcher develop person following robot for supporting elderly in shopping mall but just following may not sufficient or good enough for supporting elderly. Considering the body orientation of the customer it is better to find the appropriate positional relation between the customer and the robot.In addition to the robust person-following, the robot can more support the user if it can act in advance to meet the user's next move. For example, when the user picks up a product from a shelf, it is convenient if the robot automatically comes to the user's right hand side (if the user is right-handed) so that he or she can put it easily in the basket. To realize such functions, the robot needs to recognize the user's behavior.

The first part of the work is on developing body orientation based shopping support robot. To do that, we address the problem of real-time human pose-based robust person tracking system. We achieve this by cropping the target person's body from the image and then apply a color histogram matching algorithm for tracking a unique person. After tracking the person, we used an omnidirectional camera and LiDAR sensor to find the target person's location and distance from the robot. When the target person stop in front of shopping shelves our robot finds the target person's body movement orientation using our proposed methodology. According to the body orientation our robot assumes a suitable position so that the target person can easily put his shopping product in the basket. Our proposed system was verified in real time environments and it

shows that our robot system is highly effective at following a given target person and provides proper support while shopping the target person.

The next step was to develop an intelligent shopping support robot that can carry a shopping cart while following its owners and provide the shopping support by observing the customer's head orientation, body orientation and recognizing different shopping behaviors. Recognizing shopping behavior or the intensity of such action is important for understanding the best way to support the customer without disturbing him or her. This system also liberates elderly and disabled people from the burden of pushing shopping carts, because our proposed shopping cart is essentially a type of autonomous mobile robots that recognizes its owner and following him or her. The proposed system discretizes the head and body orientation of customer into 8 directions to estimate whether the customer is looking or turning towards a merchandise shelf. From the robot's video stream, a DNN-based human pose estimator called OpenPose is used to extract the skeleton of 18 joints for each detected body. Using this extracted body joints information, we built a dataset and developed a novel Gated Recurrent Neural Network (GRU) topology to classify different actions that are typically performed in front of shelves: reach to shelf, retract from shelf, hand in shelf, inspect product, inspect shelf. Our GRU network model takes series of 32 frames skeleton data then gives the prediction. Using cross-validation tests, our model achieves an overall accuracy of 82%, which is a significant result. Finally, from the customer's head orientation, body orientation and shopping behavior recognition we develop a complete system for our shopping support robot.

To operate our robot in a practical environment we must ensure three requirements such as speed, accuracy and cost. OpenPose based model does not ful-fill these requirements. For this reason, we replace OpenPose model with Kinect V2 depth camera. Kinect camera can detect 3D skeleton with approximately 30 frames/sec whereas OpenPose model detect skeleton approximately 5 frames/sec. The accuracy of 3D skeleton based shopping action recognition is high and it is 95% using our GRU network. Using Kinect camera we can measure the distance from robot to tracked person so extra LiDAR sensor does not need. So, Kinect based model is cost effective and does not need extra processing.

Finally, we develop a person following shopping support robot using a Kinect camera that can recognize customer shopping actions or activities. Our robot can follow within a certain distance behind the customer. Whenever our robot detects the customer performing a “hand in shelf” action in front of a shelf it positions itself beside the customer with a shopping basket so that the customer can easily put his or her product in the basket. Afterwards, the robot again follows the customer from shelf to shelf until he or she is done with shopping. We conduct our experiments in a real supermarket to evaluate its effectiveness.

Keywords:
LiDAR, Histogram matching, Omnidirectional camera, GRU, Shopping behavior, OpenPose, Head orientation, Body orientation, DNN, Kinect camera, supermarket, person following, elderly.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Recently, the proliferation of big supermarkets and shopping centers, added to the rapid development of robot technology, has produced robotic systems for helping people in these particular environments. Some specific types of applications, such as enhancing the physical capabilities of the user, helping in the transport of products by providing a mobile shopping basket, and improving the information given to the user in a more intelligent manner, have been described in scientific literature.

Assisting elder people in their environments is a way of significantly improving their quality of life. As an example, networks of sensors and actuators embedded in buildings (i.e., Ambient Intelligence) can provide useful functions. In fact, several European Union projects have focused on this particular subject. However, helping elderly people outside their usual environment is also necessary, to help them to carry out daily tasks like shopping.

Population ageing is a concern for most developed countries and it is forecast that it will accelerate in the years to come. The main goal in the field of assistance robotics is to help users in their everyday lives, in tasks such as moving around, handling objects, or interacting, possibly remotely, with other people, either friends or relatives, or professionals. In this context, robot devices are aimed not only at assisting people, instead of replacing them, but also at adopting the role of companions. Therefore, they must be designed to interact with untrained people, and a great effort must be made to enable users to accept them easily.

Moreover, these assistance robots may have to operate in unstructured environments that cannot be customized, such as homes, shops, or parks. This implies a need for the robot that is conscious of its surroundings, which requires the use of a variety of sensors. In addition, many of the things that need to be handled, such as food, do not come in standard sizes and shapes, as they often do in an industrial context.

1.2 Objectives

The above described future visions for shopping robots motivated the further work during this doctoral study. Thus, the dissertation deals with enabling shopping robots with capabilities to follow humans and recognize the shopping behavior also. More specific, the emphasis is on enabling a shopping robot to follow target person so that it can perceive customer behaviors to read their interests, intentions, and preferences inside a shopping mall using modern remote human sensing technologies.

I propose a robust targeted person following shopping support robot so that it can track target person in a large scale real public spaces even in crowded situation by combating against partial and full occlusion. Some of the challenges has been developed into questions, which are investigated through the dissertation. The research questions are:

•**Person detection and tracking:** How can shopping support robots detect and track human through any social spaces by only using skeleton joints detection?

•**Person's head and body orientation detection:** How can shopping support robots detect target person's head and body orientation by only using skeleton joints detection?

•**Shopping behavior recognition:** How can shopping support robots recognize customer shopping behavior by only using skeleton joints detection?

•**Support for the elderly:** How can shopping support robots gives proper shopping support based on body orientation and shopping action recognition?

1.3 Research Contributions

This research resulted in designing the targeted person-following shopping support robot with the ability to avoiding crowd in the crowded shopping mall environment.

The main contributions of this thesis are:

• An experimental paradigm for studying how our robot track a target person using color histogram matching algorithm.(Chapter 3)

• An experimental paradigm for studying how our robot find the location of the target person in space using LiDAR sensor and 360° camera so that it can follow the target person.(Chapter 3)

- An experimental paradigm for studying how our robot give the proper support of a target person in shopping mall depending on the head and body orientation of the target person.(Chapter 3)
- An experimental paradigm for studying how our robot recognize customer's shopping action using GRU network. (Chapter 4)
- An experimental paradigm for studying how our robot give the proper support of the target customer depending on the shopping action, head orientation and body orientation(Chapter 4)
- In last work, finally modify the robot to meet practical needs: speed and accuracy. To do that we use Kinect based 3D skeleton tracking instead of OpenPose model based 2D skeleton tracking.(Chapter 5)

1.4 Organization of Sections

Chapter 2 - Background and Literature Review

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

Chapter 3 - A Person-Following Shopping Support Robot Based on Human Pose Skeleton Data and LiDAR Sensor

In Chapter 3, We develop a targeted person following robot system that can support the elderly using body orientation recognition procedure.

Chapter 4 - An Intelligent Shopping Support Robot: Understanding Shopping Behavior from 2D Skeleton Data Using GRU Network

Only person following and body orientation recognition is not sufficient enough for supporting elderly in shopping mall. With these properties we include GRU network based customer shopping action recognition system to give proper support for the elderly.

Chapter 5 - Person-Following Shopping Support Robot using Kinect Depth Camera based on 3D Skeleton Tracking

In the last work to operate our robot in a practical environment we must ensure three requirements such as speed, accuracy and cost. For this reason, we replace OpenPose model with Kinect V2 depth camera for skeleton tracking and ensure us these three requirements.

Chapter 6 - Conclusions

Conclude the dissertation with a summary of the concepts and designed intelligent shopping support robot for the elderly followed by the potential future works and application.

Chapter 2

Background and Literature Review

This chapter provides the literature review covering the theoretical background and the state-of-the-art research topics accomplishments in the area of elderly support robot in order to frame the context of the presented thesis work.

Person following scenarios arise when a human and an autonomous robot collaborate on a common task that requires the robot to follow the human. Usually, the human leads the task and cooperates with the robot during task execution. An example application would be the service robots, which are widely used in industrial applications, e.g., in manufacturing, warehousing, health care and shopping mall. The use of companion robots in surveillance, social interaction, and medical applications has also flourished over the last decade. Numerous new applications are also emerging in the entertainment industry as robots are getting more accessible for personal use.

Many researchers use stereo cameras to track people from moving platforms [1] and achieve person following through appearance models and stereo vision using a mobile robot [2]. This is a well-known method for person following robots. To follow a person, the robot must continuously receive two types of information, such as position and distance data.

The person following robot called "ApriAttenda" and "Nurse Following Robot" was created by T. Sonoura et al and B. Ilias et al. in [3] [4]. The main task of these robots was to find and assign a person and continuously follow that person everywhere. Using a laser range finder (LRF), the OSAKA Institute of Technology has created a mobile robot named ASAHI, with semi-autonomous navigation using simple and robust person following behavior [5]. To follow a person T. Germa et al. have de-veloped a mobile robot, named Rackham [6]. This robot uses one digital camera, one ELO touch screen, a pair of loudspeakers and an RFID system.

S. Nishimura et al. [7] developed an autonomous robotic shopping cart. This shop-ping cart can follow customers autonomously and transport goods. Kokhtsuka et al. [8] provide a conventional shopping cart with a laser range sensor to measure the distance from and the position of its user and develop a system to prevent collisions. Their robotic shopping cart also follows users to transport goods.

Hu et al. [9] proposed an action recognition system to detect the interaction between the customer and the merchandise on the shelf. The recognition of the shopping actions of the retail customers was also developed by using a stereo camera from the top view [10]. Lao et al. [11] used one surveillance camera to recognize customer's actions, such as pointing, squatting, raising hand and so on.

During the last decades, several teams of roboticists have presented the idea of new shopping support robot prototypes, representing worldwide cutting edge advancements in the field. An autonomous robotic shopping cart was developed by Nishimura et al. [7]. This shopping cart can follow customers autonomously and transports the goods. Kohtsuka et al. [8] followed a similar approach: they provide a conventional shopping cart with a laser range sensor to measure distance from and the position of its user and develop a system to prevent collisions. Their robotic shopping cart also follows users to transport goods.

The study carried out in [12] concludes that elderly people interact in a better way with robots carrying the shopping basket and providing conversational facilities. In [13, 14] a shopping help system was developed and able to obtain the shopping list from a mobile device through a QR code, carry the shopping basket and show at each moment, which articles are on it, and communicate with the supermarket computer system to inform about the location of articles. It uses a laser range finder, sonar, and contact sensors (bumpers) to navigate. An indoor environment shopping cart tracking system was developed in [15]. This system needs the installation of a computer and a video camera on the shopping carts, so that they can perform self-localization and send their positions to a centralized system.

In addition to customer shopping support, the analysis of customer shopping behavior is commercially important for marketing. Usually, the records of cash registers or credit cards are used to analyze the buying behaviors of customers. But this information is insufficient for understanding the behaviors of customers for situations such as when he or she shows interest while in the front of a given merchandise shelf but does not make any purchases. The main task of customer shopping behavior recognition is to count the customers and analyze the trajectory of customers so that merchants can easily understand the interests of customers. Haritaoglu et al. [16] described a system for counting shopping groups waiting in checkout lanes. Leykin et al. [17] used a swarming algorithm to group customers throughout a store into shopping groups. For marketing and staff planning decisions, person counting is a useful tool. For understanding the hot zones and dwell time trajectories of individual customers from surveillance cameras in retail store were analyzed by Senior et al. [18]. However, customer shopping behavior includes more diverse actions, such as: stopping before products, browsing, picking up a product, reading the label of the product, returning it to the shelf or putting it into the shopping cart. Those different behaviors or combinations of them show much richer marketing information. Hu et al. [9] proposed an action recognition system to detect the interaction between customer and the merchandise on the shelf. The recognition of the shopping action of retail customer was also developed by using stereo cameras from a top view [10]. Lao et al. [11] recognize customer's actions, such as pointing, squatting and raising hand using one surveillance camera. Haritaoglu et al. [19] extracted customer behavior information whenever they watched advertisements on a billboard or a new product promotion.

2.1 Definition of Human Support Robots

The Human Support Robot (HSR) is a compact mobile manipulator for the disabled and elderly. HSR can move around the house, keep watch over family members, and fetch objects. Its creators hope to make HSR beneficial to all people in the near future.

2.2 Potential Application of Human Support Robot

The main focus of this dissertation is on "Human Support Robot" in the role of peer-type human partners in social environments. Thus, applications for human support robot would include services that are typically provided by people. This section will discuss examples of services that could be performed by human support robot in real world environments.

2.2.1 Human Support Robot in Shopping Mall

CompaRob: A person following shopping cart assistance robot based on the Erratic robotic mobile platform equipped with a shopping basket.The cart uses ultrasonic sensors and radio signals to provide a simple and effective person localization and following method. Moreover, the cart can be connected to a portable device like a smartphone or tablet, thus providing ease of use to the end user.

Figure 2.1: A person following shopping cart assistance robot [31].

RoboShop: A helper robot in a shopping mall. This robot uses Laser Range Sensor, Infra-Red Telemeters, Sonars, Odometers, Motors, Embededded PC, Touch Screen.

Figure 2.2: RoboShop: Helper Robot in a Shopping Mall [32].

