Real-Time Eye-Gaze Based Interaction for Human Intention Prediction and Emotion Analysis*

Hao He[1,2]
24320151153221@stu.xmu.edu.cn

Yingying She[1,*]
yingyingshe@xmu.edu.cn

Jianbing Xiahou[1]
jbxiahou@xmu.edu.cn

Junfeng Yao[1]
ao0010@xmu.edu.cn

Jun Li[2]
junli@fjirsm.ac.cn

Qingqi Hong[1]
hongqq@xmu.edu.cn

Yingxuan Ji[1]
24320171152286@stu.xmu.edu.cn

1 Software School, Xiamen University
2 Quanzhou Institute of Equipment Manufacturing CAS
* Corresponding author

ABSTRACT

The human eye's state of motion and content of interest can express people's cognitive status and emotional status based on their situation. When observing the surrounding things, the human eyes make different eye movements according to the observed objects which reflects human's attention and interest. In this paper, we capture and analyze patterns of human eye-gaze behavior and head motion and classify them into different categories. Besides, we compute and train the eye-object movement attention model and eye-object feature preference model based on different peoples' eye-gaze behaviors by using machine learning algorithms. These models are used to predict humans' object of interest and the interaction intention according to people's real-time situation. Furthermore, the eye-gaze behavior and head motion patterns can be used as a modality of non-verbal information in the computing of human emotional states based on the PAD affective computing model. Our methodology analyzes human emotion and cognition status from the aspect of eye-gaze behavior and head motion, understands the cognitive information that human eyes can express, and effectively improves the efficiency of human-computer interaction in different circumstances.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI);**

KEYWORDS

Eye-gaze interaction, robot and vision, machine learning

1 INTRODUCTION

Eye movement analysis is a popular research topic in the field of HCI which has been applied in many domains. However, most eye-gaze related applications are limited to data collection and statistical analysis, and there is no systematic process which combined the analysis of human eye-gaze and situated interaction. In addition, there is no further cognitive analysis of eye-gaze related user's intention and emotional state. In this paper, we present our solution of eye-gaze based interaction which can optimize the cognitive process of HCI (Human-computer interaction) and HRI (Human-robot interaction). There are three contributions of this paper. 1) The eye-object movement attention model and feature preference model are constructed based on eye-gaze behavior and head motion patterns; 2) These two models are trained and used in the process of human interaction intention prediction in real-time; 3) eye-gaze interaction models are combined with other non-verbal and verbal information to analysis people's emotional states. Our methodology can be used as the support for eye-gaze based multimodal interaction for the development of HCI, HRI and NUI (Natural user interface) applications. The specific eye-gaze learning and affective computing model can reveal valuable information for human cognitive process during interactions.

2 BACKGROUND

Eye-gaze interaction could significantly improve the efficiency of HCI, and HRI [16,21]. In general, the eye-gaze data was mainly used for the physical control in HCI and statistical analysis. Zhao proposed classification methods of eye physical movements and introduced the related application fields [17]. Microsoft embedded Tobii eye tracking technology [4] into the latest Windows 10 in order to have full-scale and hand-free interaction. Dagmar Kern [19] helped users to quickly revert previous visual concerns by recorded gaze-mark. Sundsted [6] used the eye movement to control the virtual characters in games. WADE [14] analyzed the physical eye movements of autistic patients in a VR driving scene and corrected their actions in order to help them improve their driving skills. K. Ruhland [15] examined the possible impacts of virtual characters' eye movements by simulating the behaviors of the human eyes. Yun Suen Pai [20] used head-and-eye manipulation to navigate in a virtual reality environment.

