

RESEARCH ARTICLE

Open Access



An intelligent shopping support robot: understanding shopping behavior from 2D skeleton data using GRU network

Md Matiquel Islam^{1,2*} , Antony Lam^{3†}, Hisato Fukuda^{1†}, Yoshinori Kobayashi^{1†} and Yoshinori Kuno^{1†}

Abstract

In supermarkets or grocery, a shopping cart is a necessary tool for shopping. In this paper, we have developed 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.

Keywords: GRU, Shopping behavior, OpenPose, Head orientation, Body orientation, DNN

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 [1], 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 [2], we developed an autonomous person following robot that can follow a given target person in crowded areas.

*Correspondence: matiquel@cv.ics.saitama-u.ac.jp

†Matiquel Islam, Antony Lam, Hisato Fukuda, Yoshinori Kobayashi and Yoshinori Kuno are equally contributed to this article

¹ Department of Information and Computer Sciences, Saitama University, 255 Shimo-Okubo, Sakura-ku, Saitama 338-8570, Japan

Full list of author information is available at the end of the article

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 [3] 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.

Related work

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. [4]. This shopping cart can follow customers autonomously and transports the goods. Kohtsuka et al. [5] 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 [6] concludes that elderly people interact in a better way with robots carrying the shopping basket and providing conversational facilities. In [7, 8] 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 [9]. 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. [10] described a system for counting shopping groups waiting in checkout lanes. Leykin et al. [11] 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. [12]. 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. [13] 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 [14]. Lao et al. [15] recognize customer's actions, such as pointing, squatting and raising hand using one surveillance camera. Haritaoglu et al. [16] extracted customer behavior information whenever they watched advertisements on a billboard or a new product promotion.

In this paper, we propose to combine the research on autonomous shopping cart robots and that on shopping behavior recognition to realize a shopping support robot. The robot can act in advance to meet the user's next move based on the user's behavior recognition results.

Customer behavior model for the front of the shelf

In our previous work [2], we developed a person following shopping support robot. In this paper, we focus on more intelligent shopping support robot that can recognize the customer's shopping behavior and act accordingly.

Definition of customer behavior model

Our customer behavior model captures indications of increasing interest that the customer has towards the

Table 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
Retract from shelf	If customer picked a product from shelf
Hand in shelf	If a customer is going to pick a product
Inspect product	If a customer inspects the product
Inspect shelf	If customer inspects the whole shelf

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 1. Figure 1 shows some examples of these behaviors.

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 GRU neural network is used for classification of shopping behavior. The system’s framework is shown in Fig. 2.

Customer behavior classification

Head orientation detection

In this paper the head orientation of the customer is estimated in eight directions as shown in Fig. 3. 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 Fig. 3. 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 2.

Body orientation detection

Similar to the head orientation, the body orientation is also calculated in eight directions as shown in Fig. 4.

We take four angle values $\angle EAC$, $\angle ACE$, $\angle DBE$ and $\angle AEC$ of the target person to predict body orientations.