This robot provide the following services:

- Picking in a shop or a warehouse
- Guide in a shopping mall
- Mobile interactive information booth
- Virtual shopping via tele-presence
- Announcement of news and special offers

- Distribution of flyers or free products
- Surveillance during closing hours

2.2.2 Human Support Robot in Other Disciplines

Hospital Nurse Following Cart: This robot can carry a load of 20 kg and used dc geared motor to move. The mobile platform is able to rotate at axial axis with the construction of special wheel and the placement of the motor. A suitable ultrasonic sensor bank is selected so that robot can detect obstacle around the mobile platform and avoid the obstacle. The robot control and obstacle avoidance system is designed by adopting the facilities of Basic ATOM microcontroller.

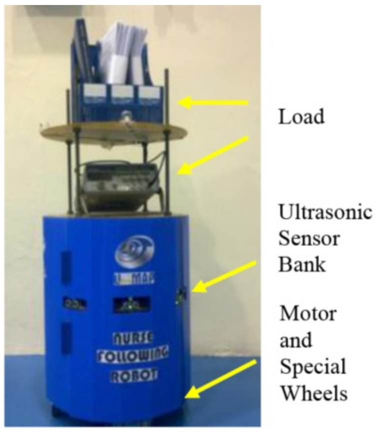


Figure 2.3: A nurse following robot [4].

Luggage Carrying Cart: This work introduces a visual based sensor cart follower to ease the mobility of wheelchair user in carrying their luggage. The cart will track and follows the wheelchair without having any physical connection to the wheelchair. The cart is equipped with a microcontroller, distance sensor, ultrasonic sensors, vision sensor, and several motors. The cart will move and follow the wheelchair if vision sensor (CMUcam5^{TM}) detects the predefined color pattern.

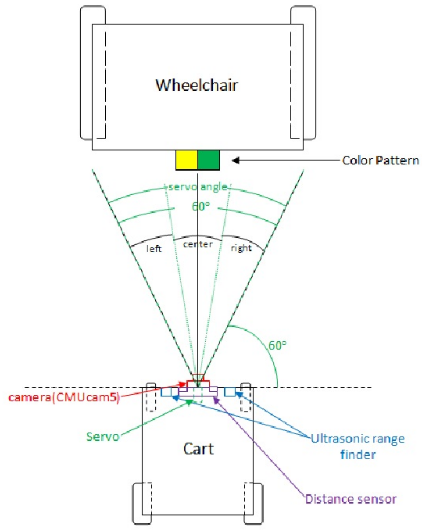


Figure 2.4: A luggage carrying cart [21].

2.3 Customer Shopping Behavior Recognition

Consumer behaviour is the study of human response to products and services. It is important to understand “Why” and “ How’s” of buyers behave so manufacturer can do a better job of developing quality of product, charging reasonable price,improvement in distributing product through various channels, and promoting goods and services with various promotion measure for the group of consumers.

2.3.1 Shopping Behavior Understanding

Behavior understanding, or recognition, aims at classifying time varying feature data. The goal is, in most cases, to match test sequences with labeled sequences used as reference. There are many types of methods used to recognize human behavior.

Human model based methods: This model use a full 3D or 2D model of human body parts. Action recognition is achieved using information on body parts pose and motion

Holistic methods: This type of method use knowledge about the localization of humans in video and learn an action model from global body motion, without any notion of body parts.

Local feature methods: Only use descriptors of local regions in a video. No prior knowledge about the position of human or its limb is given. Human models require a prior knowledge and can not be applied on various scenes.

Dynamic time warping: It is simple and robust template-based matching method [33]

Finite state machines (FSM): Finite state machines are made of states and transition function between them. These transition functions are the most important feature of a FSM. The machine can be deterministic and can recognize behavior without learning involved [34], unlike the other methods.

Human model based methods: This model use a full 3D or 2D model of human body parts. Action recognition is achieved using information on body parts pose and motion

Hidden markov model (HMM): HMM is a stochastic state machine that allows analysis of spatio-temporal varying data. Given a number of states, HMM optimizes state transitions as well as output probabilities. HMM is extensively applied since it outperforms DTW. Later approaches use couple hidden Markov model [35].

Neural network: It is a non-linear statistical data modeling technique. Time-delay neural network adds delay units to a general static network [36]. Self-organizing neural network allows the description of unrestricted object motion [37].

Human model based methods: This model use a full 3D or 2D model of human body parts. Action recognition is achieved using information on body parts pose and motion

Support vector machine (SVM): SVM is a discriminative classifier defined by a separating hyperplane [38]. SVMs have been successfully applied to various recognition tasks.

2.3.2 Shopping Actions

There are many aspects of recognizing shopping actions.The customer behavior information allows for enhanced customer experience, optimized store performance, reduced operational costs, and ultimately higher profitability. Traditionally, retailers use the records of cash registers or credit cards to analyze the buying behaviors of customers. However, this information cannot reveal the behaviors of customer when he or she shows interest on the front of the merchandise shelf but does not buy. For example, a customer stops on the front of the merchandise shelf but does not pick anything, or a customer pick an item but then return it to the shelf. Some related works regarding to shopping actions recognition shown in below:

Customer behavior classification using surveillance camera for marketing: Liu et al. developed a behavior model for recognizing of customer shopping actions recognition using surveillance camera for marketing [39]. Behavior model reveals the increasing interest of customer to the product. If a customer has no interest to the product, he or she will neither look at the shelf nor turn to the shelf. When the customer has initial interest, his or her gaze will fall upon the product. If the customer has more interest, he or she will turn to shelf. These viewing and turning behaviors reveal two basic interest levels. Table 2.1 shows the customer shopping actions and corresponding action definitions

Figure 2.5 shows the examples of different customer behaviors.

Analysis of Shopping Behavior based on Surveillance System: Popa et al. developed a model of shopping behavior based on surveillance system [40]. They have classify the following shopping behavior.

1. Goal Oriented Shopper: A goal-oriented shopper has a shopping list, knows the location of the product and walks directly to that place at a high speed.

2. Disoriented Shopper: The disoriented shopper has no specific idea about what he wants, doesn't know if the product is available or where to find it.

3. Looking for help behavior: Looking for the shop assistant or waiving their hand, asking information about the location of the products, alternatives, and characteristics of the product.

Figure 2.5: Examples of different customer behaviors: (a) no interest; (b) viewing; (c) turning to shelf; (d) touching;(e) picking and returning; (f) picking and putting [39].

Table 2.1: Customer behavior model [39].

Shopping actions	Definition
No interest	A customer neither look at the shelf nor turn to the shelf.
Viewing	A customer stops on the front of shelf and looks at a certain product but does not turn to shelf.
Turning to shelf	The customer turns to the shelf but with no arm action.
Touching	The customer touches the product but does not pick it from the shelf.
Picking and returning	The customer picks a product then puts it back to the shelf.
Picking and putting	The customer picks a product then puts it into the shopping basket.

Figure 2.6 shows customers behaviors in shopping mall.

Figure 2.6: Customer shopping behaviors [40].

Kinect Sensing of Shopping Related Actions: Popa et al. developed a model of kinect sensing of shopping related actions [41]. They have classify the following shopping behavior. Figure 2.7 shows their system architecture and Figure 2.8 shows the basic shopping actions descriptions.

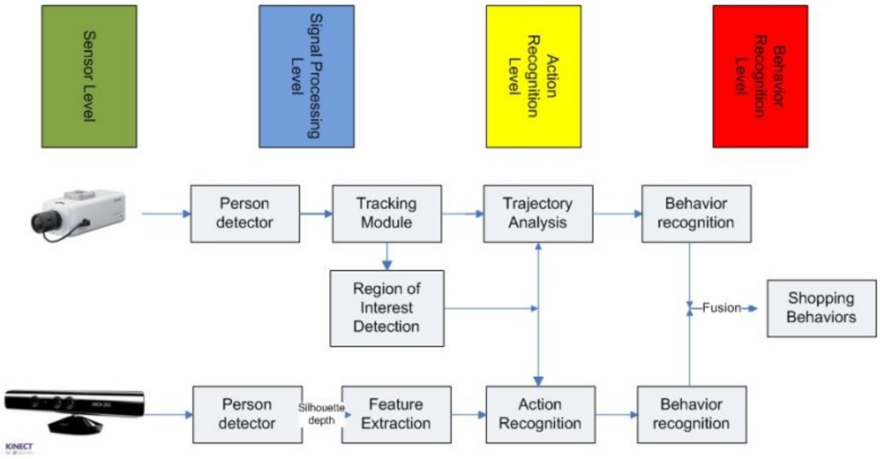


Figure 2.7: System architecture [41].

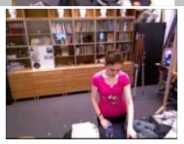
Actions	Description	Key Pose(s) – single shopper	Key Pose(s) – duo shopper
Browsing	The person goes through the products on the table using the hands.		
Examining	The person holds one or more products in her/his hand, looking at them more closely.		
Trying on	The person is putting on a product, like a jacket, scarf, vest, or sun glasses etc.		
Picking	The person leans to the table, picks a product, and takes the hand back.		
Looking for help (waving one hand)	Having one or two arms in the air, the person waves at some direction, trying to get the attention of an assistant.		
Shopping Cart Interaction	The person is holding the shopping cart, drives it in shop (pushes or pulls it) and interacts with it.		

Figure 2.8: Basic shopping actions descriptions [41].

2.4 Chapter Summery

This chapter covers different smart technologies related to human support robot in shopping mall and other discipline. From the review of many published papers, it is concluded that researchers are continuously trying to build powerful and helpful human support robot by recognizing human behavior.

Chapter 3

A Person-Following Shopping Support Robot Based on Human Pose Skeleton Data and LiDAR Sensor

3.1 Introduction

In the field of computer vision, research in human detection and tracking is a challenging area with decades of efforts. This area has numerous applications in robot vision [20]. A prime example is in human following robots, where such types of robots must localize the walking person and avoid obstacles. These human following robots also are used in many applications such as wheel-chairs [21] that carry belongings when travelling, automated shopping cart robots [22], nurse following carts [4] and a line following intelligent shopping cart controlled by a smartphone that can guide the cart to the locations of the wanted items to be purchased [23].

Our goal of this research is to track a target person based on a pose skeleton model and a color histogram matching algorithm and compute the target person's body movement orientation for shopping support. Specifically, we use the OpenPose model [24] implemented on a 360° camera and crop the body image of target person using the left shoulder, right shoulder, left hip and right hip key points, then further apply a color histogram matching algorithm in each frame to track the target person. After tracking the target person, we take the left or right ankle key points and calculate the angle value (0° to 270°) of the person's position in the image frame according to Figure 3.4. When we get the angle value of person's position, we take the rotation and translation value of the robot using LiDAR (Hokuyo UTM-30LX) sensors.

3.2 Design Approach

The major issue of person-following is how to localize the target person by tracking. A common approach is to localize the target person over time and identify them based on motion continuity. Figure 3.1 shows our system model. In this model, we first detect the person's pose using a 360° camera. This wide-angle camera allows us to detect different people for all 360° viewing angles. Among these detected people, we apply the color histogram intersection algorithm to track the target person. After we get the target person, we record that person's ankle x coordinate points. Using these points, we calculate the angle value (0° to 270°) of the person's position in the image frame according to Figure 3.3. By this angle and LiDAR sensor data, we finally calculate the robot rotation and translation values to follow the target person.

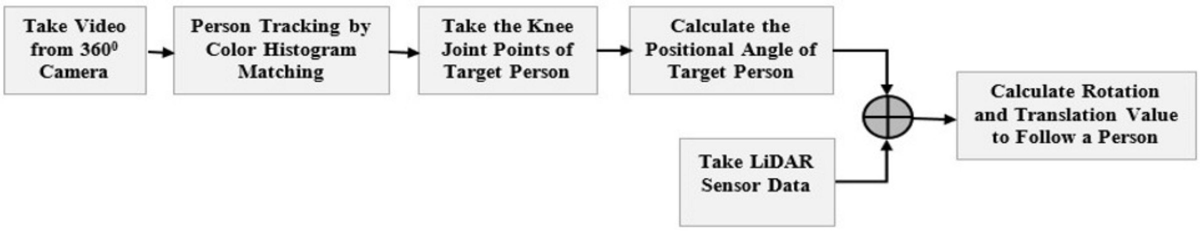


Figure 3.1: Block diagram of person following procedure.

3.3 Person Tracking

In our system, we use the histogram intersection algorithm to track the target person. From the human pose skeleton data, we crop the body using the left shoulder, right shoulder, left hip and right hip key points as shown in Figure 3.2. Then in each frame, we compare each person's body image to the cropped image using the histogram intersection algorithm.

The histogram intersection algorithm was proposed by Swain and Ballard in [30]. Here the author considered two histograms I and M, each containing$n$ bins, and the intersection is defined as,

$$\sum_{j=1}^{n} min(I_j, M_j) \tag{3.1}$$

We get the output of the histogram intersection model with an image histogram as the number of pixels from the model that have corresponding pixels of the same color in the image. The match value is given by,

$$\frac{\sum_{j=1}^{n} min(I_j, M_j)}{\sum_{j=1}^{n} M_j} \tag{3.2}$$

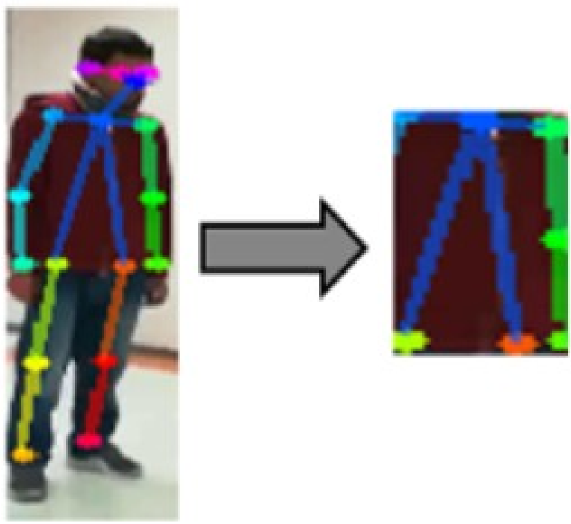


Figure 3.2: Cropped image from skeletal key points information.