In addition to exploring the relationship between eye movements and physiological status, the hidden psychological information of eye-gaze movement was also of great value [9]. Vicente [5] analyzed the states of the eye movement to determine the concentrations of the driver while driving. Zheng [18] improved the UI(User interface) design of car navigation system by analyzing the drivers' eye movement information. M Borys [7] analyzed the relationship between the eye movement status and the level of fluency in presentation based on classification methodologies. In the field of psychology [1], researchers combined eye attention with other methods to study how the brain works and to analyze brain-controlled behavioral patterns in different states [8], such as eye attention differences among peoples for crafts, advertisements, and films. When people were in an interactive environment, they showing different behavioral trends based on the different information presented by the environment [24]. Lance's analyzed the mapping of the gaze states and the user's emotional states, and revealed that the human eye movement behaviors could reflect the human emotional states [2,26].

3 METHODOLOGY

The human eye can acquire and process visual information within a certain range. Since most of the visual contents in the real world are presented on two-dimensional planes (such as screen, billboard, etc.), our method focuses on user behavioral intention prediction and user emotional state analysis based on eye movements, eye-gaze interaction and interactive objects on 2D screen. Firstly, we develop the classification algorithms by extracting the eye movement and head motion features to classify the eye movement and head pose categories. Secondly, the relationship between the object's movement categories and states of eye attention are analyzed to construct the eye-object movement attention model, which can be applied on the prediction of the moving objects that the eyes are focusing. Furthermore, since the human eyes may focus on different categories of moving objects in the same region at the same time, the features of objects are being considered as well. The object features contain the elements that describe the characterizations of the objects, including color, shape, and the moving category. By analyzing the preferences of different people on different object features, the eye-object feature preference model is constructed. It can be combined with the object movement attention model to predict the concentration object from user's eyes. The result of user's eye attention object as well as the context information from the environment can be computed together to predict people's behavioral intentions during interaction. In addition, computed with the PAD model [27], the real-time eye movements and head poses can express the emotional states of people. The framework of our methodology is shown in Fig. 1.

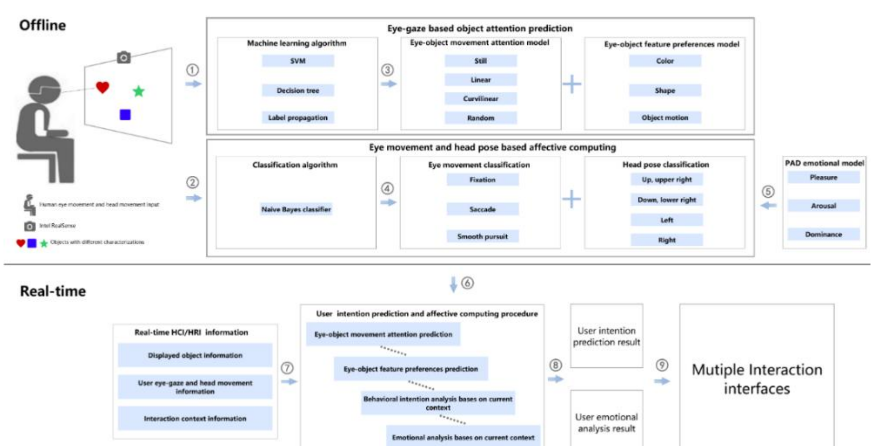


Figure 1: Real-time human intention prediction and affective computing based on eye-gaze analysis main framework.

In Fig. 1, step 1 & step 2, our approach captures and analyzes the user's eye-gaze and head motion information. Step 3, the machine learning classification algorithms are used to construct the eye-object movement attention model and the eye-object feature preference model, which are make of the eye-gaze based object attention prediction module. Step 4 & step 5, at the same time, we use the Naive Bayesian classification algorithm to construct the eye movement and the head pose classifiers. Then, the classification results combine with the PAD emotion model form the eye-gaze and head pose based affective computing module. Step 6 & step 7, with the support of the offline created and trained modules, we can obtain the HCI elements in the application scenario in real-time, including the object information displayed on the scene, the eye-gaze and head motion input of the user and the interaction context information from the environment. Step 8, in the user behavioral intention prediction and affective computing procedure, the user's potential attention moving objects are predicted first, and the final attention object is determined according to user's object feature preferences. Based on the attention object, user's behavioral intention can be predicted according to the current interactive context. At the same time, user's emotional state can be analyzed with the eye movement and head pose as well as the PAD emotion model. Finally, the analysis results of the user behavioral intention and emotional state is obtained. In step 9, the output results can be used as the user intention prediction and emotional analysis feedback. Also, it can be served as the input of the decision-making module in different interaction systems.

