Design Studies and Intelligence Engineering
L.C. Jain et al. (Eds.)

doi:10.3233/FAIA220752

Robot Aided Intelligent Kitchen Assistance System Using Eye Tracking Based on Object Recognition

Junhui Guan[a #], Zheng Liu[b,c#], Xiaodong Zhao[a] and Shiwei Cheng[a*1]
[a] *Zhejiang University of Technology, School of Computer Science, Hangzhou 310023, China*
[b] *Zhejiang Provincial Key Laboratory of Integration of Healthy Smart Kitchen System, Ningbo 315336, China*
[c] *China Academy of Art, Hangzhou 310000, China*

Abstract. How to get the potential interaction intent is the key to design a personalized intelligent kitchen assistance system based on Human Robot Interaction (HRI), which can facilitate the elder people in their daily life. Eye movement includes abundant user intent information, eye tracking technology can be used to obtain and analyze this information, thus identifying user's potential intent. This paper proposed an eye tracking based object recognition method, which could recognize objects that the user was watching. Firstly, multiple fixation points were clustered into a fixation point that best represented the user's intent by research on the study of the clustering algorithm. Then the object recognition algorithm based on PP-ShiTu platform was used to obtain the specific location of the object in the scene image. Finally, this paper proposed a fixation based object recognition module, and used the module to design and implement a personalized intelligent kitchen assistance system, and verified the effectiveness through experiments. The experimental results showed that when there was only single food recognition, the average recognition accuracy of the personalized intelligent kitchen assistance system reached over 80%, and the average time delay of the system was about 1400ms.

Keywords. Human robot interaction, Personalized intelligent kitchen assistance system, Eye tracking, Fixation clustering, Object recognition

1. Introduction

The increasing trend of aging of the world population requires attention to the lives of elder people in our society. Assistive robotics plays a key role in technological solutions to address this issue[1]. From the HRI perspective, the design of assistive interaction systems has the potential to reduce the burden of caregivers in hospitals and nursing homes, as well as to take on simple assistive tasks in the home and help the empty nesters. In addition, for the health issues of the elder people, it is even more important. Therefore, how to use assistive robots to help elder people to perform specific tasks in specific scenarios is a worthwhile research question.

* Corresponding author, E-mail: swc@zjut.edu.cn. #Contributed Equally

In recent years, there have been more researches on the modeling and recognition of user intent in the field of psychology and cognitive science, which inspires the researchers on Human Computer Interface (HCI) and HRI to develop new paradigms[2]. Efficient HRI requires the accurate recognition of user interaction intent. However, the current intelligent service robots have limited recognition capabilities. They lack the capability to recognize user's potential intent. The existing HRI interaction methods are based on the user's explicit intent such as voice, button and touch, which is difficult for users with limited mobility, such as disabled people and elder people, to accomplish by themselves. Hence, a new interaction method is needed to enhance the capability of HRI to recognize user's potential intent.

Researchers began to use eye tracking to recognize the potential intent of users. In fact, eye movement contains abundant user intent information and eye tracking can be used to obtain and analyze this information[3]. Previous studies have focused on fixation estimation on two-dimensional screens. For example, the estimation of the fixation position is used to replace the function of the mouse to control the computer cursor[4]. Studying HRI based on eye tracking can enhance the robot's capability to recognize user's potential intent, making HRI more efficient and natural, and enhancing user's interactive experience. However, HRI involves the interaction between humans and robots, and robots have complex hardware and software systems with autonomy and cognition. Since the robots can be used in complex real-world environments, it will bring challenges to the HRI research based on eye tracking.

2. Related work

There are several current research on assistive systems for the elder people[5]. For example, some research integrates smart furniture assistive systems with various household technologies (home appliances, sensors, etc.) to increase the autonomy of elder people and disabled people in kitchen-related activities[6]. And other studies have made recommendations for the design of kitchen aids for patients with cognitive deficits[7]. However, these researchers did not address the problem of how to make the interactive experience of the assistive systems more natural.