When an unknown object image is given as input, we compute the histogram intersection for all the stored models, the highest value is the best match.

3.4 Positional Angle Value of Tracked Person

We calculate the positional angle value of the tracked person based on the 360° wide angle camera. However, in practice, we find it sufficient to calibrate the total width of the image to be from 0° to 270° as shown in Figure 3.3. Given the width of the image is W and the x coordinate value, A then we can easily calculate the ankle angle value by the following equation,

$$Person's\ positional\ angle\ value = \frac{270}{W} * (W - A) \tag{3.3}$$

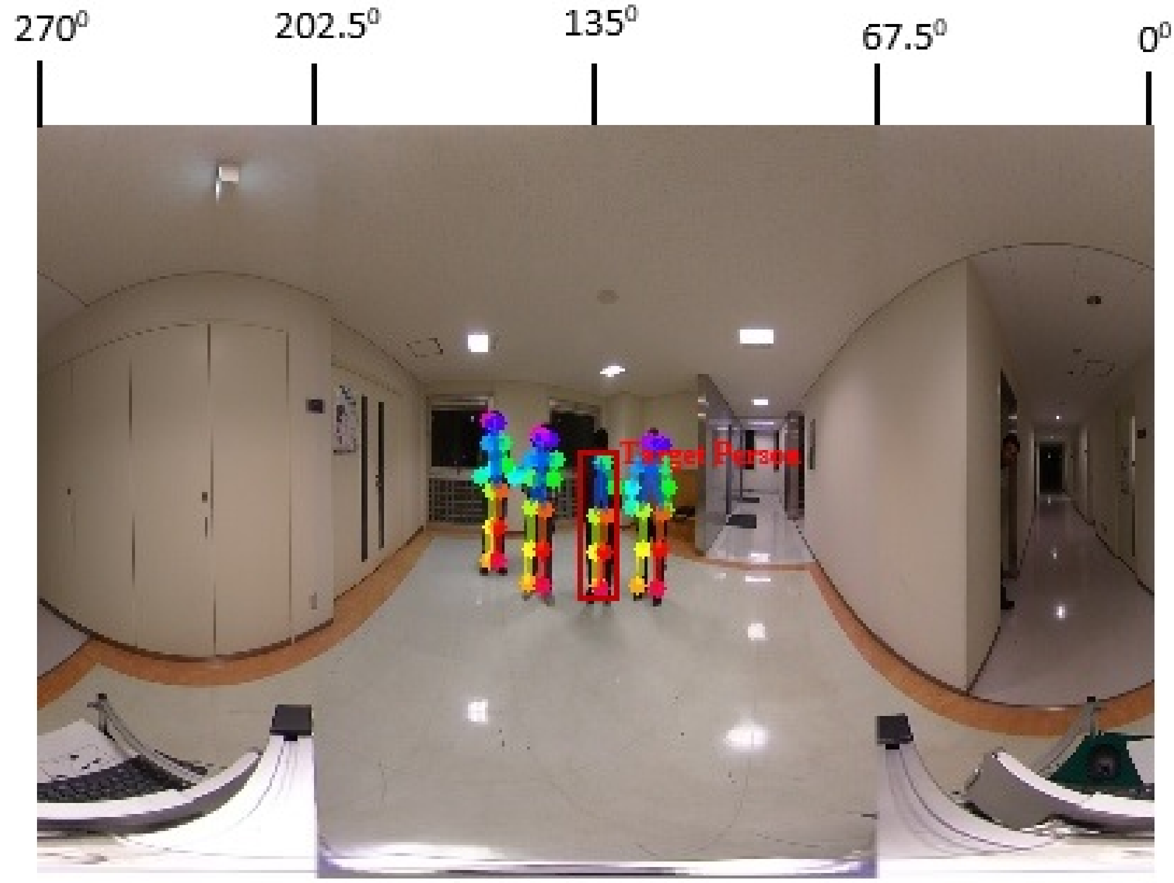


Figure 3.3: Calibrating the width of the image 0° to 270°.

3.5 Calculate the Rotation and Translation Value

Figure 3.4 shows how we calculate the rotation and translation values of our mini cart robot. The whole procedure is given below:
Step 1: Track the target person among multiple people using the histogram intersection algorithm described in Section 3.3
Step 2: Take the tracked person's ankle x coordinate value.
Step 3: Calculate the tracked person's ankle angle using Equation (3.3) as described in Section 3.4
Step 4: If the person's angle value is not in the range $(100 - 135)^{\circ}$ of the LiDAR sensor we calculate the rotation angle θ needed to be in the range $(100 - 135)^{\circ}$, as illustrated by the example in Figure 3.4
Step 5: After rotation of our mini cart at the θ we calculate the distance between the

tracked person's ankle points to the mini cart robot. We also set the threshold value such that if the distance is greater than $1m$, we operate the translation operation. We calculate the distance using the Euclidean distance as,

$$Distance = \sqrt{(0-x)^2 + (0-y)^2} \tag{3.4}$$

Where (x, y) is the coordinate value of the left or right ankle from the sensor output. Using the above procedure, the robot follows its target person effectively.

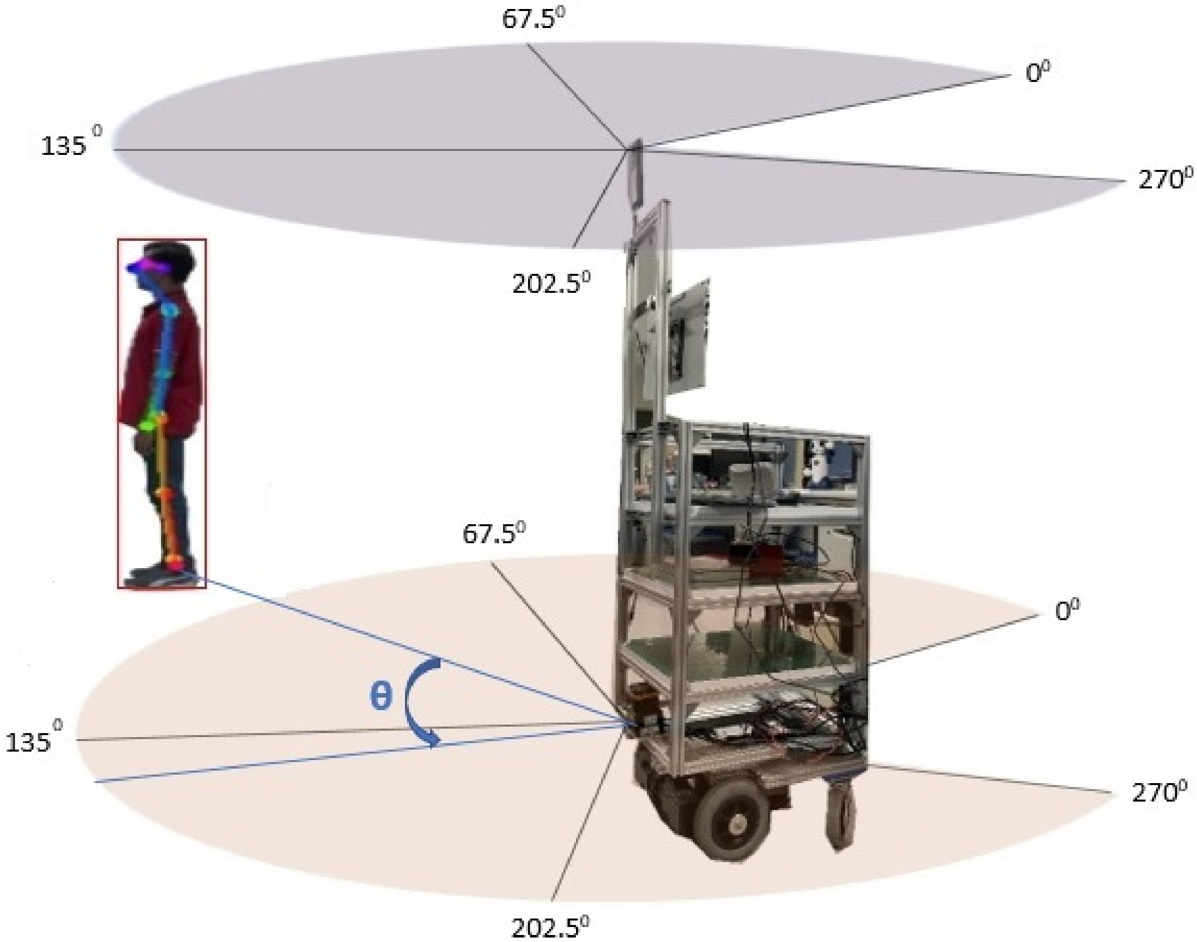


Figure 3.4: Rotation and translation of mini cart robot.

3.6 Body Orientation Angle of Target Person

The body orientation of the target person is estimated in eight directions as shown in Figure 3.5(b). We take four angle values of the target person to predict body orientations. As for example to calculate the angle of $\angle EAC$ as shown in Table 1, we use the following formula,

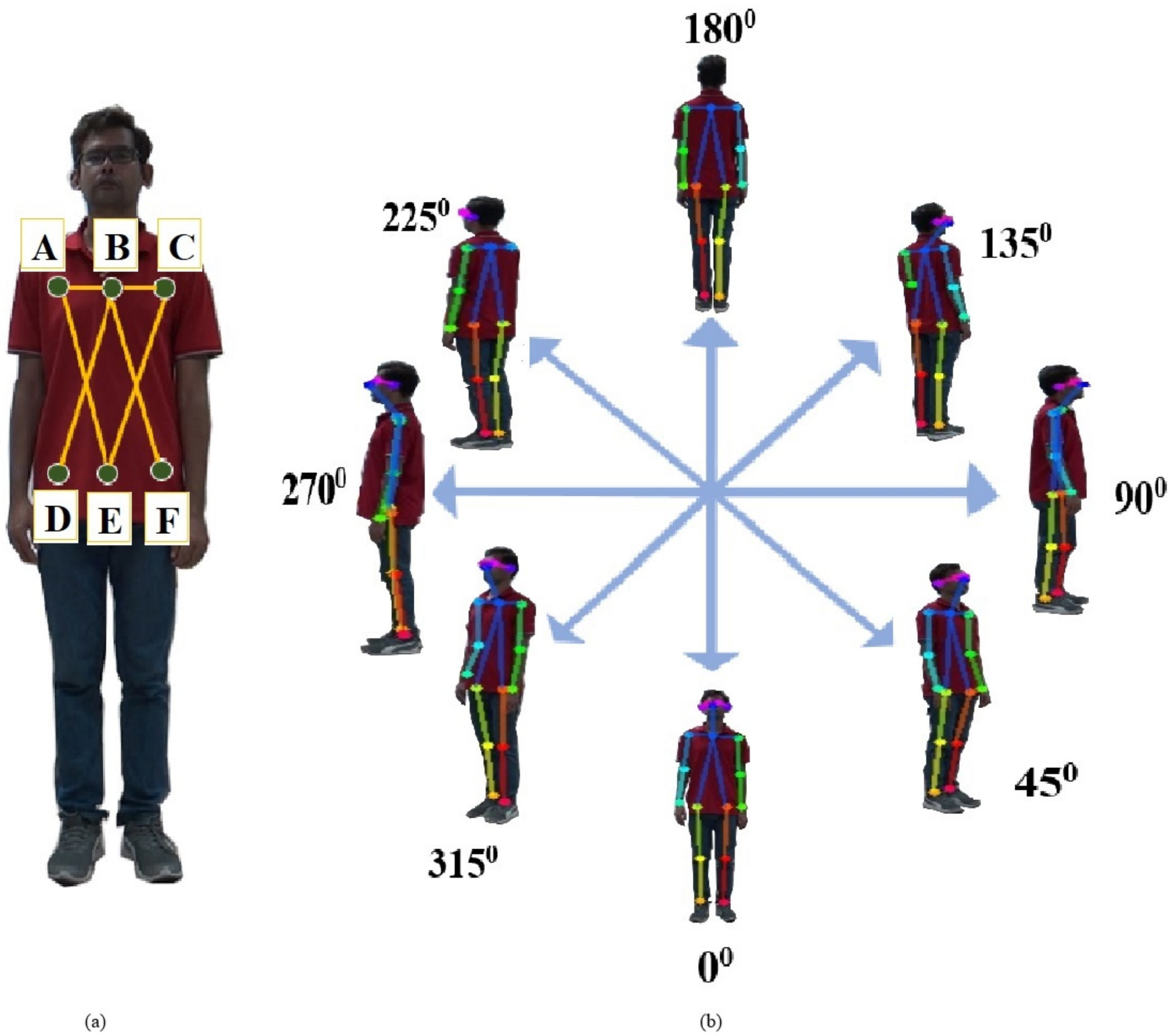


Figure 3.5: Examples of different body orientations.

First, we generate the vector E and C as,

$$E = EA = (e1 - a1, e2 - a2) = [E1, E2] \tag{3.5}$$

$$C = CA = (c1 - a1, c2 - a2) = [C1, C2] \tag{3.6}$$

The dot product of vector E and C is defined as,

$$E.C = E1C1 + E2C2 \tag{3.7}$$

The magnitude of a vector E is denoted by $||E||$. The dot product of vector A with itself is

$$E.E = ||E||^2 = E_1^2 + E_2^2 \tag{3.8}$$

The magnitude of a vector E is denoted by $||E||$. The dot product of vector A with itself is

$$||E|| = \sqrt{E.E} = \sqrt{E_1^2 + E_2^2} \tag{3.9}$$

Similarly, we can calculate

$$||C|| = \sqrt{C.C} = \sqrt{C_1^2 + C_2^2} \tag{3.10}$$

The dot product of two non-zero Euclidean vectors A and B is given by

$$E.C = |E||C|cos\theta \tag{3.11}$$

Where θ is the angle between E and C.