3.1 Eye-object movement attention model and eye-object feature preference model

For the construction of eye-object movement attention model, we adapt three supervised learning classification algorithms, label propagation algorithm [10], SVM algorithm [11], and decision tree algorithm [12]. In order to learn relationships between eye-gaze patterns and different object movements, we construct the attention models for objects with different moving categories. According to the result of kinematics analysis [13], we can specify the moving category of objects as still, linear, curvilinear, and random, which described in Table 1.

Table 1: The object movement category list

Object movement categories	Descriptions
Still	Object keeps still
Linear	Object moves in straight line
Curvilinear	Object moves in curve line
Random	Object appears randomly

Firstly, we collect eye-gaze data for still, linear, curvilinear, and random moving objects from human. Next, the eye-gaze point coordinates as well as the moving object central point coordinates are treated as the learning features. The object of concentration is treated as the learning label. And classifiers described above is used to train the eye-object movement attention model for specific moving category of objects on screen. In real-time, the model can predict whether eyes are paying attention to specific object by analyzing the eye-gaze data from human.

For the construction of eye-object feature preference model, we use the same classification methods as the attention model. The feature attributes, such as color, shape, and movement category of objects are treated as learning features, as shown in Table 2. The feature preference of human is treated as the learning label to train the eye-object feature preference model for different individuals. In real-time, the model can predict which object has specify features that people may be interested in.

Table 2: The object feature list

Object features	Descriptions
Shape	The shape of the object
Color	The color of the object
Movement	The current movement category of the object

In the next step, we explain the training processes of these two models based on three machine learning classification algorithms which are label propagation algorithm, SVM algorithm, and decision tree algorithm. In detail, for the attention model training, the training feature includes the gaze point coordinates as well as the object coordinates, and the corresponding positive attention state is '1'. The feature model training includes the object color, object shape and movement category, and the corresponding positive attention state is '1'. After the training process, we obtain the eye-object movement attention model and eye-object feature preference model, these two models are combined in order to form the eye-gaze based object attention prediction module.

3.2 Eye movement and head pose classification method

In previous work, the definition and classification of eye movements [22] are defined. We use the threshold-based method to segment the eye elements such as iris and pupil. After that, the eye movement features are extracted, including unit-time translation and the moving velocity of eyes. Finally, the eye movement features are classified by the Naive Bayes classification algorithm and generated the classification result. The eye movement categories are divergence, fixation, saccade, and smooth pursuit, as shown in Table 3. They correspond to different eye-gaze behaviors and reflect different physical status of the eyes.

Table 3: The categories of eye movement

Eye movement categories	Descriptions
Divergence	Pupils have not focus point

Fixation	Pupils nearly have not actions
Saccade	Eyes move up-side down or left to right, the translation values of them change quickly
Smooth pursuit	Eyes keep moving in a certain speed, but pupils nearly stay still

As the definitions of eye movement category, the Naive Bayes classifier is used to classify each category based on eye movements. In order to improve the classification accuracy, two feature values, the translation vector T and the velocity vector V of pupil center point, is added to the classification algorithm. The feature combination described as follows:

$$F = V + T \tag{1}$$

The detail steps of the classification are showed in the algorithm below. The eye movement category can be classified in real-time with the Naïve Bayes classifier.