In the past decades, the eye tracking technology has shown its potential in HRI field. Eye tracking based HRI research currently focuses on the user's fixation at the specified position on the interactive interface, which is used to trigger the corresponding instructions to control the robot's movement. For example, the user controls the travel of a robotic wheelchair by looking at the associated control interface[8]. Similarly, some studies have also attempted to use eye tracking for teleoperation of mobile robots[9]. Eye tracking techniques are used in these HRIs to trigger some simple, fixed robot control commands. However, in daily life, users will interact with other objects besides robots, inevitably. If the robots can identify these objects the user is gazing at and give them feedbacks, it will bring users with better experience and improve the intelligence of the robots.

From the above researches, it can be seen that the current methods can not accurately recognize the object which the user is watching. It will lead to the low accuracy for robots to recognize the user's potential intent. This paper aims to build a personalized intelligent kitchen assistive system for the elderly people based on user fixation object recognition. This system can improve the elder people's understanding of the health effects of different diets and thus reduce the risk of accidentally eating contraindication foods.

3. Personalized intelligent kitchen assistance system

The digestive ability of the elder people are growing poor, so it is important to select the foods for them before eating. We designed and developed a personalized intelligent kitchen assistance system to introduce the food composition for them, and by introducing the user knowledge graph and the foods knowledge graph, the robot will give reasonable suggestions. This system is divided into two modules: user fixation based object recognition module and personalized assistance module.

3.1. User fixation based object recognition module

3.1.1. Fixation point clustering algorithm

Efficient HRI requires the accurate recognition of user interaction intent. However, the current intelligent system lack the capability to recognize user's potential intent. On this score, eye movement contains abundant user intent information to satisfy the efficient intent recognition. For example, eye movement information allows intelligent systems to know what the user is currently gazing at. Since the sampling speed of user's raw eye tracking fixation is relatively fast, a large amount of fixation data can be generated in a short time. However, only when the eye fixation duration is greater than 100ms, it is considered that this fixation behavior can reflect the real activity of the human brain. When the user's eyes are fixed on certain object, the positions of the raw fixation points are scattered. In order to reduce the data size, facilitate data analysis and use, in this paper we used clustering algorithm to cluster the user's raw fixation points.

The fixation clustering algorithm first collects the coordinates of a group of fixation points and stores them. Then it calculates the mean value of the coordinates of these fixation points. The calculation is shown in Equation 1:

$$\begin{cases} \acute{o}_x = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^{N} (x_i - x_o)} \\ \acute{o}_y = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^{N} (y_i - y_o)} \end{cases} \tag{1}$$

where (x_i,y_i) represents the coordinates of the ith fixation point, and (x_o,y_o) represents the mean value of a group of fixation points. Finally, the standard deviation of these points is calculated. The larger the standard deviation, the more scattered the fixation points are. If the value is greater than the threshold, the set of data is discarded and collected again. Otherwise, it means that the mean value of the set of fixation points can be further used in user fixation based object recognition.

3.1.2. Object recognition algorithm

Various algorithms have involved the neural networks such as YOLO[10], Mask R-CNN[11], SSD[12] and FCOS[13]. However, these algorithms requires high-end hardware and are not suitable in assisted systems. In this paper, we have introduced PP-ShiTu platform, it is a lightweight image recognition platform provided by PaddlePaddle[14], an open source deep learning framework providing a complete set of image recognition functions[15]. This platform doesn't require high-end hardware.

The PP-ShiTu platform includes three parts: subject detection, feature extraction and vector retrieval. First, an image library needs to be constructed, and the features of the image are extracted to generate a feature library. Subject detection is then performed on the input image, which detects one or more subject regions in the image. Features will then be extracted from these detected subject regions using the relevant CNN models. These feature values are composed of float vectors or binary vectors. According to Metric Learning Theory (MLT)[16], the similarity of two objects can be calculated by feature values. The shorter the distance between the two features, the more similar the two objects are. Finally, the vector search algorithm is used to find the features with highest similarity in the pre-generated image feature library, and provide the corresponding label, confidence and object position as the result of this recognition.

As can be seen from Figure 1, the method based on the PP-ShiTu platform can accurately recognize some food objects with a confidence level of up to 0.92, which can basically meet the needs of user fixation based object recognition in this paper.