Table 3.1: Different angles according to Figure 3.5(a) to predict target person's body orientations.

Body Orientations	$\angle EAC$	$\angle ECA$	$\angle AEC$	$\angle DBF$
0°	-72.65	81.22	26.12	-15.83
45°	-75.31	87.45	17.23	-11.50
90°	127.90	-48.48	-3.61	2.95
135°	81.41	-78.05	-20.55	14.45
180°	76.57	-75.73	-27.70	16.72
225°	72.65	-90.83	-16.52	11.91
270°	-80.98	95.14	3.88	-4.74
315°	-77.47	83.09	19.44	-12.69

Table 3.1 shows examples of four angle data to classify eight direction body orientations. We take different person's four angle data according to Figure 3.5(a) and make a dataset for SVM classifier to classify person's body orientations.

Table 3.2 shows the body orientations classification result using SVM classifier.

Table 3.2: Evaluations of body orientation classification using SVM classifier.

Body Orientations	Precision	Recall	F1 Score	Support
0°	1	1	1	22
45°	0.68	0.65	0.67	23
90°	1	1	1	18
135°	1	1	1	24
180°	1	1	1	20
225°	1	1	1	21
270°	1	1	1	15
315°	0.56	0.59	0.57	17
Avg/Total	**0.90**	**0.91**	**0.91**	**160**

3.7 Head Orientation Angle of Target Person

The head orientation of the customer is estimated in eight directions as shown in Figure 3.6. We propose a simple method of detecting head orientation from OpenPose [17] results. In OpenPose, the whole body pose is represented by [0, 1, 2, ... , 17] joints as shown in the right hand side of Figure 3.6. Depending on the detected skeleton joint numbers, we can easily classify our head orientation. For example, for detecting the 0° head orientation, all head skeleton joint points [0, 14, 15, 16, 17] are detected. For other head orientations, detection of the corresponding head skeleton joint numbers are shown in Table 3.3.

3.8 Experimental Results

Figure 3.7 shows our histogram intersection algorithm-based person tracking results in different frames. Using this tracking result, we take the tracked person's left and right ankle points and calculate his or her positional angle value. Our LiDAR sensor (Hokuyo UTM-30LX) can estimate the distance by calculating the phase difference with in a range of 30m. The LiDAR sensor gives us 1080 distance points. We

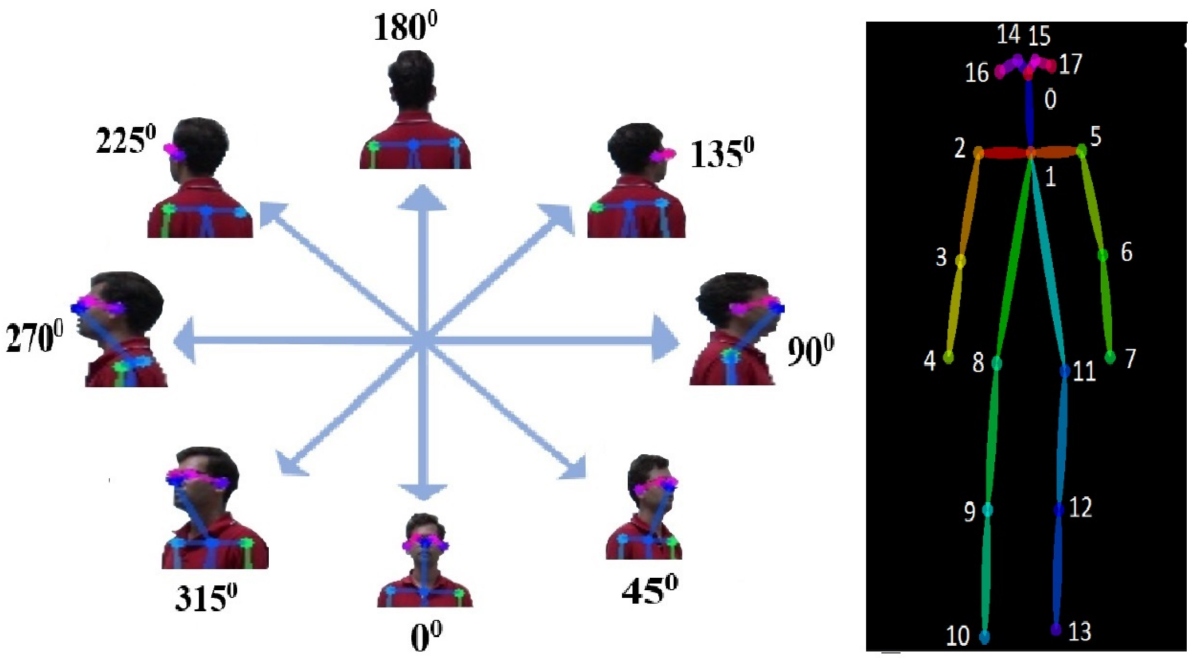


Figure 3.6: Examples of different head orientation detection.

Table 3.3: Detected head joint points for classification of head orientation.

Head orientations	Detected head joints
0°	[0,14,15,16,17]
45°	[0,14,15,16]
90°	[0,14,16]
135°	[14,16]
180°	[No points detected]
225°	[15,17]
270°	[0,15,17]
315°	[0,14,15,17]

calibrate these 1080 distance points to the $(0-270)^{\circ}$range. According to Equation 3.3 we take the target person's ankle angle and find the distance between robot and target person. In Figure 3.8 the red dots give the distance per degree. The blue and

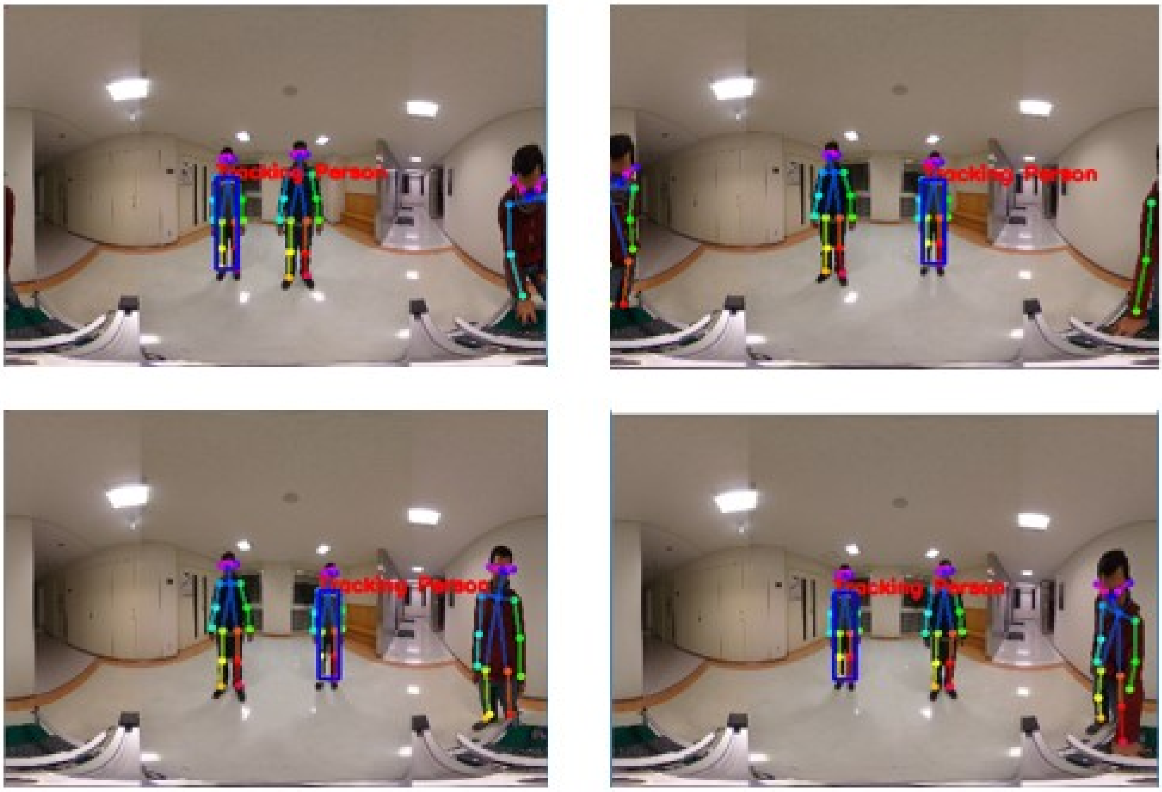


Figure 3.7: Result of tracking a person.

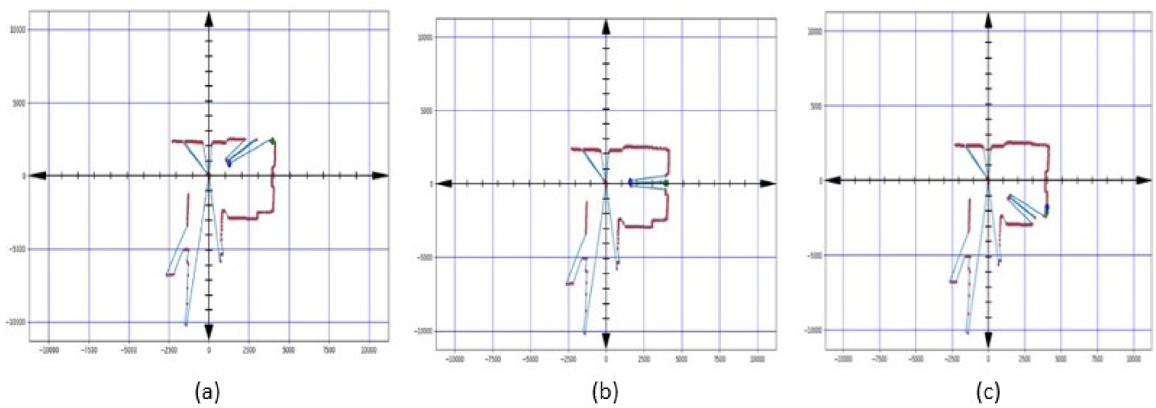


Figure 3.8: The LiDAR sensor output for calculating the target person's positional angle and corresponding distance.

green solid circles are the plot from the tracked person's left and right ankle angle using the 360° camera. The black square shows the actual blob of the left and right ankle angle detected by the sensor. We see that both solid circles and the detected

Table 3.4: Positional angle value of the left ankle by laser sensor and 360° camera

Fig. No.	Left ankle angle detected by laser sensor	Left ankle angle detected by 360° camera	Error
Figure 3.8(a)	167°	165°	2°
Figure 3.8(a)	136°	138°	2°
Figure 3.8(a)	107°	111°	4°

Table 3.5: Positional angle value of the right ankle by laser sensor and 360° camera

Fig. No.	right ankle angle detected by laser sensor	right ankle angle detected by 360° camera	Error
Figure 3.8(a)	169°	167°	2°
Figure 3.8(a)	140°	138°	2°
Figure 3.8(a)	111°	115°	4°

black square are very close.

Table 3.4 shows the person's positional left ankle angle value using the LiDAR sensor and 360° camera. We see that the detected angles of the LiDAR sensor and 360° camera are almost the same and varies only by $(2-4)°$which is an acceptable amount of variation. The cause of this variation is mainly due to the variation of detected left ankle skeleton data. In Table 3.5, we see that identical results are found for the right ankle by the laser sensor and 360° camera like for the left ankle angles.

Figure 3.9 shows that when the target person's body orientation is 180°, it follows the person with-in a certain distance. But when the person's distance is fixed, and his body orientation is 90°, the robot changes its direction according to the target person's body orientation.

Figure 3.9: Changing the orientation of our shopping support robot.

3.9 Chapter Summary

The goal of this thesis work is to design a targeted person following shopping support robot and provide support by taking the appropriate relative position with the customer. We have used color histogram matching algorithm to track the target person then we develop a method to calculate the target person's body orientation. According to the body orientation our robot take relative position.

Chapter 4

An Intelligent Shopping Support Robot: Understanding Shopping Behavior from 2D Skeleton Data Using GRU Network

4.1 Introduction

Nowadays, there are many applications in mobile robotics that are becoming a part of everyday life. Among them, shopping centers is one of the sectors where automated robots can be utilized to facilitate shopping activities. Shopping carts are broadly used in modern shopping centers, supermarkets and hypermarkets. However, pushing a shopping cart and moving it from shelf to shelf can be tiring and laborious job, especially for customers with certain disabilities or elderly. Sometimes, if a customer has one or more child it is difficult to push the cart as he or she has to hold his or her child's hand at the same time. In this situation, sometimes they need caregivers for support. To overcome this, an intelligent shopping support robot is a good replacement. In [25], Kobayashi et al. show the benefits of robotic shopping trolleys for supporting the elderly. In general, some of the core functionality of a shopping support robot are: following its user (customer), navigating through the paths that a customer takes during his or her shopping time and avoiding collisions with obstacles or other objects. Shopping malls or supermarkets typically have many crowded regions. For this reason, in our previous research [26], we developed an autonomous person following robot that can follow a given target person in crowded areas.

In addition to the robust person-following, the robot can more support the user if it can act in advance to meet the user's next move. For example, when the user picks up a product from a shelf, it is convenient if the robot automatically comes to the user's right hand side (if the user is right-handed) so that he or she can put it easily in the basket. To realize such functions, the robot needs to recognize the user's behavior. To recognize the user's behavior, we have used GRU (Gated Recurrent Unit) network [27] instead of LSTM network because the GRU network performance is better than LSTM. GRU has a simpler structure and can be computed faster. The three gates from LSTM are combined into two gates, respectively updating gate and resetting gate in GRU.