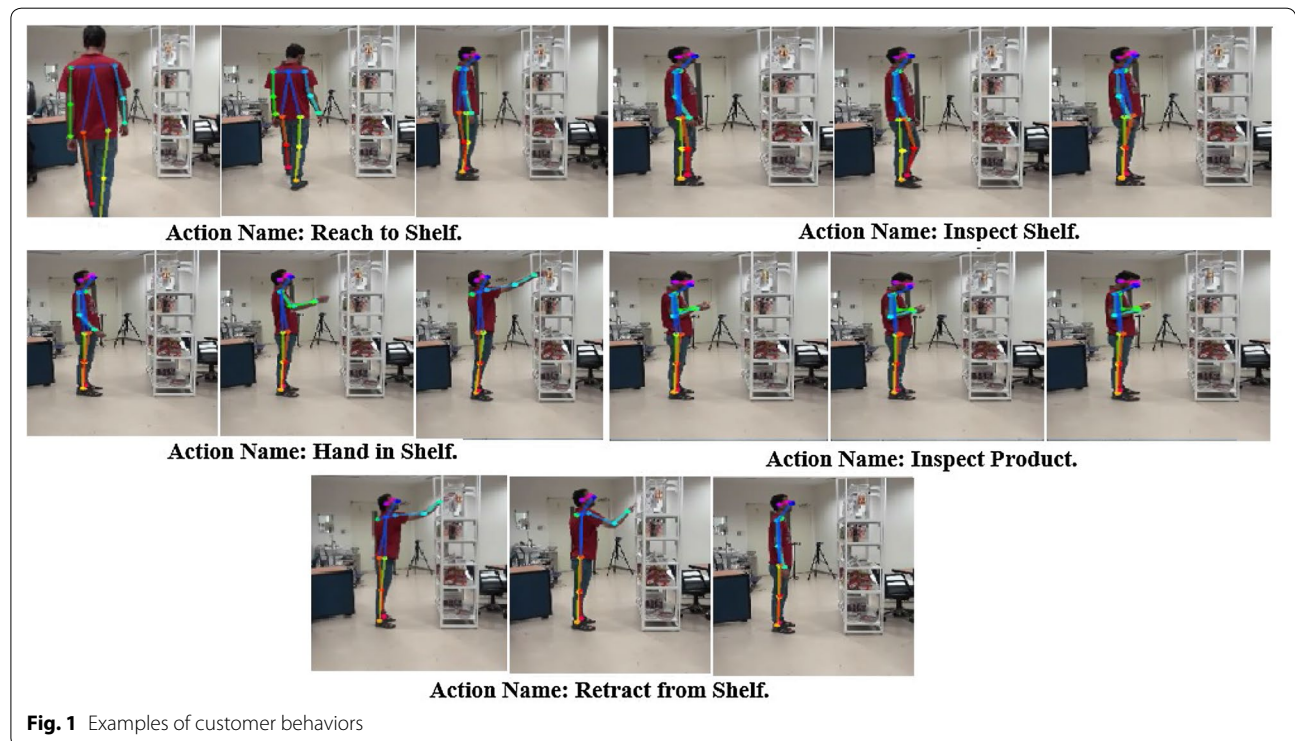


Fig. 1 Examples of customer behaviors

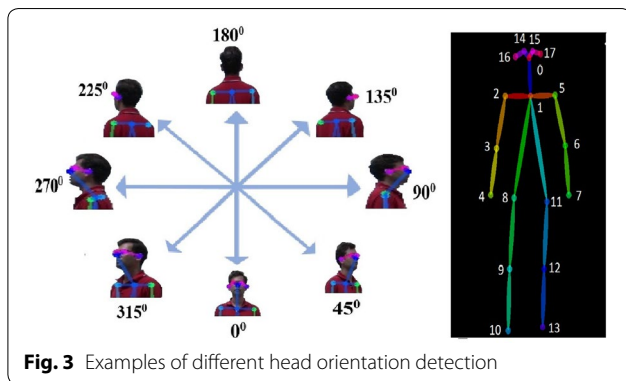
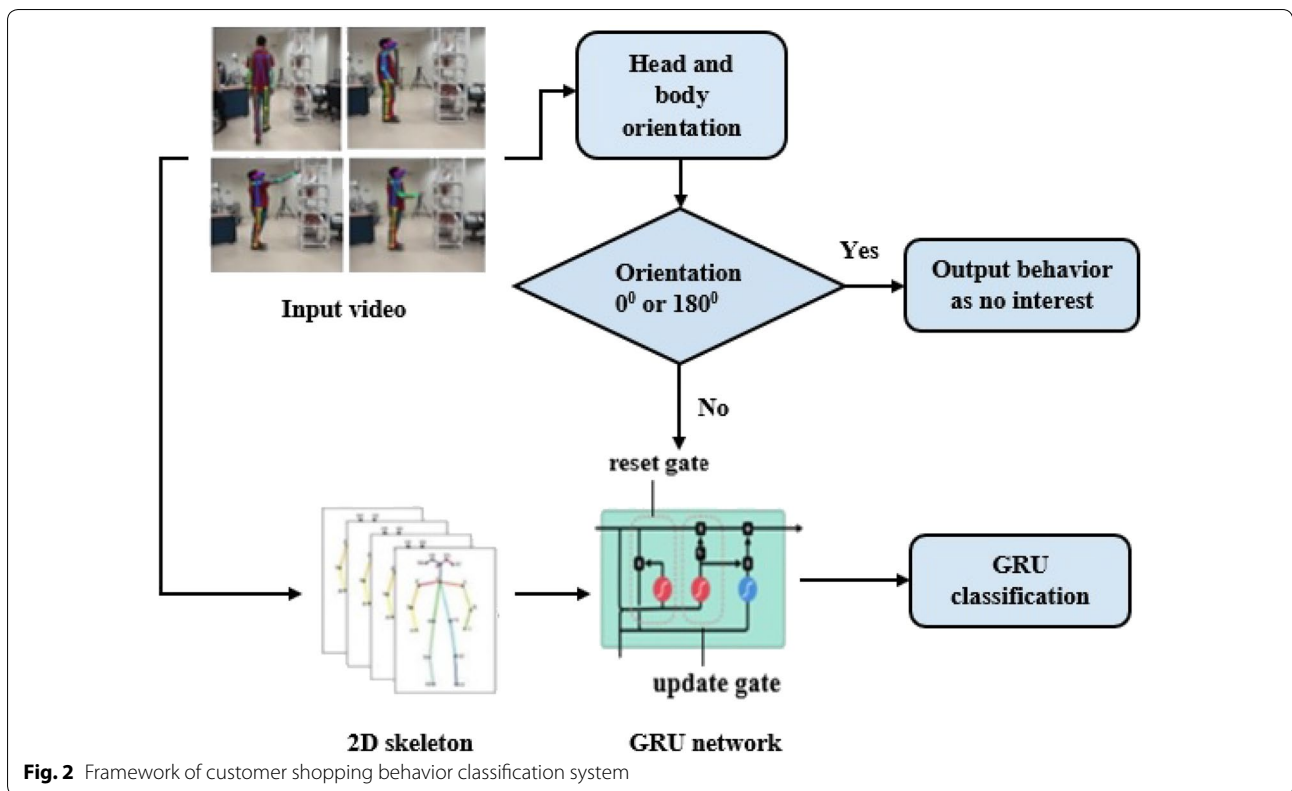
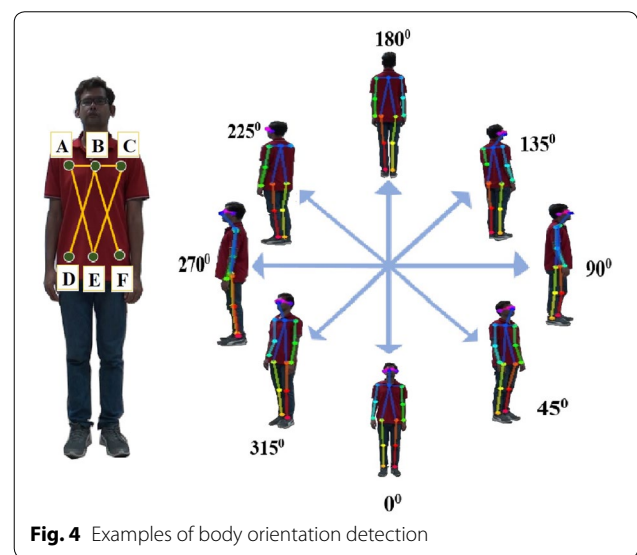


Table 2 Detected head joint points for classification of head orientation

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]



More details for body orientation detection is shown in our previous work [1].

For example Fig. 5 shows our head and body orientation detection results. It can be seen that our method clearly identifies different head and body orientations.

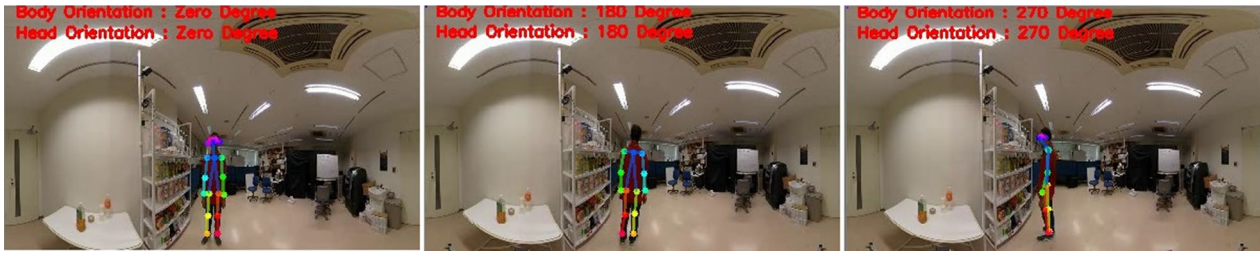


Fig. 5 Detection results of head and body orientation

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.

As shown in Fig. 6, 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 \tilde{h}_t^j :

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

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

$$z_t^j = \sigma(Wx_t + Uzh_{t-1})^j \tag{2}$$

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

$$\tilde{h}_t^j = \tanh(W_zx_t + U(r_t \odot h_{t-1}))^j \tag{3}$$

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:

$$r_t^j = \sigma(W_r x_t + U_r h_{t-1})^j \tag{4}$$

And the output is given by

$$y = \text{sigmoid}(W_y h_{32} + b_y) \tag{5}$$

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 Eq. (5). We use the sigmoid function as the network output activation function.