Figure 1. Recognition results of the PP-ShiTu platform

3.1.3. Fixation based object recognition

Fixation based object recognition, which combines the user's eye tracking information with object recognition, allowing the platform to recognize the object the user is currently gazing at. The object recognition platform can recognize multiple objects (the red rectangle) and their coordinates in each image, and then utilizes the user's clustered fixation point information to locate the object they are gazing at.

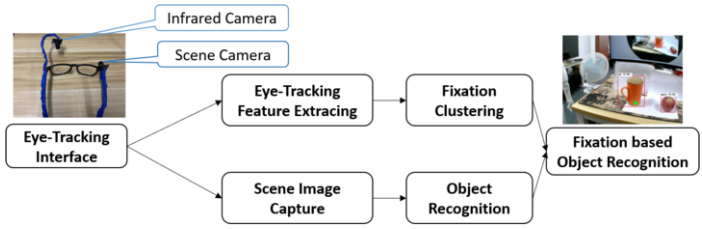


Figure 2. Process of user fixation based object recognition module

Figure 2 shows the process of user fixation based object recognition module. First, the image collected by the scene camera of the eye tracker can detect objects and their coordinates through the object recognition platform. To evaluate whether the user is gazing at the object, here we present a strategy: as shown in Figure 3, taking the center of the rectangle as the center of the circle, half of the diagonal as the radius, we draw a circle on the object. Finally, we calculate whether the current clustering fixation point is within this circle (E_1). If so, we believe the user is gazing at the object. Otherwise, the fixation based object recognition process will be repeated on the rest objects until we find the object.

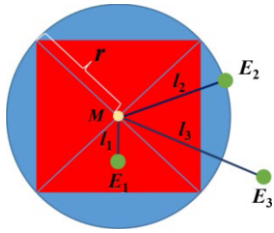


Figure 3. The fixation based object recognition strategy

3.2. Personalized assistance module

3.2.1. User identification

In the assistive system, user identification is required first. In this paper, we use the Dlib toolkit to implement face recognition[17] to identify the users.

The implementation of face recognition includes two steps: face registration and face login. Firstly, the system needs to detect the face as well as calculate the features of the face. During the face registration, the user's face features and name are stored. During the face login, the current user's face features are extracted and compared with the face library, the system will return the message whether login is successful or not.

3.2.2. User knowledge graph building

For the complex relationships among users, foods and diseases, relational databases are not suitable for storing such information, so in this paper we chose Neo4j graph database[2] to build the knowledge graph.

Users were required to enter personal information on the tablet of Pepper robot, which includes user's disease, gender, age, *etc*. The data will be stored in the Neo4j database to complete the user knowledge graph, which provided the basis for personalized assistance.

3.2.3. Food knowledge graph building

The personalized intelligent kitchen assistance system in this paper will assist the user with the preparation of various dishes. Take the scrambled egg with tomato as an example, to prepare this dish required the building of a food knowledge graph first. We analyzed the food, including its categories, effects and side effects. After that, a food knowledge graph of this dish was built. It mainly contained the node of dish, food disease as well as effect.

3.2.4. Human robot interaction

In this paper we built a personalized assisted cooking system to achieve HRI in intelligent kitchen. To assist users to improve their understanding of various foods during cooking, the Pepper robot gives personalized voice prompts by combining the foods and user's own health conditions.

[2] https://neo4j.com

4. Experiments

4.1. Participants

We recruited 10 participants (7 male and 3 female, aged 25 to 65 years). All experiments were approved by the laboratory ethics committee of the author's University, all participants were informed about the procedures of the experiment, and signed the consent form before the experiment. In addition, they were asked to fill in the user information questionnaire, including name, age and physical condition. According to ages, the participants were divided into young group (aged 25 to 33 years) and elder group (aged 60 to 65 years), and each group has five participants.