Before presenting the details of our methods, we would like to summarize our contributions of our paper. Firstly, we integrate head orientation, body orientation, GRU network for customer shopping behavior recognition and then, provide the shopping support to the customer. Here, we propose a GRU network to classify five types of shopping behavior: reach to shelf, retract from shelf, hand in shelf, inspect product and inspect shelf. Head and body orientations are used to classify customer gaze and interest in any given shelf.

4.2 Definition of Customer Behavior Model

Our customer behavior model captures indications of increasing interest that the customer has towards the store's products. If a customer has no interest in a given product, he or she will neither look at the shelf nor product and will likely not turn towards the shelf. We classify this behavior by our head and body orientation methods.

Other shopping behaviors such as reach to shelf, retract from shelf, hand in shelf, inspect product, inspect shelf are classified by our proposed GRU network. These behaviors indicate increasing interest levels to the product. These behaviors are defined in Table 4.1. Figure 4.1 shows some examples of these behaviors.

4.3 Framework of Customer Behavior Classification

Head orientation and body orientation are relevant to our shopping behavior recognition model. According to our previous work, a robot with shopping cart can be made to effectively follow a person. If the person's body orientation is $0°$ or $180°$ it just follows that person and the behavior is recognized as "no interest in the products". If the person's body and head orientation is neither $0°$ nor $180°$, then our proposed

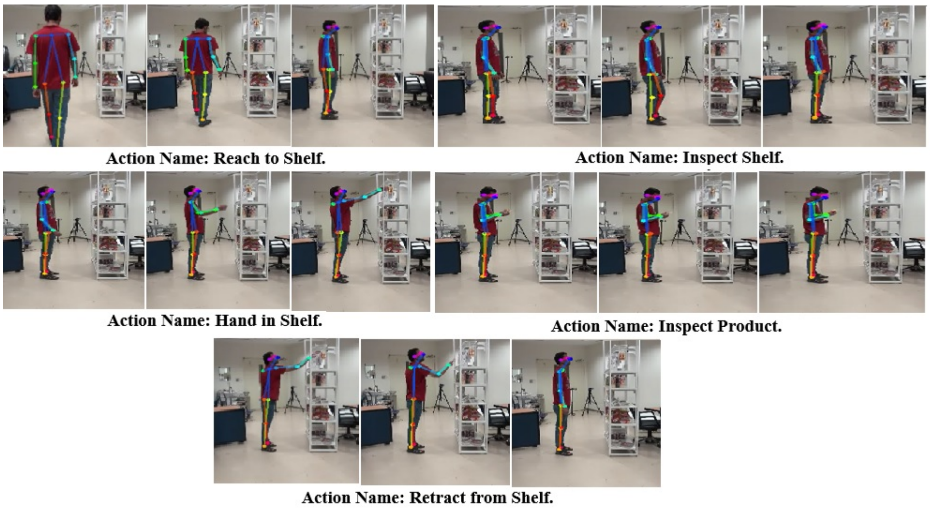


Figure 4.1: Examples of customer behaviors.

Table 4.1: Customer behavior model.

Customer behavior	Definition
No interest	If a customer does not look at the shelf or does not turn to the shelf.
Reach to shelf	If a customer reached to the shelf.
Inspect shelf	If customer inspects the whole shelf.
Hand in shelf	If a customer is going to pick a product.
Retract from shelf	If a customer picked a product from shelf.
Inspect product	If a customer inspects the product.

GRU neural network is used for classification of shopping behavior. The system's framework of customer shopping behavior classification is shown in Figure 4.2.

Action recognition in shopping environments is an important and challenging computer vision task. We introduce a framework for integrating human pose to detect

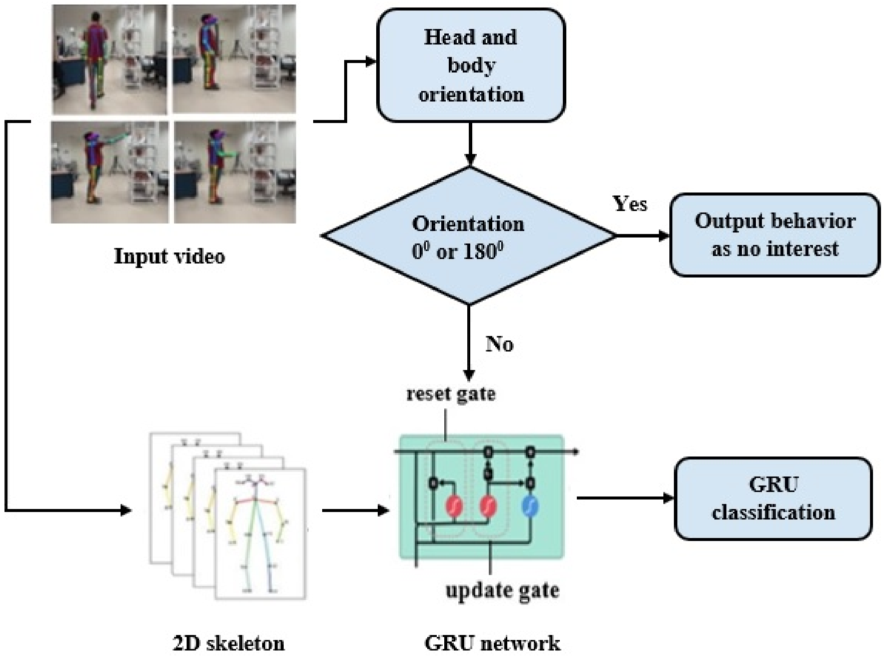


Figure 4.2: Framework of customer shopping behavior classification system.

and classify the activities in a shopping mall (very short and similar activities). Our system recognize six shopping actions according to Table 4.1. This shopping actions are important in shopping mall. The importance of these actions are describe below:

No interest: When our system get ”no interest” action of the customer it just follows the customer.

Reach to shelf: This action confirms us that the customer reach in shelf also confirm us that the customer growing interest on that shelf.

Inspect shelf: This action confirms us that the customer searching a product in that shelf.

Hand in shelf: This action represents that the customer try to grab a product from the shelf.

Retract from shelf: This action shows that the customer picking the product from the shelf.

Inspect product: This action represents that the customer checking the product quality, price and any other necessary things. It is important for analysis of customer-product interaction.

4.4 Gated Recurrent Neural Network (GRU)

The GRU is a similar network to the well-known LSTM. A GRU network has two gates, a reset gate and an update gate. The reset gate determines how to combine new inputs with the previous memory, and the update gate determines how much of the previous memory remains.

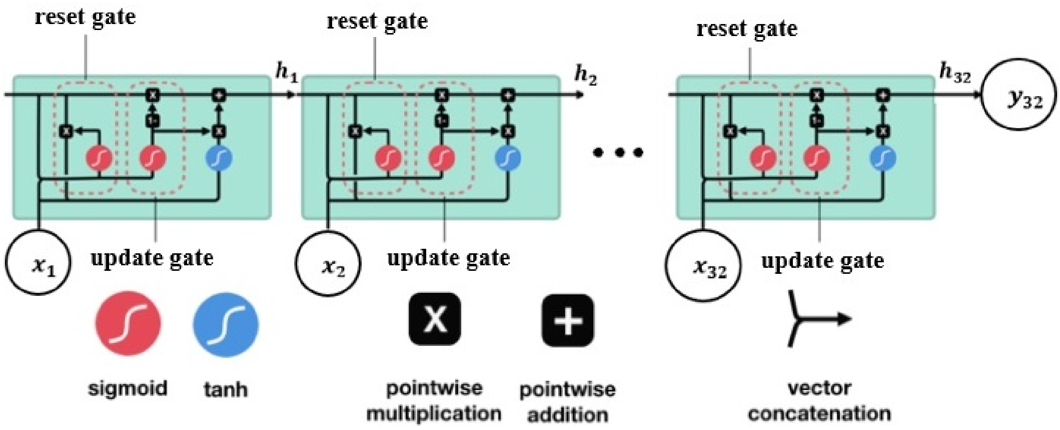


Figure 4.3: Proposed GRU network for shopping behavior classification [42].

As shown in Figure 4.3, our designed GRU network consists of 32 GRU cells. In our GRU model, the number of GRUs reflects the length of the activity video frames of skeleton data.

The activation h_t^j of the GRU at time t is a linear interpolation between the previous activation h_{t-1}^j and the candidate activation $\widetilde{h}_t^j$:

$$h_t^j = (1 - z_t^j)h_{t-1}^j + z_t^j\widetilde{h}_t^j \tag{4.1}$$

where an update gate z_t^j decides how much the unit updates its activation, or content. The update gate is computed by:

$$z_t^j = \sigma(Wx_t + U_z h_{t-1})^j \tag{4.2}$$

where x_t is the input sequence, W denotes the weight matrices and σ is the logistic sigmoid function. The candidate activation $\widetilde{h}_t^j$ is computed similarly to that of the traditional recurrent unit.

$$\widetilde{h}_t^j = \tanh(W_z x_t + U(r_t \odot h_{t-1}))^j \tag{4.3}$$

where r_t^j is a set of reset gates and $\odot$ is an element-wise multiplication. When off (r_t^j close to 0), the reset gate effectively makes the unit act as if it is reading the first symbol of an input sequence, allowing it to forget the previously computed state.

The reset gate r_t^j is computed similarly to the update gate:

$$r_t^j = \sigma(W_r x_t + U_r h_{t-1}))^j \tag{4.4}$$

And the output is given by

$$y = sigmoid(W_y h_{32} + b_y) \tag{4.5}$$

The output vector of y comes from the hidden state vector of h_{32} at the last time step of 32 which is multiplied by the weight matrix and added a bias as expressed in Equation (4.5). We use the sigmoid function as the network output activation function.

4.4.1 Dataset Construction

We built one dataset, that recorded five different kinds of shopping behavior for a certain period equally distributed among them. The actions consisted of: reach to shelf, retract from shelf, hand in shelf, inspect product and inspect shelf.

For creating the dataset, we constructed shopping shelves in our lab environment and put different items or products on the shelves. Then we setup four cameras for four angle views in recording videos. A total of 20 people took part in the video recording sessions. Each participant performed our desired shopping actions for 10 minutes. So, the total length of our video sequences is (20x10x4) 400 minutes as four cameras were used for each person. Then we ran the OpenPose model to extract skeleton data for each action.

We obtained 211,872 skeleton data of different actions. A single frame's input (where j refers to a joint) is stored in our dataset as:

$[j0_x, j0_y, j1_x, j1_y, j2_x, j2_y, j3_x, j3_y, j4_x, j4_y, j5_x, j5_y, j6_x, j6_y, j7_x, j7_y, j8_x, j8_y,$
$j9_x, j9_y, j10_x, j10_y, j11_x, j11_y, j12_x, j12_y, j13_x, j13_y, j14_x, j14_y, j15_x, j15_y,$
$j16_x, j16_y, j17_x, j17_y]$

4.4.2 Experiments Description

All the experiments were performed using a GPU NVIDIA GTX TITAN X, with 12 GB of global memory and with Nvidia Digits. We divided our dataset into two parts: 80% of the total data as training data and 20% of the total data as a testing data. Using this data, we trained our GRU network to classify our shopping behaviors. A fixed learning rate of 0.000220 was used. Our model was trained using 50000 epochs. The training took around 5 hours to finish. Other training specifications are given in Table 4.2.

Table 4.2: Training specification for our proposed GRU network.

Training parameters	value
Batch size	512
Epochs	50000
Timesteps	32
No. of hidden layer	34
Learning rate	0.000220
Optimizer	adam
Momentum	0.9

Figure 4.4 shows the plot of the model's loss and accuracy over 50000 iterations. Table 4.3 shows the detailed layer information for our proposed GRU network struc-

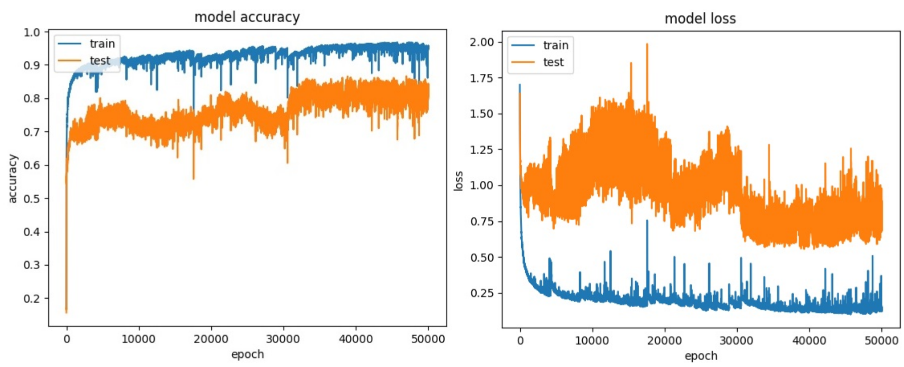


Figure 4.4: The model accuracy and loss over 50000 iterations.

ture. It has three layers. The first layer is GRU layer and it is the main layer containing two gates, a reset gate and an update gate. The second layer is a Dropout layer and it reduces overfitting. The last layer is a Dense layer and it is a fully connected layer.

Table 4.3: Detailed layer information for the proposed GRU structures.

Layer(type)	Output Shape	Param #
gru_1 (GRU)	(512, 32, 34)	7038
dropout_1 (Dropout)	(512, 32, 34)	0
dense_1 (Dense)	(512, 32, 5)	175

4.5 Architecture of the Shopping Support Robot Based on the User's Behavior Recognition

Figure 4.5 shows our proposed shopping support robot. First, it detects the nearest person as the user and starts following him/her. It can robustly follow the target person in crowded places. The details of our person tracking and following system are discussed in our previous paper [2]. Our shopping support robot uses a LiDAR sensor about 20 cm high from the floor. So, the sensor can cover customers of any height. Then our subsequent task is to develop a shopping support robot that can recognize the customer's behavior and intensity of interest in the products.