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

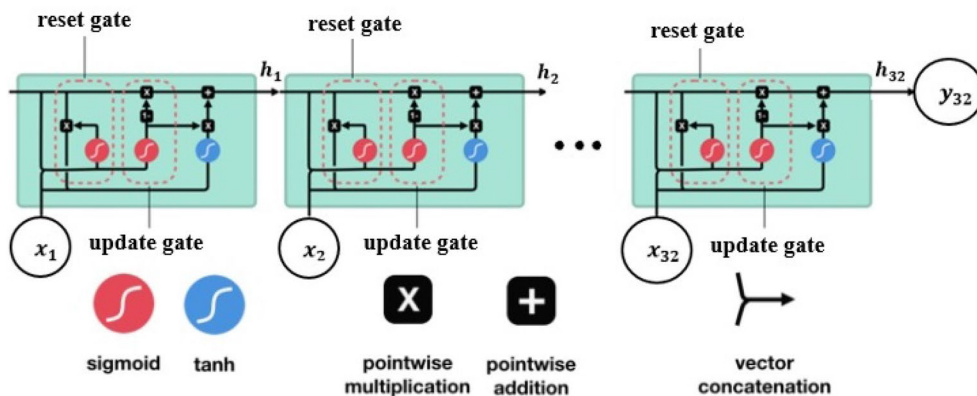


Fig. 6 Proposed GRU network for shopping behavior classification

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 $(20 \times 10 \times 4) 400$ min 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: $[j0_x, j0_y, j1_x, j1_y, j2_x, j2_y, j3_x, j3_y, j4_x, j4_y, j5_x, j5_y, j6_x, j6_y, j7_x, j7_y, j8_x, j8_y, j9_x, j9_y, j10_x, j10_y, j11_x, j11_y, j12_x, j12_y, j13_x, j13_y, j14_x, j14_y, j15_x, j15_y, j16_x, j16_y, j17_x, j17_y]$

Table 3 Training specification for our proposed GRU network

Training parameters	Value
Batch size	512
Epochs	50,000
Timesteps	32
No. of hidden layer	34
Learning rate	0.000220
Optimizer	Adam
Momentum	0.9

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 50,000 epochs. The training took around 5 h to finish. Other training specifications are given in Table 3.

Figure 7 shows the plot of the model's loss and accuracy over 50,000 iterations.

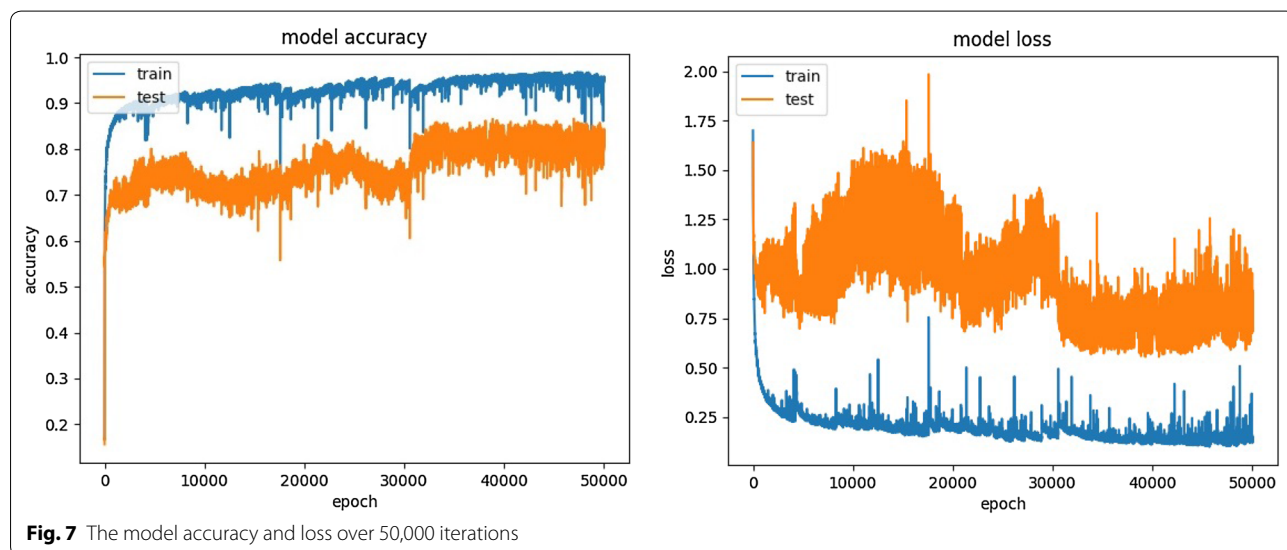
Table 4 shows the detailed layer information for our proposed GRU network structure. It has three layers. The first layer is GRU layer and it is the main layer containing two gates, a reset gate and an update gate. The second layer is a Dropout layer and it reduces overfitting. The last layer is a Dense layer and it is a fully connected layer.