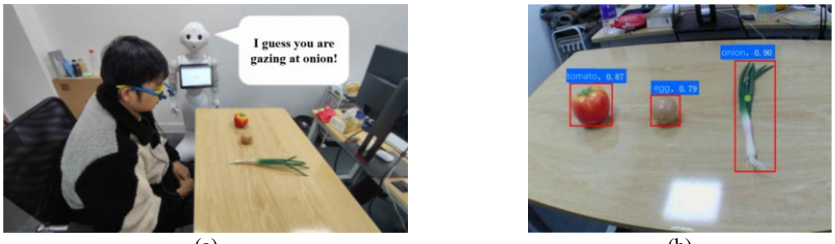
4.2. Experimental equipment

In this paper, we introduced a head-mounted eye tracker developed in previous work[18], it contains several LEDs and 2 infrared cameras: eye camera and scene camera. Each camera has 800×600 pixel resolution and has high sensitivity to infrared rays. In addition, we used the SoftBank's Pepper as the assistant robot.

4.3. Experiment of user fixation based object recognition module

In this experiment, each group were asked to gaze 1-3 objects at a distance of 30-50 cm between them and the objects, and the distance between the objects was set at 15 to 20 cm. Each object was gazed at 20 times.

The experimental scene is shown in Figure 4(a), the participant was gazing at the object on the table and the Pepper robot was saying the object that participant was gazing at from the side. The scenes in Figure 4(b) were collected by the scene camera on eye tracker. The red rectangles were the objects recognized by the system, and the green dots represented the clustered fixation points. By recognizing the object participant was gazing at, this module would send the object to the Pepper robot.

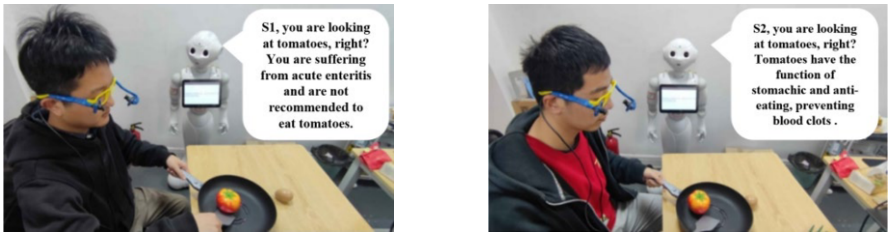


(a) (b)

Figure 4. Experimental scene of user fixation based object recognition

4.4. Experiment of personalized intelligent kitchen assistance system

The participants were required to complete the cooking task of scrambled egg with tomato, and the Pepper robot gave personalized health tips for different users to reduce the adverse impacts of eating, as shown in the Figure 5.



(a) (b)

Figure 5. Experiment scene of robot giving health tips

The recipe of the dish contained 10 steps, as was listed in Table 1, each participant should follow the steps until they finished cooking, and they were asked to gaze at seven kinds of food in order (except the Step 5,6 and 10). The assistant system would recognize the food and provided personalized cooking tips.

Table 1. The recipe, food and instructions

Step Number	Food	Instruction
1	Tomato	Wash and slice the tomatoes
2	egg	Beat the eggs and put in a bowl
3	onion	Wash and slice the onion
4	cooking oil	Turn on the heat of the pan, pour in the appropriate amount of cooking oil
5	none	Pour in the egg mixture and cook
6	none	Heat the pan with oil, add the onions and tomatoes, stir-fry
7	salt	Pour in some salt
8	sugar	Pour in some sugar
9	soy sauce	Pour in some soya sauce
10	none	Add the scrambled eggs to the pan, stir evenly

Each participant was asked to complete 10 tasks. Note that in this paper, we marked it as one successful completion of the task, only when the system recognized user's every fixation object and gave the correct cooking tips through all 10 steps.

Finally, participants were asked to rate the following subjective evaluation questions using a 5-point Likert scale:

- A. Please score your satisfaction with the personalized kitchen intelligent assistance system in the experiment. (1=least satisfied, 5=most satisfied)
- B. Please score the level of fatigue caused by using the eye-tracking device. (1=very fatigued, 5=not fatigued)
- C. Please score how easy it is to use the robot application. (1=very hard, 5=very easy)
- D. Please score your adaption to use the robot for interaction. (1=not adapt to, 5= very adapt to)

5. Result

5.1. Experiment of user fixation based object recognition module

Table 2 shown the recognition accuracy of the user's fixation on different amounts of objects. Recognizing one object reached the highest recognition accuracy, with the average accuracy of 92.00% for the young group and 88.00% for the elder group, which outperformed that of two and three objects. The young group performed better than the elder group in all tasks, this was probably because the elder people could not focus as much as the young group did. The recognition accuracy of all tasks reached above 80%, verifying the feasibility of the module to recognize intents in HRI.