Figure 4.6 shows the flowchart of our behavior based shopping support robot. The total working procedure is given below:

Step 1: Track the target customer.

Step 2: Recognize the customer's body orientation. If the customer's body orientation is 0°, go to step 3. If body orientation is 180°, go to step 4. Otherwise, go to step 5.

Step 3: Recognize the customer's head orientation. If the customer's head orientation is 0°, take a suitable position in front of the customer. Otherwise, go to step 5.

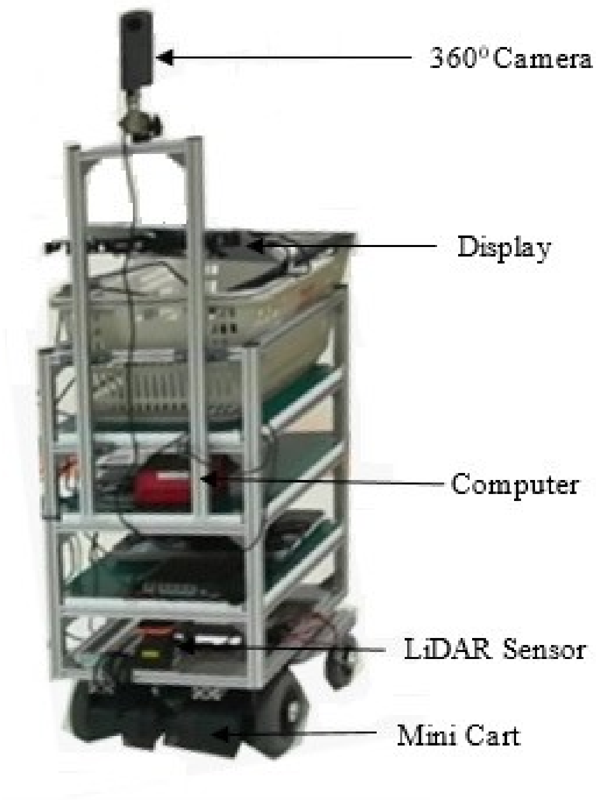


Figure 4.5: Our proposed shopping support robot.

Step 4: If the customer's head orientation is 180°, follow the target customer at a certain distance. Otherwise, go to step 5.

Step 5: Recognize the customer's shopping behavior actions using GRU network.

4.6 Results

4.6.1 Evaluation of Behavior Recognition

4.6.1.1 Performance Metrics

To verify the performance of behavior recognition, we employed four widely used evaluation metrics for multi-class classification.

Precision The precision or positive predictive value (PPV) is defined as the proportion of instances that belongs to a class (TP: True Positive) by the total instances, including TP and FP (False Positive) classified by the classifier as belong to this

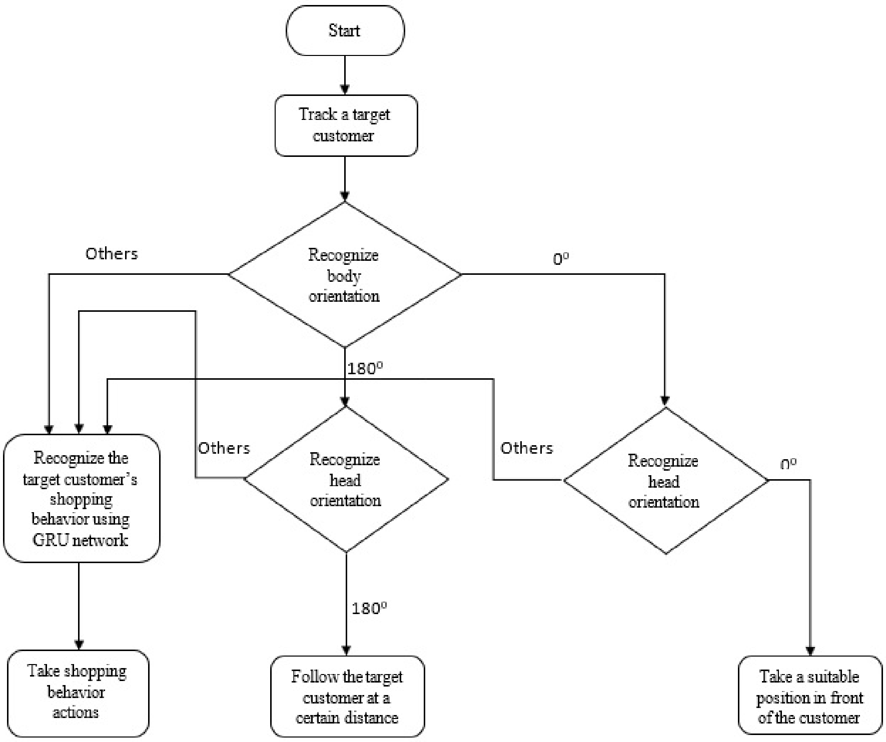


Figure 4.6: . Flowchart of our proposed shopping support robot.

particular class.

$$Precision = TP/(TP + FP) \tag{4.6}$$

Recall The recall or sensitivity is defined as the proportion of instances classified in one class by the total instances belonging to that class. The total number of instances of a class includes TP and FN (False Negative).

$$Recall = TP/(TP + FN) \tag{4.7}$$

Accuracy Measures the proportion of correctly predicted labels over all predictions:

$$Over \quad all \quad accuracy = (TP + TN)/(TP + TN + FP + FN) \tag{4.8}$$

F1 measure A weighted harmonic means of precision and recall. The F1 score can be interpreted as a weighted average of the precision and recall, where an F1 score ranges from 0 to 1. The relative contribution of precision and recall to the F1 score equal. The formula for the F1 measure is:

$$F1 \quad measure = (2 * Precision * Recall)/(Precision + Recall) \tag{4.9}$$

In Table 4.4 we review and compare the performance of different shopping behavior action classification using our proposed GRU network.

Table 4.4: Performance evaluation of shopping behavior classification.

Shopping behavior	Precision	Recall	F1 Score	Support
Reach to shelf	0.86	0.92	0.89	128
Inspect shelf	0.93	0.90	0.91	411
Hand in shelf	0.79	0.83	0.81	336
Retract from shelf	0.44	0.69	0.54	109
Inspect product	0.92	0.73	0.81	377
Avg/Total	**0.79**	**0.82**	**0.79**	**1361**

Figure 4.7: Confusion matrix of different shopping behavior.

Figure 4.7 shows the confusion matrix for shopping behavior action classification of our proposed network. It can be found that only 243 samples are misclassified out of 1361 samples, which means our accuracy is 82.1%. Hand in shelf and inspect product are less discernible from retract from shelf in this case.

4.6.2 Evaluation of Shopping Support Robot

The prediction time of recognizing head orientation, body orientation and shopping behavior recognition is five frames per second, which means 200 ms per frame. Using this processing speed, we evaluate our shopping support robot in two ways. In the first case, our shopping support robot is on the right side of the customer. In the second case, our shopping support robot is on the left side of the customer.

Figure 4.8: Evaluation of shopping support robot.

For the first case, in the first frame of Figure 4.8 we see that our shopping support robot observes the customer inspecting the products with 45° head and body orientation from a distance. In the second frame, we see that the customer's head and body orientation is 0° with respect to the shopping support robot. In this situation, our shopping support robot decides to move closer to the customer and change its orientation to a suitable position so that the customer can easily put his product in the shopping basket. The last frame shows that the customer is putting his product in the basket.

Figure 4.9: Evaluation of shopping support robot.

For the second case, the procedure is similar to the first case except the head and body orientations are different while inspecting the product. In the first frame of Figure 4.9 we see the customer inspecting the product with 270° head and body orientation. After inspecting the product, we see the customer looking towards the robot and his head and body orientation are both 0° with respect to the shopping support robot. Then the robot decides to move close to the customer and assumes the proper orientation so that the customer can put his product into the shopping basket.

In this way, our shopping support robot provides proper support to the customer by carrying his shopping product and following the customer until he or she is finished shopping.

4.7 Chapter Summary

Only person following and body orientation is not sufficient enough for supporting elderly in shopping mall. This chapter introduces GRU network based targeted person's shopping behavior recognition. According to the shopping behavior recognition our robot gives proper support to the customer.

Chapter 5

Person-Following Shopping Support Robot using Kinect Depth Camera based on 3D Skeleton Tracking

5.1 Introduction

With the advancement of robotics technologies researchers have started to explore the application of service robots to our daily life. A shopping mall is also one of them. Many researchers develop robotic shopping trolley Y. Kobayashi et al. [25] and shows the benefits of supporting the elderly. But if the robotic shopping trolley intelligently follows the customer by understanding the behavior recognition it can more helpful. By understanding the behavior, it can take an appropriate position behind the customer. In our previous paper [28] we developed an intelligent shopping support robot that can understand customer shopping behavior using our developed GRU (Gated Recurrent Unit) neural network. To develop these robots, we used OpenPose [24] neural network model to detect a person's skeleton for person tracking and the LiDAR sensor to measure the distance from robot to person. But in a practical situa-tion, this robot is not so appropriate because of its fastness.

To operate our robot in a practical environment we must ensure three requirements such as speed, accuracy and cost. OpenPose based our previous model does not ful-fill these requirements. For this reason, in this paper, we replace OpenPose model with Kinect v2 depth camera [29] to get the following advantages:

1. The processing speed of OpenPose model-based skeleton tracking is not so fast (5 frames/sec) whereas Kinect based skeleton tracking is so fast (30 frames/sec). So, for real time application Kinect based system is better.

2. By integrating GRU network with OpenPose model based skleton tracking the overall processing speed is approximately 1 frame/sec where as integrating GRU network with Kinect model based skeleton tracking the overall processing speed is approximately 25 frames/sec.
3. The accuracy of shopping behavior recognition using OpenPose model- based 2D skeleton data is not so good as compared to Kinect 3D skeleton data-based model. For OpenPose model we get 82% accuracy whereas using Kinect 3D skeleton based model we get 95% accuracy.
4. In our previous model, we used the LiDAR sensor to measure the distance between robot and customer where Kinect itself can measure the distance. So, it reduced the extra processing and cost.

Figure 5.1 shows our proposed robot. In our proposed system, we used a Kinect camera to find the location of the customer in a shopping mall and measure the distance from the robot to the customer so that it can easily follow the customer. We recognize the customer shopping behavior using 3D skeleton data and according to the recognized hand in shelf action, our robot takes proper position to the customer for putting a product on the shopping basket.

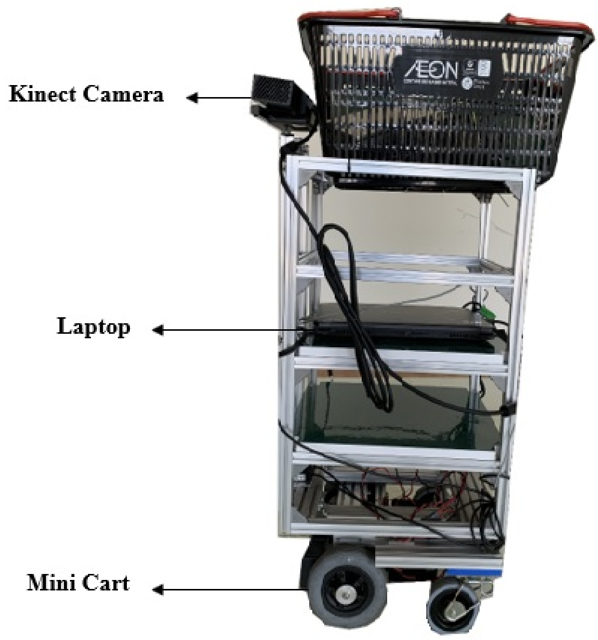


Figure 5.1: Our person following a mini cart robot.

5.2 Design Approach

We developed our system using the Xbox One Kinect V2 camera. Using this camera first we track the customer's 3D skeleton and recognize the customer's different shopping actions in front of the shelf using the GRU network. At the same time, we calculate the robot to customer distance using the Z-coordinate value of the middle of the spine. If the distance is less than 1.5 meters our robot performs action recognition otherwise it just follows the customer. If our robot recognizes a "hand in shelf" action by the customer it takes the proper position and helps the customer. Figure 5.2 shows the block diagram of our shopping support robot.

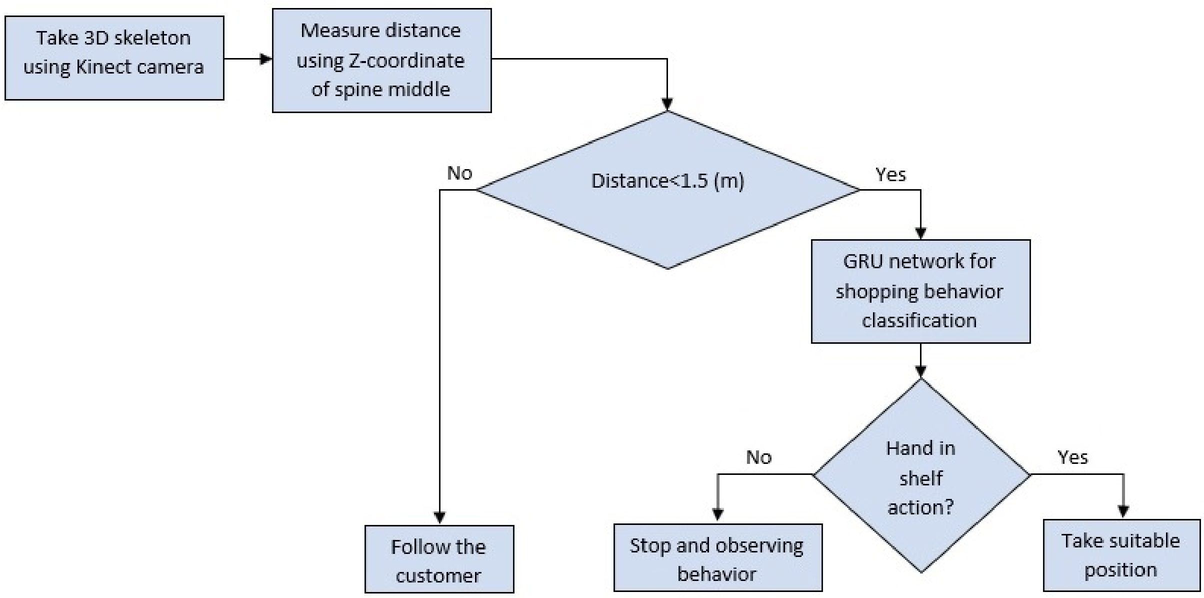


Figure 5.2: Block diagram of our proposed shopping support robot.