Architecture of the shopping support robot based on the user's behavior recognition

Figure 8 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

Table 4 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



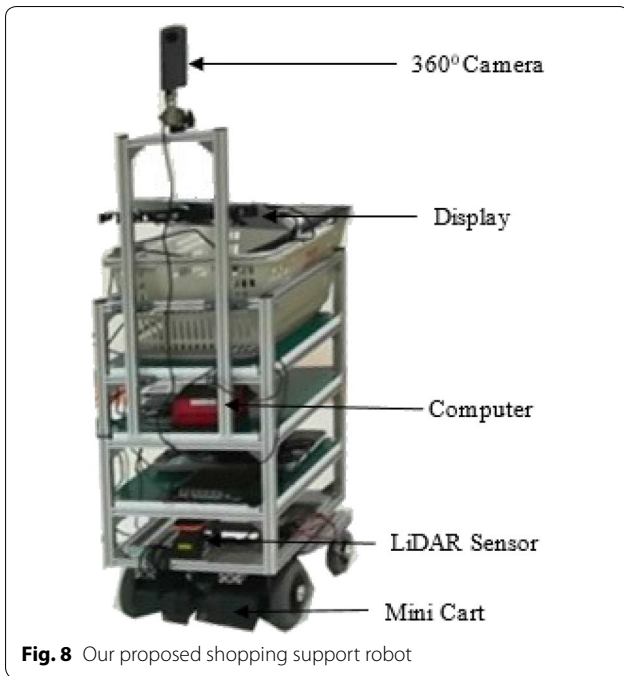


Fig. 8 Our proposed shopping support robot

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 9 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.