ALGORITHM 1: The eye movement classification algorithm based on Naïve Bayes classifier

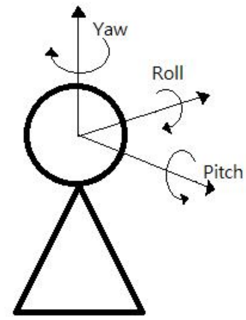
Input: Eye movement feature data sets and trained Naïve Bayes classifier

Output: Category of eye movement

Steps:

1. *Assume the feature data set as: {D1, D2, D3, …, Dn}, n as the iteration upper limit, t as the iteration count;*
2. *Counter: t = 1;*
3. *Category = Bayes. predict {D1, D2, D3, …, Dn}, calculate the velocity V and translation T of the feature data, if those factors match the value of corresponding movement category, return it as the predict category.*
4. *t = t + 1;*
5. *if t<=n, repeat step 3;*
6. *Figure out the mode of {D1, D2, D3, …, Dn} from all the iterations, if the mode exists, then it is the category of the eye movement;*
7. *Return the category of the eye movement.*

In order to enhance the accuracy and efficiency of eye-gaze interaction, the head motion information is also considered in our method. The Euler angles of the head, yaw, roll, and pitch can identify the current pose of the head [23]. Fig. 2 shows how Euler angle describes the head pose and as well as the existing head poses.



Head poses
Up,
Down
Right
Left

Figure 2: The relationship illustration between head pose and Euler angles as well as the head pose list.

After computation, we can identify the head pose through two rotation dimensions yaw and pitch. The calculation formula is as follows:

$$\mathrm{Pose}_{yaw} = \begin{cases} left\ shift, & if(\theta_i - \theta_0) > 0 \\ right\ shift, & if(\theta_i - \theta_0) < 0 \\ origin, & otherwise \end{cases}$$
$$\mathrm{Pose}_{pitch} = \begin{cases} up\ shift, & if(\theta_i - \theta_0) > 0 \\ low\ shift, & if(\theta_i - \theta_0) < 0 \\ origin, & otherwise \end{cases} \tag{2}$$

Where θ_i denotes the angle of yaw or pitch of the i_{th} current time, θ_0 denotes the angle of yaw or pitch of the first time. The different value of yaw angles determines whether the head turned left or right, and the difference value of pitch angles determines whether the head turned up or down. When the difference between θ_i and θ_0 is greater than 0 in the yaw dimension, the head turns to the left, or turns to the right. When the difference between θ_i and θ_0 is greater than 0 in the pitch dimension, it determines the head turns up or turns down.

3.3 Eye movement and head pose based affective computing

From classifiers explained in previous sections, we can obtain the real-time eye movements and head poses. Then, we use the PAD affective computing model to build the eye movement and head pose based affective computing module. The categories of emotional state and related emotional dominance in PAD emotional space are shown in the following Table 4 and Table 5.

Table 4: The emotional categories

Emotional categories
Contempt
Anger
Excitement
Fear

Guilt
Sadness
Surprise

Table 5: The emotional dimension rating scales

Emotional dimensions	Rating statements
Hight dominance	The character is dominant
Low dominance	The character is submissive
Hight arousal	The character is agitated
Low arousal	The character is relaxed
High valence	The character is pleased
Low valence	The character is displeased

There is a significant relationship between the PAD model emotional dimensions and eye movements as well as the head poses [3]. As shown in the following Table 6, when users make the corresponding head poses and eye movements, we can identify users' current emotion state based on the PAD model.

Table 6: Emotional categories and significantly related behavior combinations

Emotional categories	Head poses
Contempt	Raised
Anger	Neutral
Excitement	Neutral
Fear	Neutral
Guilt	Bowed
Sadness	bowed
Surprise	Raised

Through the mapping between the head poses, eye movements and the PAD emotion model, the eye movement and head pose based affective computing module is generated.