5.2.1 Person's Skeleton Tracking

In our system, we use the Kinect Version 2 camera to detect the 3D skeleton of a person. The skeleton has 25 joints as shown in Figure 5.3.

To find the 3D points of the joints in space the Kinect's camera coordinates use the Kinect's infrared sensor. The coordinate system is defined as according to Figure 5.4.

The origin(x=0,y=0,z=0) is located at the center of the IR sensor on Kinect.
Positive X values tend towards to the sensor's left [from the sensor's POV]
Y moves in an upward direction (note that this direction is based on the sensor's tilt)

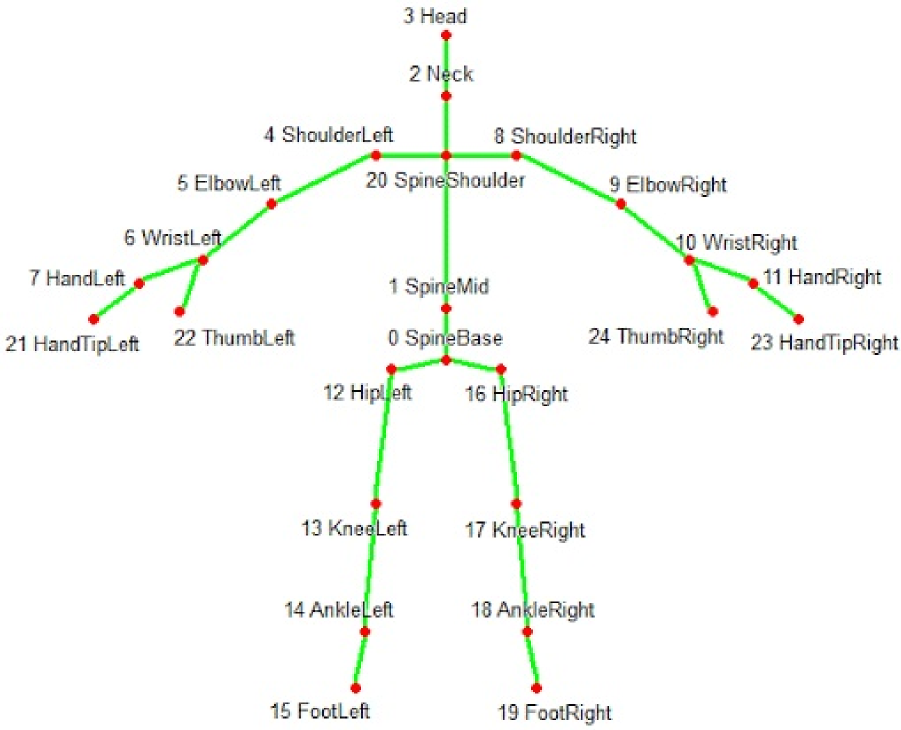


Figure 5.3: Kinect V2 25-joint skeleton [43].

Z moves out in the direction the sensor is facing
1 unit = 1 meter

Figure 5.4: Camera space coordinates from the Kinect SDK [44].

5.2.2 Person Following Procedure of Our Robot

When we run our system, the Kinect camera track the 3d skeleton of a person who is in front of the camera.

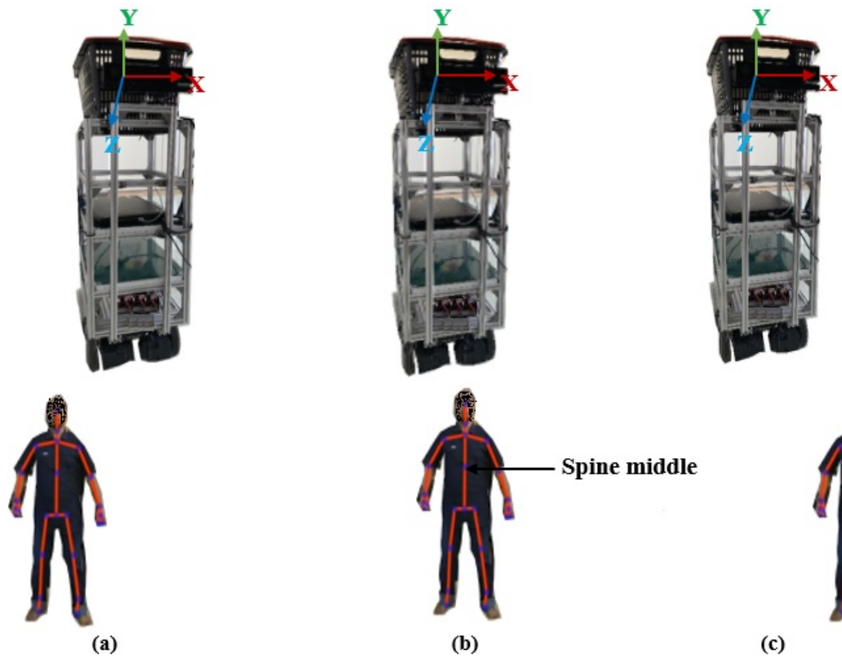


Figure 5.5: A person following procedure.

To follow the customer, we take two values from detected skeleton joint points, sm.X and sm.Z, where sm.X is the “mid-spine” value of the X coordinate that represents the position of the person and sm.Z is the mid-spine value of the Z coordinate and represents the distance the robot to customer. Depending on this value, the appropriate command issued to our robot is given in Table 5.1.

Table 5.1: Different command conditions for our robot.

Command	position
Forward	0.218678<sm.X<-0.179835
Stop	sm.Z<1.5 (m)
Left	sm.X<-0.179835
Right	sm.X>0.218678

We set the threshold value for the mid-spine value of X in the range (-0.179835 to 0.218678). If the tracked person’s mid-spine value of X is in this range, we assume that the person’s position is in the middle of the robot as shown in Figure 5.5(b). In this situation, the robot follows the person measuring the Z coordinate value of the mid-spine. If the mid-spine value of X>0.218678, the robot assumes that the person is on the right side of the robot as shown in Figure 5.5(c) and the robot then rotates right until it reaches the threshold value. When it reaches the threshold value, it again follows the person measuring the Z value of the mid-spine. If the mid-spine value of X<-0.179835, the robot assumes that the person is on the left side of the robot as shown in Figure 5.5(a) and the robot then rotates left until it reaches the threshold value. When it reaches the threshold value, it again follows the person measuring the Z value of the mid-spine. In this way, our person following robot works using a Kinect camera.

5.2.3 Shopping Behavior Action Recognition

5.2.3.1 Dataset Construction

We make a dataset to train our GRU network. The details of our GRU network are shown in our previous paper [28]. We take 3D joints of the skeleton to make the data-set for different shopping behaviors. We take 4 camera views to take the 3D skeleton joints data. We take 114,464 joints data for training and 33,984 joints data for the testing set.

5.2.3.2 Training the GRU Network

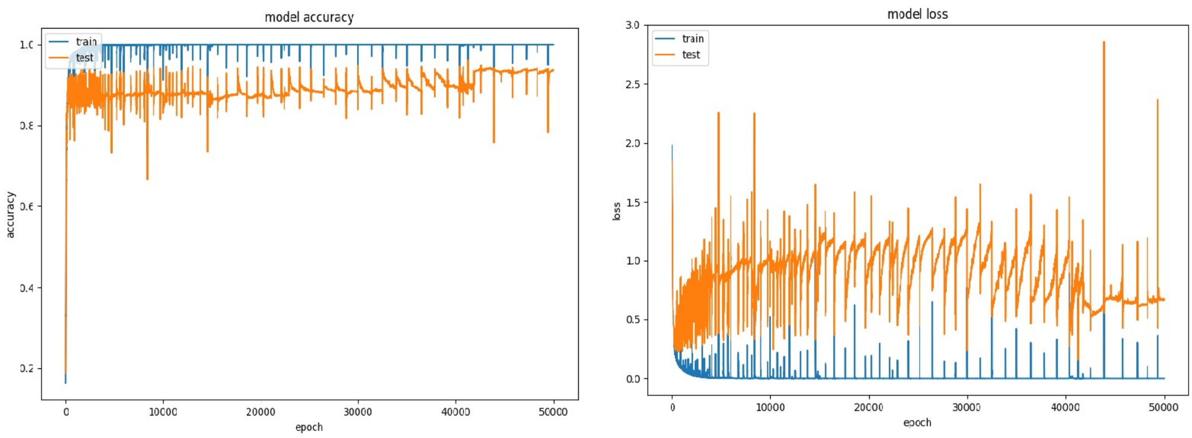


Figure 5.6: The model accuracy and loss over 50000 iterations.

The details of the training specification are shown in Table 5.2. Figure 5.6 shows the plot of the model accuracy and loss over 50000 iterations. The results of the GRU network-based shopping behavior recognition OpenPose 2D skeleton data compared to the Kinect 3D skeleton data is shown in Table 5.3.

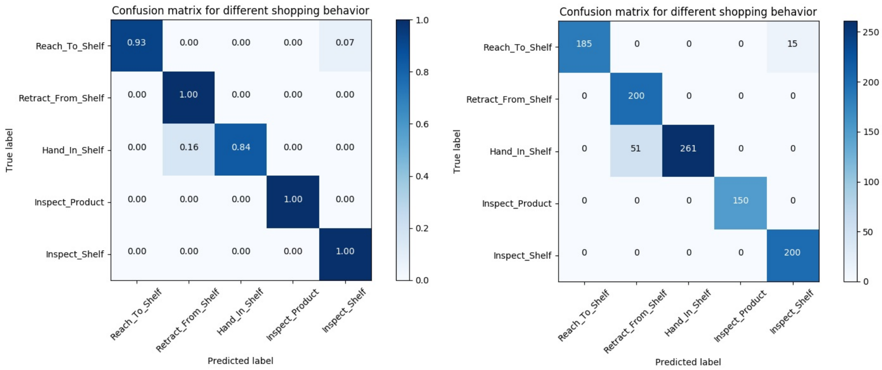


Figure 5.7: Shows the confusion matrix of different shopping behavior.

Table 5.2: Training specification for our proposed GRU network.

Training parameters	value
Batch size	512
Epochs	50000
Timesteps	32
No. of hidden layer	34
Learning rate	0.000220
Optimizer	adam
Momentum	0.9

5.3 Experiments

We experimented at an actual supermarket. The area was arranged with a two-sided shelf with different items. A person moved between the shelves and our robot followed that person. When the person performed a "hand in shelf" action, our robot took a suitable position so that the person could easily put his items in the basket.

Table 5.3: Performance evaluation of shopping behavior classification.

Shopping behavior	Precision	Recall	F1 Score	Support
Reach to shelf	1.00	0.93	0.96	200
Inspect shelf	0.93	1.00	0.96	200
Hand in shelf	1.00	0.84	0.91	312
Retract from shelf	0.80	1.00	0.89	200
Inspect product	1.00	1.00	1.00	150
Avg/Total	**0.95**	**0.94**	**0.94**	**1062**

5.3.1 Experimental Conditions

We conducted our experiment in four conditions. We define the different conditions below: In the first condition, we assume that the robot is on the left side and the customer turns to the front shelf and makes a hand in shelf action as shown in Figure 5.8(a).

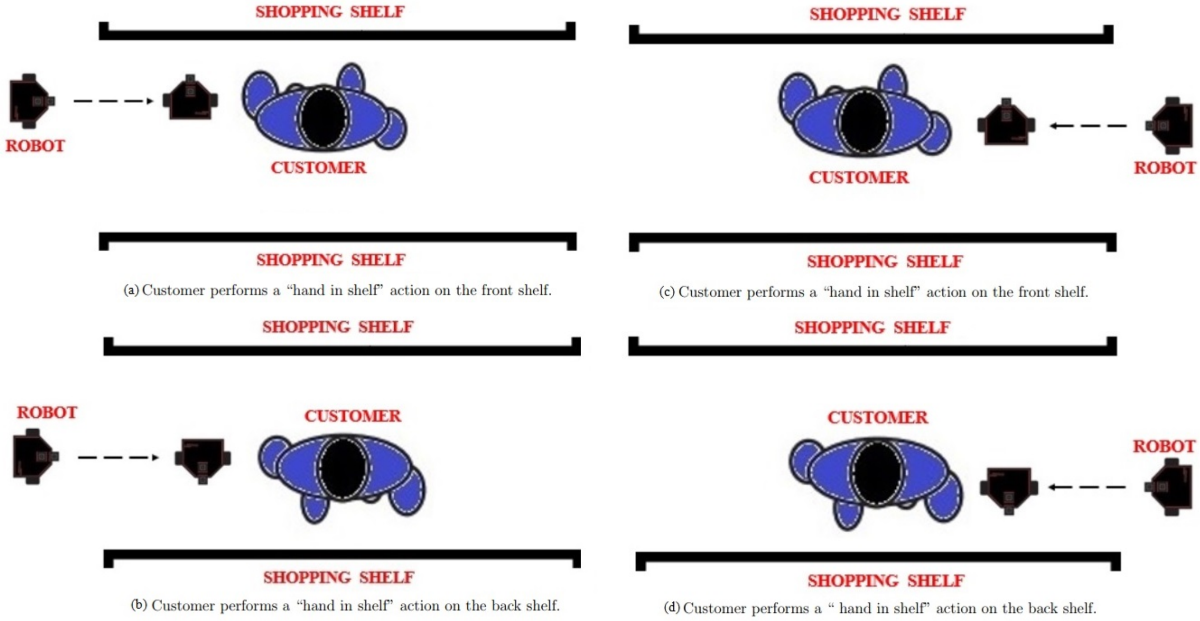


Figure 5.8: Different experimental conditions.