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

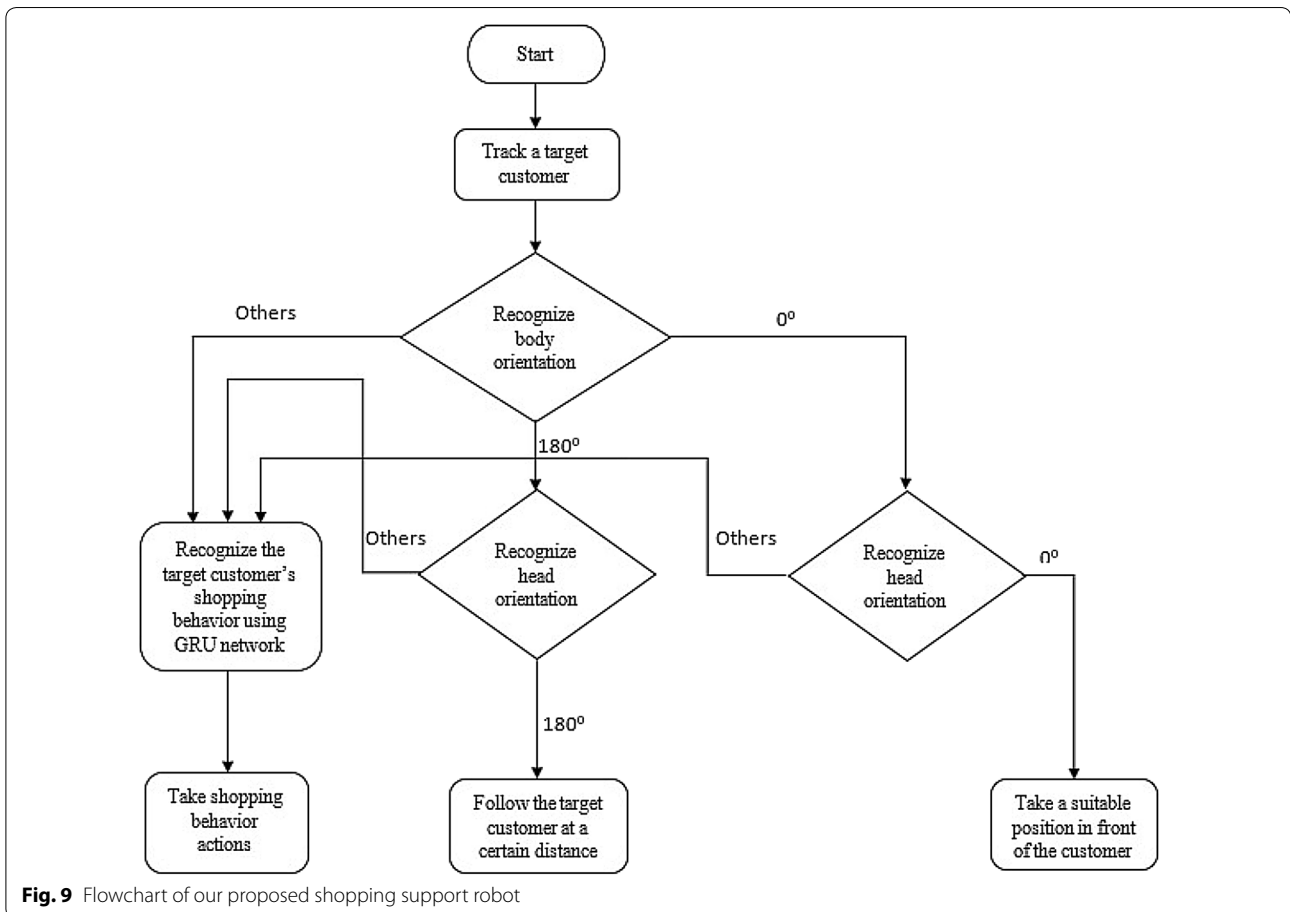


Fig. 9 Flowchart of our proposed shopping support robot

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

Results

Evaluation of behavior recognition

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 particular class.

$$Precision = TP / (TP + FP) \tag{6}$$

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).

Table 5 Performance evaluation of shopping behavior classification

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

$$Recall = TP / (TP + FN) \tag{7}$$

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

$$Over\ all\ accuracy = (TP + TN) / (TP + TN + FP + FN) \tag{8}$$

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:

$$F1\ measure = (2 * Precision * Recall) / (Precision + Recall) \tag{9}$$

In Table 5 we review and compare the performance of different shopping behavior action classification using our proposed GRU network.

Figure 10 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.