After constructing the offline modules, the eye-gaze based object attention prediction module can be used to predict the user's attention object, and further combined with the context information from the environment to predict the user's behavioral intention. Besides, the eye movement and head pose based affective computing module can be applied to analyze the user's emotional status. In the following parts, we describe the process of the offline modules computing. In order to fulfill the practical usages of the offline modules, four application scenarios are presented to demonstrate the accuracy and usage of our method in real time, including the multiple users eye-gaze attention content sharing scenario, the cursor-aided shooting game scenario, the hand-free interaction for social robot scenario, and the nonverbal affective computing for social robot scenario.

4 EXPERIMENT

At first, we build the eye-object movement attention model of objects in different moving categories, as well as the eye-object feature preference model of different objects with different characterization features for different individuals, which form the eye-gaze based object attention prediction module. Secondly, based on the Naive Bayesian classification algorithm, we analyze the eye movement and head pose features, design the corresponding classification algorithms to classify the user's eye movements and head poses. At the same time, with the support of the PAD emotion model, we map the classification results of the eye movement category and the head pose with the user's emotional state to form the eye movement and head pose based affective computing module. At the end, in order to test the accuracy and practical usages of the modules in real time, we design four application scenarios, including the multi-user eye-gaze attention content sharing scenario, the cursor-aided shooting game scenario, the hand-free interaction for social robot scenario, and the nonverbal affective computing for social robot scenario. During the real-time test, we propose a human-computer interaction approach based on eye-gaze behaviors and analyze the practicality and convenience of it according to its functional accuracies. At the same time, we use the eye movement and head pose based affective computing results as the input of the social robot's non-verbal affective computing procedure in order to help the robot understand the user's emotional state, and analyze the user's cognitive state in interactive context.

4.1 Offline construction—eye-gaze based object attention prediction module

The variables controlling method is the key point during the training process of our models. The types of obtained data including the user's ID, the user's eye-gaze point coordinates on the 2D screen, the coordinates of the moving object central point on the 2D screen, the categories of the object's movement, the colors of the object, the shapes of the object, and the user's subjective eye-object attention state, as shown in Table 7.

Table 7: The obtained data types

Obtained data types	Obtained data types
User ID	User ID
Eye-gaze point x	Eye-gaze point x
Eye-gaze point y	Eye-gaze point y
Object point x	Object point x
Object point y	Object point y
Object movement category	Object movement category
Object color	Object color
Object shape	Object shape
User attention state	User attention state

We use machine learning algorithms, including label propagation algorithm, SVM algorithm and decision tree algorithm to learn the relationships among the collected data and generate the trained models, the feature data types are shown in the following Table 8.

Table 8: The machine learning algorithm training data types

Varieties	Explanations	Data types
Object movement category	The movement categories of the object	enum {Still, Random, Curve, Line}
Object shape	The shape of the object	enum {Square, Circle, Triangle}
Object color	The color of the object	enum {Red, Blue, Yellow}
Time	Current time	int: int: int: int
Object point x, Object point y	X and Y axis coordinates of object	int
Gaze point x, Gaze point y	X and Y axis coordinates of gaze point	int
User attention state	The current attention state of user	bool

The entire eye-object attention experiment must be carried out in a specific physical environment. The experiment uses Intel RealSense as the capture device for eye-gaze data, which is placed above a 2D screen. The tester put the head on a stand, and keep it in an immovable state. The tester's visual range is the entire 2D screen whose eye-to-screen distance remains constant. After setting up the experimental environment, the eye-gaze data is collected.

1) For the construction of the eye-object movement attention model, except for different categories of movement, the object's characterization features are unchanged. We arrange for 100 adult testers, 50% male and 50% female, to experiment in the physical environment that has been introduced above. The testers observe the objects movements on the 2D screen and follow the content of the presented objects. At the same time, we mark the current positive state of attention through the keyboard as '1'. Each experiment time is strictly 5 seconds, and each tester performs 10 repeat tests for every category of the movement object. During the experiment, the eye-gaze data generated by the testers and the movement trajectories of the objects on the 2D screen are recorded in a time-aligned manner and stored in the file labeled with the tester's ID tag for further data collection and processing. In the construction of this model, the gaze point coordinates as well as the object coordinates on the two-dimensional displayer are used as the learning features, and the eye-object attention state are used as the learning labels. After the data collection and model training, we can use the eye-object movement attention model to predict which objects the user is following or paying attention to. Table 9 shows the testing accuracies of the eye-object movement attention model during its training process.