Table 2. The recognition accuracy of different amount of objects across young and elder group.

Group	Participant ID	One object	Two objects (%)	Three objects (%)
Young	S1	95.00	90.00	85.00
	S2	90.00	85.00	85.00
	S3	90.00	85.00	80.00
	S4	95.00	90.00	85.00
	S5	90.00	90.00	85.00
	AVG±SD	92.00±2.73	88.00±2.58	84.00±2.04
Elder	S6	90.00	90.00	85.00
	S7	90.00	85.00	85.00
	S8	90.00	85.00	80.00
	S9	85.00	85.00	80.00
	S10	85.00	80.00	80.00
	AVG±SD	88.00±2.58	85.00±3.76	82.00±2.73

5.2. Experiment of personalized intelligent kitchen assistance system

The average accuracy of cooking tip for young group and elder group were 92.29% and 86.01%, respectively. And a Mann-Whitney U test showed that the average accuracy from young group was significantly higher than that of elder group (Z=-2.619, p<0.05). The average accuracy in young group all reached above 90%, which proved the reliability of the personalized kitchen intelligent assistance system proposed in this paper. However, most of the elder participants have poor vision, which cost more time to focus on the food, making the average accuracy worse than young group. The average running duration of giving the personalized tips for young group and elder group were 1355.76ms and 1461.34ms, respectively, which indicated that giving tips was efficient with a low delay. Similarly, the elder group did not perform as well as the young group, because they could be distracted or discomforted with the AI Pepper robot[19].

Table 3. Average accuracy and runtime duration of personalized cooking tips.

Group	Participant ID	Average accuracy (%)	Runtime duration(ms)
Young	S1	90.00	1379.12
	S2	92.86	1348.95
	S3	91.43	1355.31
	S4	92.86	1382.43
	S5	94.29	1312.98
	AVG±SD	92.29±1.73	1355.76±26.79

Group	Participant ID	Average accuracy (%)	Runtime duration(ms)
Elder	S6	85.71	1458.72
	S7	88.57	1437.81
	S8	82.86	1495.67
	S9	88.60	1429.11
	S10	84.29	1485.39
	AVG±SD	86.01±2.29	1461.34±25.93

In addition, Figure 6 showed the subjective evaluation results (the vertical lines represent the standard deviation intervals). It can be seen that most participants were satisfied with the assistant system designed in this paper. Moreover, the interaction process required them to wear an eye-tracker for a long time, which would cause some fatigue to the eyes and they would feel uncomfortable. In this respect there was a mainly impact on the elderly. Therefore, we will take this into consideration in our future work. Then, the operator interface of Pepper robot was developed based on a tablet, so it was less difficult for most participants to use. For the adaption, most participants in both groups gave high scores. Additionally, at the end of the experiment an elder participant told us that this system helped them to better understand various foods and that he looking forward to using this type of system in the future.

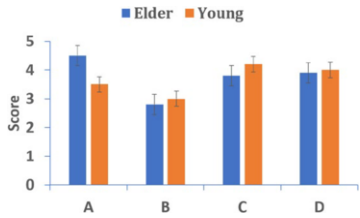


Figure 6. User subjective evaluation(the error bar is standard deviation)

6. Conclusion and Future Work

This paper focuses on how to design a intelligent kitchen assistance system for elder people to use, and facilitate their healthy diet. Firstly, the fixation point clustering method is proposed, and multiple fixation points are clustered into a fixation point that can best represent the user's intent. Secondly, the object recognition method based on PP-ShiTu platform is developed. Third, a module for user fixation based object recognition based on eye-tracking is proposed by combining the first two methods. Finally, the module is used in the intelligent kitchen assistance system and the effectiveness of the system is verified by experiments.