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

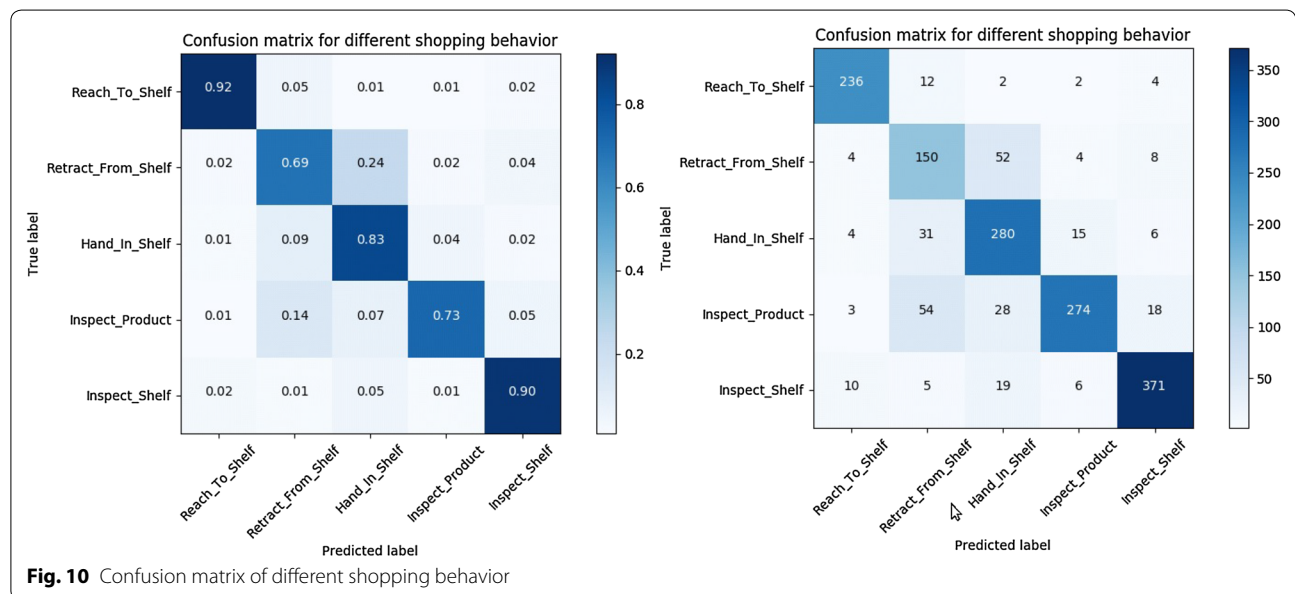


Fig. 10 Confusion matrix of different shopping behavior

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.

For the first case, in the first frame of Fig. 11 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.

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 Fig. 12 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.

Conclusion and future work

In this paper, 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. To do so, a person following robot was developed in our previous work [2]. In this work, 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. However, the current visual processing speed is not fast enough for the robot to move smoothly and to be used in practice. We will improve the processing speed and perform experiments in actual shopping situations to evaluate the total system. Then, based on the experimental results, we will further modify a robot system that can provide practical support for the elderly in shopping.

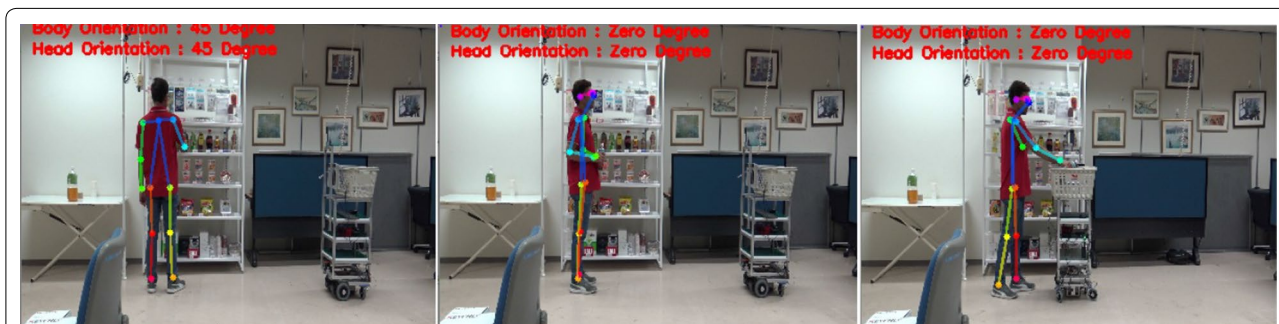


Fig. 11 Evaluation of shopping support robot

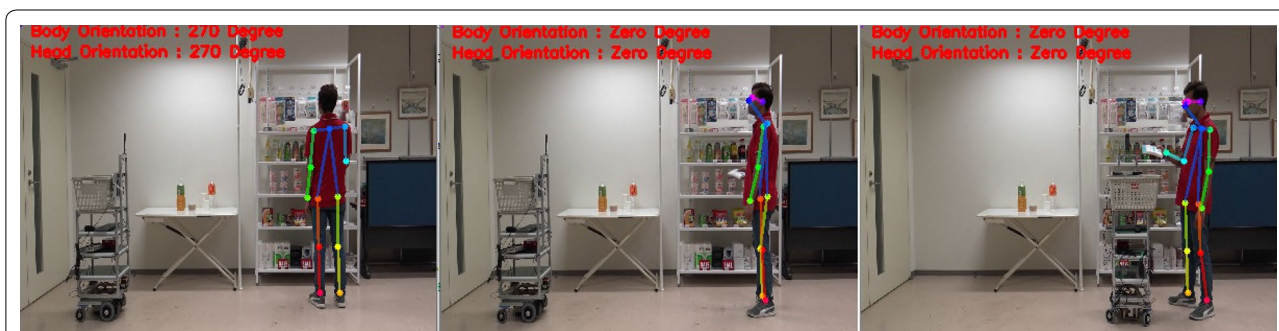


Fig. 12 Evaluation of shopping support robot

Acknowledgements

We thank JSPS for their research grant to make this research happened. We also thank all lab mates of Computer Vision Lab of Saitama University, Japan for their continuous support and encouragement over the period.

Authors' contributions

MMI designed the experiment, carried out all experiments, analyzed data and wrote the paper. AL was conducted while he was at Saitama University. AL, HF, YKo and YKu, initiated this project and advised the design of the experiments, analysis of data and paper writing. All authors read and approved the final manuscript.