In the second condition, we assume that the robot is on the left side and the customer turns to the back shelf and performs a “hand in shelf” action as shown in Figure 5.8(b).

In the third condition, we assume that the robot is on the right side and the customer turns to the front shelf and performs a “hand in shelf” action as shown in Figure 5.8(c).

In the fourth condition, we assume that the robot is on the right side and the customer turns to the back shelf and performs a “hand in shelf” action as shown in Figure 5.8(d).

5.3.2 Experimental Results

Figure 5.9 shows the experimental result according to our first condition. In Figure 5.9(a) we see that our robot follows a customer with a shopping basket within a certain distance. When the customer performs a “hand in shelf” action on the front shelf, we see that the robot is on the left side of the customer as shown in Figure 5.9(b). Then our robot moves closer to the customer and changes its orientation according to the customer and the customer easily puts his product in the basket.

(a) (b) (c)

Figure 5.9: Experimental results according to the first condition.

5.4 Chapter Summary

In this chapter, we have developed a customer-following robot in the shopping mall to give support for an elderly person with high-speed vision systems by using Kinect 3D skeleton data. The robot can follow a person (i.e. customer in the shopping mall) successfully and at the same time, our robot always tries to recognize the customer’s different shopping actions. When it gets the hand in shelf action, it can take a suitable position to give the proper support of the elderly person with the shopping basket. Our robot gives support by carrying the goods of the customer until he or she finishes the shopping.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

In this thesis we presented a complete work, started from scratch on how our robot follow a target person and gives proper support. we address the design considerations for an intelligent shopping support robot. One of the objectives of the work was to develop a simple, reliable and easy to use system that could provide freedom of movement for elderly and handicapped people. We provide shopping support facilities for the elderly. We have confirmed that our vision system can understand the shopping behaviors necessary for supporting the user and developed a robot system.

6.2 Future Work

In this dissertation, attempts has been taken to solve a highly complex and unconventional design problem in HRI-designing humans' shopping behaviors tracking system to estimate their interests, intentions, and preferences concerning surrounding environments incorporation with a psychology, sociology, cognitive science and human-robot interaction fields. The ultimate goal is to give a better support in shopping mall for the elderly person. There are still several issues that have not addressed in the current model. Some of the issues are discuss in the following.

• The main drawback of color histograms for classification is that the representation is dependent of the color of the object. In future we replace color histogram matching based person tracking system with any other good person tracking algorithm.

• As future works, we need to analyze the customer behavior with different postures, such as bending over or squatting down. Also, to solve the similarity problem of touching and picking and putting in our system.

- In the future, we plan to propose a mapping system to more effectively prevent the shopping support robot from colliding with obstacles and to move in a crowded environment so that it can follow the customer automatically in any direction.

Publication List

[1] Islam, M.M., Lam, A., Fukuda, H., Kobayashi, Y. and Kuno, Y.: A person-following shopping support robot based on human pose skeleton data and LiDAR sensor. In: International Conference on Intelligent Computing (ICIC), Springer LNCS11645, pp. 9-19 (2019).

[2] Islam, M.M., Lam, A., Fukuda, H., Kobayashi, Y. and Kuno, Y.: An intelligent shopping support robot: understanding shopping behavior from 2D skeleton data using GRU network. In: ROBOMECH Journal, 6(1), 18 pages (2019).

[3] Islam, M.M., Lam, A., Fukuda, H., Kobayashi, Y. and Kuno, Y.: Person following shopping support robot using kinect depth camera based on 3D skeleton tracking. In: International Conference on Intelligent Computing (ICIC), Springer LNCS, 11 pages (2020). (to appear)

Bibliography

[1] Beymer, D. and Konolige, K.: Tracking people from a mobile platform. In: Experimental Robotics VIII, Springer, Berlin, Heidelberg, pp. 234-244 (2003).

[2] Calisi, D., Iocchi, L. and Leone, R.: Person following through appearance models and stereo vision using a mobile robot. In: Workshop on Robot Vision (VISAPP), pp. 46-56 (2007).

[3] Sonoura, T., Nakamoto, H., Nishiyama, M., Matsuhira, N., Tokura, S. and Yoshimi, T.: Person following robot with vision-based and sensor fusion tracking algorithm. In: Computer Vision, Book edited by: XiongZhihui, pp. 519-538 (2008).

[4] Ilias, B., Nagarajan, R., Murugappan, M., Helmy, K., Awang Omar, A.S. and Abdul Rahman, M.A.: Hospital nurse following robot: Hardware development and sensor integration. In: International Journal of Medical Engineering and Informatics, 6(1), pp. 1-13 (2014).

[5] Hiroi, Y., Matsunaka, S. and Ito, A.: A mobile robot system with semi-autonomous navigation using simple and robust person following behavior. In: Journal of Man, Machine and Technology (JMMT), 1(1), pp. 44-62 (2012).

[6] Germa, T., Lerasle, F., Ouadah, N., Cadenat, V. and Devy, M.: Vision and RFID-based person tracking in crowds from a mobile robot. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 5591-5596 (2009).

[7] Nishimura, S., Takemura, H. and Mizoguchi, H.: Development of attachable modules for robotizing daily items-Person following shopping cart robot. In: IEEE International Conference on Robotics and Biomimetics (ROBIO), New York, pp. 1506–1511 (2007).

[8] Kohtsuka, T., Onozato, T., Tamura, H., Katayama, S. and Kambayashi Y.: Design of a control system for robot shopping carts. In: International Conference on

Knowledge-based and Intelligent Information and Engineering Systems, Springer, Berlin, Heidelberg, pp. 280–288 (2011).

[9] Hu, Y., Cao, L., Lv, F., Yan, S., Gong Y., and Huang TS.: Action detection in complex scenes with spatial and temporal ambiguities. In: IEEE 12th International Conference on Computer Vision, pp. 128–135 (2009).

[10] Haritaoglu, I., Beymer, D., and Flickner, M.: Ghost 3d: detecting body posture and parts using stereo. In: IEEE Workshop on Motion and Video Computing, pp. 175–180 (2002).

[11] Lao, W., Han, J., and De With, P.H.: Automatic video-based human motion analyzer for consumer surveillance system. In: IEEE Trans. on Consumer Electron, 55(2), pp. 591-598 (2009).

[12] Iwamura, Y., Shiomi, M., Kanda, T., Ishiguro, H. and Hagita, N.: Do elderly people prefer a conversational humanoid as a shopping assistant partner in supermarkets? In: Human-Robot Interaction (HRI), pp. 449-456 (2011).

[13] Garcia-Arroyo, M., Marin-Urias, L.F., Marin-Hernandez A. and Hoyos-Rivera, G.D.J.: Design, integration and test of a shopping assistance robot system. In: Human-Robot Interaction (HRI), pp. 135–136 (2012).

[14] Marin-Hernandez, A., de Jesús Hoyos-Rivera, G., Garcia-Arroyo, M. and Marin-Urias, L.F.: Conception and implementation of a supermarket shopping assistant system. In: 11th IEEE Mexican International Conference on Artificial Intelligence, New York, pp. 26–31 (2012).

[15] Zimmerman, T.G.: Tracking shopping carts using mobile cameras viewing ceiling-mounted retro-reflective bar codes. In: 4th IEEE International Conference on Computer Vision Systems (ICVS'06), New York, pp. 36-36 (2006).

[16] Haritaoglu, I. and Flickner, M.,: Detection and tracking of shopping groups in stores. In: Computer Vision and Pattern Recognition (CVPR), Vol.1, pp.I-I (2001).

[17] Leykin, A., and Tuceryan, M.: Detecting shopper groups in video sequences. In: IEEE Conference on Advanced Video and Signal based Surveillance, New York, pp. 417–422 (2007).

[18] Senior, A.W., Brown, L., Hampapur, A., Shu, C.F., Zhai, Y., Feris, R.S., Tian, Y.L., Borger, S. and Carlson, C.: Video analytics for retail. In: IEEE Conference on Advanced Video and Signal Based Surveillance, New York, pp. 423–428 (2007).

[19] Haritaoglu, I. and Flickner, M.: Attentive billboards: towards to video based customer behavior understanding. In: 6th IEEE Workshop on Applications of Computer Vision, pp. 127-131 (2002).

[20] Moeslund, T.B.: Is there anybody out there?. In: Visual Analysis of Humans, Springer, London, pp. 3-9 (2011).

[21] Ahmad, M.F., Alhady, S.S.N., Kaharuddin, S. and Othman, W.A.F.W.: Visual based sensor cart follower for wheelchair by using microcontroller. In: IEEE International Conference on Control System, Computing and Engineering (ICCSCE), pp. 123-128 (2015).

[22] Rastogi, S., Agarwal, V. and Singh, P.: An intelligent sensing follower cart (automatic shopping trolley). In: Journal of Controller and Converters, 2(2) (2017).

[23] Ng, Y.L., Lim, C.S., Danapalasingam, K.A., Tan, M.L.P. and Tan, C.W.: Automatic human guided shopping trolley with smart shopping system. In: Jurnal Teknologi, 73(3) (2015).

[24] Cao, Z., Simon, T., Wei, S.E. and Sheikh, Y.: Realtime multi-person 2d pose estimation using part affinity fields. In: Computer Vision and Pattern Recognition (CVPR), pp. 7291-7299 (2017).

[25] Kobayashi Y., Yamazaki S., Takahashi H., Fukuda H. and Kuno Y.: Robotic shopping trolley for supporting the elderly. In: International Conference on Applied Human Factors and Ergonomics. Springer, Cham, pp. 344–353 (2018).

[26] Islam, M.M., Lam, A., Fukuda, H., Kobayashi, Y. and Kuno, Y.: A person-following shopping support robot based on human pose skeleton data and LiDAR sensor. In International Conference on Intelligent Computing (ICIC), Springer LNCS11645, pp.9-19 (2019).

[27] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H. and Bengio, Y.: Learning phrase representations using RNN encoderdecoder for statistical machine translation. arXiv preprint arXiv:1406.1078 (2014).

[28] Islam, M.M., Lam, A., Fukuda, H., Kobayashi, Y. and Kuno, Y.: An intelligent shopping support robot: understanding shopping behavior from 2D skeleton data using GRU network. In: ROBOMECH Journal, 6(1), 18 pages (2019).

[29] Microsoft Kinect SDK [Online]. Available: http://www.microsoft.com/en-us/kinectforwindows/.

[30] Swain, M.J. and Ballard, D.H.: Color indexing. In: International Journal of Computer Vision, 7(1), pp. 11-32 (1991).

[31] Sales, J., Martí, J.V., Marín, R., Cervera, E. and Sanz, P.J.: CompaRob: the shopping cart assistance robot. In: International Journal of Distributed Sensor Networks, 12(2):4781280 (2016).

[32] Bouraqadi, N., Fabresse, L., Laval, J. and Bragagnolo, S.: RoboShop Project Delivrable 1 PHAROS: Concepts and Architecture.

[33] Takahashi, K., Seki, S., Kojima, E. and Oka, R.: November. Recognition of dexterous manipulations from time-varying images. In: IEEE Workshop on Motion of Non-rigid and Articulated Objects, pp. 23-28 (1994).

[34] Bremond, F. and Medioni, G.: Scenario recognition in airborne video imagery. In: DARPA Image Understanding Workshop, pp. 211-216 (2108).

[35] Oliver, N.M., Rosario, B. and Pentland, A.P.: A Bayesian computer vision system for modeling human interactions. In: IEEE Trans. on Pattern Analysis and Machine Intelligence, 22(8), pp. 831-843 (2000).

[36] Meier, U., Stiefelhagen, R., Yang, J. and Waibel, A.: Towards unrestricted lip reading. In: International Journal of Pattern Recognition and Artificial Intelligence, 14(05), pp. 571-585 (2000).

[37] Hu, W., Xie, D. and Tan, T.: A hierarchical self-organizing approach for learning the patterns of motion trajectories. In: IEEE Trans. on Neural Networks, 15(1), pp. 135-144 (2004).

[38] Chang, C.C. and Lin, C.J.: A library for support vector machines. In: ACM Trans. on Intelligent Systems and Technology (TIST), 2(3), pp. 1-27 (2011).

[39] Liu, J., Gu, Y. and Kamijo, S.: Customer behavior classification using surveillance camera for marketing. In: Multimedia Tools and Applications, 76(5), pp. 6595-6622 (2017).

[40] Popa, M., Rothkrantz, L., Yang, Z., Wiggers, P., Braspenning, R. and Shan, C.: Analysis of shopping behavior based on surveillance system. In: IEEE International Conference on Systems, Man and Cybernetics, pp. 2512-2519 (2010).

[41] Popa, M., Koc, A.K., Rothkrantz, L.J., Shan, C. and Wiggers, P.: Kinect sensing of shopping related actions. In: International Joint Conference on Ambient Intelligence . Springer, Berlin, Heidelberg, pp. 91-100 (2011).

[42] https://towardsdatascience.com/$illustrated-guide-to-lstms-and-gru-s-a-step-by-step-explanation-44e9eb85bf21$

[43] https://www.researchgate.net/figure/$The-25-body-joints-Kinect-v2-can-capture$_fig2_334778881

[44] https://homes.cs.washington.edu/ edzhang/tutorials/kinect2/kinect3.html?fbclid =IwAR0DFvN6LQPkLCs674423mUeQN0ZQDQiMXbmvOkG3mem8HW23ktuwz Wzbcw