Table 9: The testing accuracies of eye-object movement attention model

Object categories	Times	Accuracies
Still	1000	93.3%
Linear	1000	92.1%
Curvilinear	1000	89.2%
Random	1000	88.4%

2) For the construction of the eye-object feature preferences model, objects with different movement categories, colors and shapes are presented on a 2D screen. There are 100 adult testers, 50 males and 50 females. Testers observe the objects moving on the screen and follow the contents of the objects, they label the object which is currently paying attention to by pressing the keyboard as the number showed on the objects. Each experiment time is strictly set as 5 seconds. Each tester performs 10 times repeat tests as only one feature attributes of the object showed differently on the 2D screen, and the other attributes remain unchanged. The characterization feature attributes, including color, shape, movement of different objects are used as the learning features. The object that eyes are paying attention to is used as the learning label. During the training process, we use this model to predict which object that the user prefers to or pay attention to as the difference in characterization features. Table 10 shows the testing accuracy of the eye-object feature preferences model.

Table 10: The testing accuracies of eye-object feature preferences model

Feature categories	Times	Accuracies
Shape	1000	85.6%
Color	1000	90.3%
Movement category	1000	82.3%

By testing the accuracy of classifiers, it shows that the accuracy of the three classification algorithms are very close, as a result, we use three models to predict the eye attention object at the same time and use the result of the averaging as the final prediction result.

4.2 Offline construction—eye movement and head pose based affective computing module

In the methodology section, we introduce the classification methods of eye movements and head poses. During the test, we invite 10 testers, 5 males and 5 females, to perform different types of eye movements and head poses for 100 times. The classification algorithms categorize the user's eye-gaze and head motion input and record the classification results in the text. The following Table 11 and Table 12 show the classification accuracies of the eye movement categories and head poses.

Table 11: The classification accuracy of the eye movement

Eye movements	Classificatio n success rates
Fixation	100%
Saccade	87.3%
Smooth Pursuit	72.3%
Divergence	100%

Table 12: The classification accuracy of the head pose

Head poses	Classificatio n success rates
Up	98.3%
Down	97.4%
Left	98.4%
Right	99.1%

4.3 Real-time test—multi-user eye-gaze attention content sharing

In a multi-user meeting, participants who want to speak can project what they want to show on the main screen by paying attention to the content on their screens. To realize the functions described above, we firstly obtain the screen content that the speaker pays attention to by using eye-gaze based object attention prediction module, and then judge whether the speaker's head's direction of rotation is in line with the direction of the main screen according to the head pose classification method. If the conditions are met, the attention content of the speaker on the screen is projected to the main screen. Fig. 3 and Fig. 4 show the interaction illustration as well as the scenario photos.

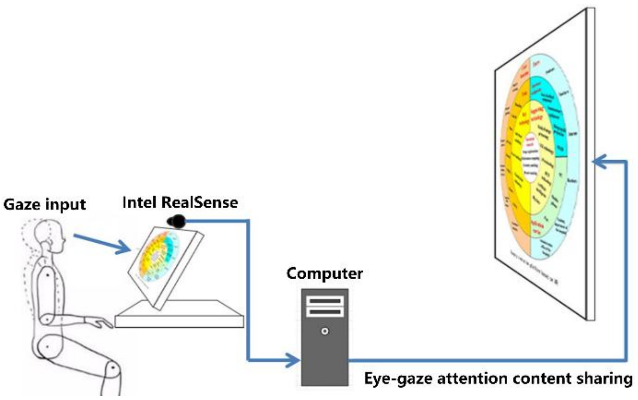


Figure 3: The interaction illustration of multi-user eye-gaze attention content sharing.