Funding

This research was supported by a research Grant from JSPS KAKENHI Grant Number JP26240038.

Availability of data and materials

All data generated or analyzed during this study are included in this published article.

Competing interests

The authors declare that they have no competing interests.

Author details

¹ Department of Information and Computer Sciences, Saitama University, 255 Shimo-Okubo, Sakura-ku, Saitama 338-8570, Japan. ² Department of Information and Communication Engineering, University of Rajshahi, Rajshahi 6205, Bangladesh. ³ Mercari, Inc., Roppongi Hills Mori Tower 18F, 6-10-1 Roppongi Minato-ku, Tokyo 106-6118, Japan.

Received: 30 September 2019 Accepted: 6 December 2019

Published online: 16 December 2019

References

- Kobayashi Y, Yamazaki S, Takahashi H, Fukuda H, Kuno Y (2018) Robotic shopping trolley for supporting the elderly. In: International conference on applied human factors and ergonomics. Springer, Cham, pp 344–353
- Islam MM, Lam A, Fukuda H, Kobayashi Y, Kuno Y (2019) A person-following shopping support robot based on human pose skeleton data and LiDAR sensor. In: International conference on intelligent computing. Springer, Cham, pp 9–19
- Cho K, Van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, Bengio Y (2014) Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint [arXiv:1406.1078](https://arxiv.org/abs/1406.1078)
- Nishimura S, Takemura H, Mizoguchi H (2007) Development of attachable modules for robotizing daily items-Person following shopping cart robot. In: 2007 IEEE international conference on robotics and biomimetics (ROBIO), IEEE, New York, pp 1506–1511
- Kohtsuka T, Onozato T, Tamura H, Katayama S, Kambayashi Y (2011) 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
- Iwamura Y, Shiomi M, Kanda T, Ishiguro H, Hagita N (2011) Do elderly people prefer a conversational humanoid as a shopping assistant partner in supermarkets? In: Proceedings of the 6th international conference on human-robot interaction, ACM, New York, pp 449–456
- Garcia-Arroyo M, Marin-Urias LF, Marin-Hernandez A, Hoyos-Rivera GDJ (2012) Design, integration and test of a shopping assistance robot system. In: 2012 7th ACM/IEEE international conference on human-robot interaction (HRI), pp 135–136
- Marin-Hernandez A, de Jesus Hoyos-Rivera G, Garcia-Arroyo M, Marin-Urias LF (2012) Conception and implementation of a supermarket shopping assistant system. In: 2012 11th Mexican international conference on artificial intelligence, IEEE, New York, pp 26–31
- Zimmerman Thomas G (2006) Tracking shopping carts using mobile cameras viewing ceiling-mounted retro-reflective bar codes. In: Fourth IEEE international conference on computer vision systems (ICCVS'06), IEEE, New York, pp 36
- Haritaoglu I, Flickner M (2001) Detection and tracking of shopping groups in stores. In: Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition, vol 1. CVPR, IEEE, New York, pp 1
- Leykin A, Tuceryan M (2007) Detecting shopper groups in video sequences. In: 2007 IEEE conference on advanced video and signal based surveillance, IEEE, New York, pp 417–422
- Senior AW, Brown L, Hampapur A, Shu CF, Zhai Y, Feris RS, Tian YL, Borger S, Carlson C (2007) Video analytics for retail. In: 2007 IEEE conference on advanced video and signal based surveillance, IEEE, New York, pp 423–428
- Hu Y, Cao L, Lv F, Yan S, Gong Y, Huang TS (2009) Action detection in complex scenes with spatial and temporal ambiguities. In: 2009 IEEE 12th international conference on computer vision, IEEE, New York, pp 128–135
- Haritaoglu I, Beymer D, Flickner M (2002) Ghost/sup 3D: detecting body posture and parts using stereo. In: Workshop on motion and video computing, IEEE, New York, pp 175–180
- Lao W, Han J, De With PH (2009) Automatic video-based human motion analyzer for consumer surveillance system. IEEE Trans Consumer Electron 55(2):591–598
- Haritaoglu I, Flickner M (2002) Attentive billboards: towards to video based customer behavior understanding. In: Sixth IEEE workshop on applications of computer vision, proceedings, pp 127–131
- Cao Z, Simon T, Wei SE, Sheikh Y (2017) Realtime multi-person 2D pose estimation using part affinity fields. In: CVPR

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)