Although the user fixation based object recognition module studied in this paper has achieved certain results and applied it to HRI in intelligent kitchens, in addition, the experimental results show that the system has great potential, it is still in the exploratory stage. How to make HRI in intelligent kitchens more practical still needs to to be researched deeply. For example, the current knowledge graph is relatively single, and it will be extended in the future to implement more functions for the robot. Then, there should be more interaction design when the system does not recommend the food. For example, let the user choose to continue using or change it, instead of only voice prompts. In addition, the head-mounted eye-tracker we used in this paper needs to connect with

the system via USB, which limits the user's scope of activity in the experiment. In the future, the system will be connected wirelessly to allow the user to be free.

Acknowledgement

We thank all the volunteers, and the grant from Natural Science Foundation of China (No.62172368, 61772468), Fundamental Research Funds for the Provincial Universities of Zhejiang (No. RF-B2019001), and Zhejiang Provincial Key Laboratory of Integration of Healthy Smart Kitchen System (No.2020F04).

References

[1] Christoforou E G, Avgousti S, Ramdani N, et al. The upcoming role for nursing and assistive robotics: Opportunities and challenges ahead. Frontiers in Digital Health, 2020: 39.
[2] Wang W, Li R, Chen Y, et al. Human intention prediction in human-robot collaborative tasks. Companion of the 2018 ACM/IEEE international conference on human-robot interaction. 2018: 279-280.
[3] Majaranta P, Bulling A. Eye tracking and eye-based human–computer interaction. Advances in physiological computing. Springer, London, 2014: 39-65.
[4] Zhang X, Liu X, Yuan S M, et al. Eye tracking based control system for natural human-computer interaction. Computational intelligence and neuroscience, 2017.
[5] Kuoppamki S, Tuncer S, Eriksson S, et al. Designing Kitchen Technologies for Ageing in Place: A Video Study of Older Adults' Cooking at Home. Proceedings of the ACM on Interactive Mobile Wearable and Ubiquitous Technologies, 2021, 5(2):1-19.
[6] Blasco R, Marco Á, Casas R, et al. A smart kitchen for ambient assisted living. Sensors, 2014, 14(1): 1629-1653.
[7] Kosch, Thomas, et al. "Smart kitchens for people with cognitive impairments: A qualitative study of design requirements." Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018.
[8] Shinde S, Kumar S, Johri P. A Review: Eye Tracking Interface with Embedded System & IOT. 2018 International Conference on Computing, Power and Communication Technologies (GUCON). IEEE, 2018: 791-795.
[9] Bolarinwa J, Eimontaite I, Mitchell T, et al. Assessing the role of gaze tracking in optimizing humans-in-the-loop telerobotic operation using multimodal feedback. Frontiers in Robotics and AI, 2021, 8.
[10] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788.
[11] He K, Gkioxari G, Dollár P, et al. Mask r-cnn. Proceedings of the IEEE international conference on computer vision. 2017: 2961-2969.
[12] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector. European conference on computer vision. Springer, Cham, 2016: 21-37.
[13] Tian Z, Shen C, Chen H, et al. Fcos: Fully convolutional one-stage object detection. Proceedings of the IEEE/CVF international conference on computer vision. 2019: 9627-9636.
[14] Ma Y, Yu D, Wu T, et al. PaddlePaddle: An open-source deep learning platform from industrial practice. Frontiers of Data and Domputing, 2019, 1(1): 105-115.
[15] Wei S, Guo R, Cui C, et al. PP-ShiTu: A Practical Lightweight Image Recognition System. arXiv preprint arXiv:2111.00775, 2021.
[16] Kaya M, Bilge H Ş. Deep metric learning: A survey. Symmetry, 2019, 11(9): 1066.
[17] Suwarno S, Kevin K. Analysis of face recognition algorithm: Dlib and opencv. Journal of Informatics and Telecommunication Engineering, 2020, 4(1): 173-184.
[18] Cheng, S., Fan, J., & Dey, A. K. (2018). Smooth gaze: a framework for recovering tasks across devices using eye tracking. Personal and Ubiquitous Computing, 22(3), 1-13.
[19] Sharkey A, Sharkey N. Granny and the robots: ethical issues in robot care for the elderly. Ethics and information technology, 2012, 14(1): 27-40.