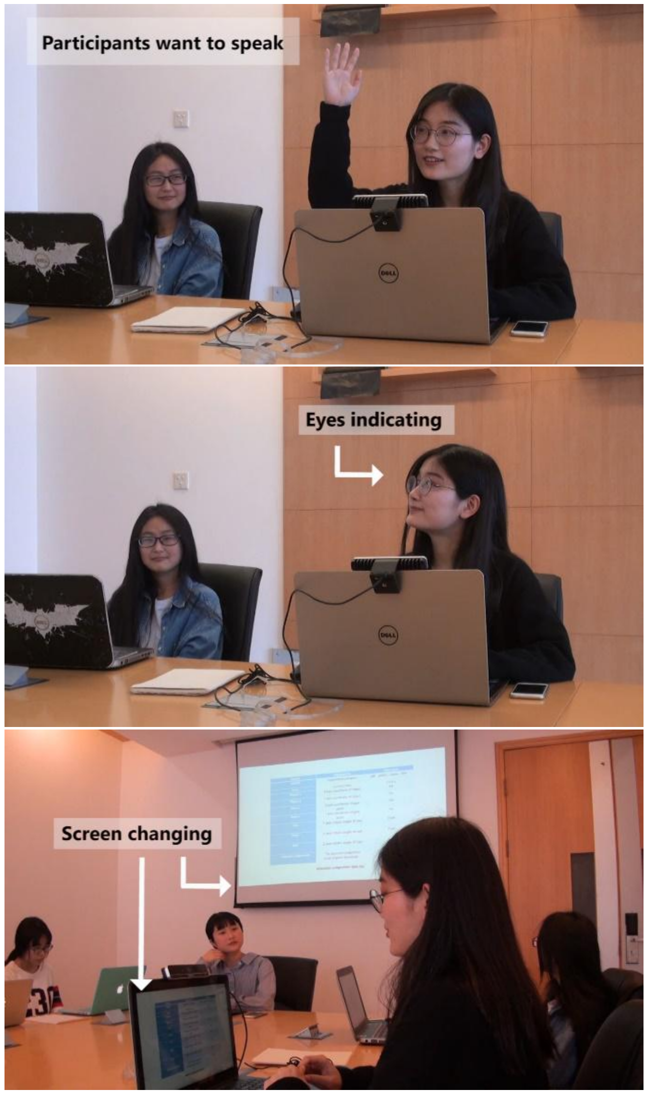


Figure 4: The scenario photos of multi-user eye-gaze attention content sharing

4.4 Real-time test—cursor-aided shooting game

In this game scenario, by analyzing the object of attention of the player in the current time, the system helps the user to perform quick cursor-assisted aiming to the locked object, users only need to use the mouse or the handle to perform an easy quasi-lock. Fig. 5 shows the procedures of this game.

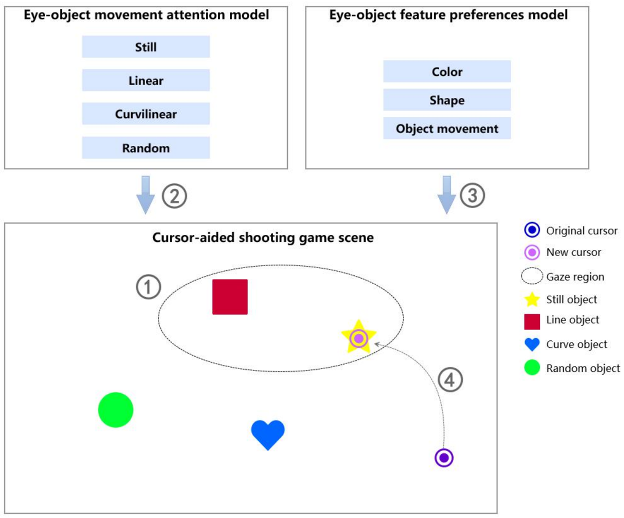


Figure 5: The cursor-aided shooting game interaction illustration.

Figure explanation: step 1, mark out the eye-gaze region around the gaze point. Step 2 & step 3, use eye-gaze based object attention prediction module to predict the user's attention object. Step 4, output the result of the attention object and move the cursor automatically to it.

We invite 5 male players to experience this shooting game. The procedure firstly predicts the player's attention object with the eye-gaze based object attention prediction module, and next moves the shooting cursor to the corresponding object to complete the shooting. During this process, the player will record their attention objects by pressing the keys. The recorded results are compared with those predicted by our method to test the accuracy of the module.

4.5 Real-time test—hand-free interaction for social robot

The social robot can predict the user's behavioral intention by analyzing the user's eye-gaze information with the support of eye-gaze based object attention prediction module and refer to its prior knowledge library to generate the proper interactive feedback. When the user is busy doing things with both hands, he can communicate with the robot through his eye-gaze behaviors. For example, when the user is making Kungfu tea according to the tutorial displayed on the screen and cannot free his hands to control the screen. At this moment, the robot can perceive the area of interest from the user's eye-gaze behavior to help the user switch the instruction steps on the screen, providing the user with a very convenient and efficient interaction experience. The scene photos are showed in Fig. 6. We invite 1 male tester to interact with the robot for 20 times, who judge the robot's feedback during the interactions. Besides, the interaction operation time of the traditional screen touch and the eye-gaze behavior interaction are collected to compare the efficiencies between these eye-gazed based or no eye-gazed based interaction paradigms.

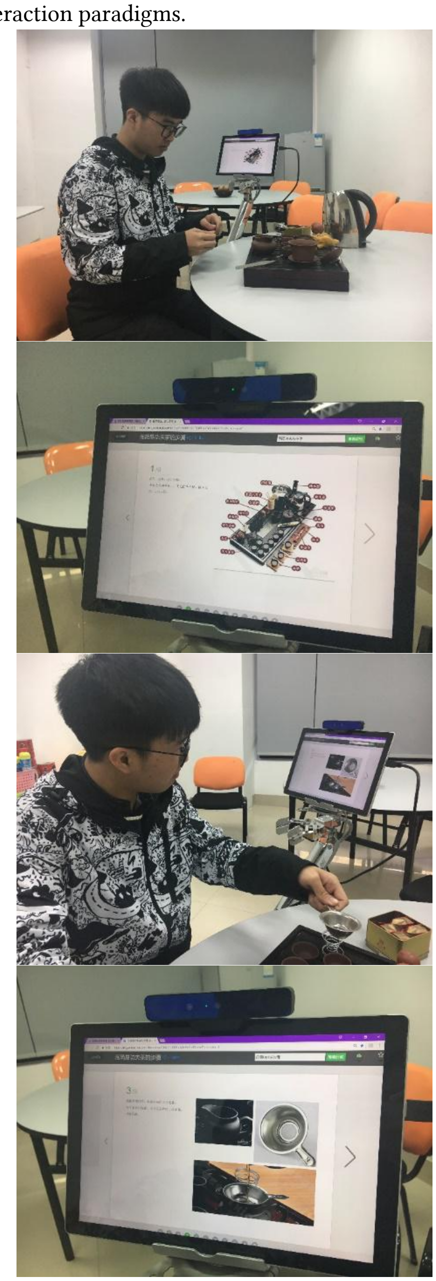


Figure 6: Hand-free interaction with eye-gaze behavior illustration.

4.6 Real-time test—nonverbal affective computing for social robot

In addition to help users free their hands during HRI, we also analyze the user's current emotional state by controlling the robot to capture the user's head poses as well as eye movements, in order to understand the user's cognitive states in a concise way by combining the multimodal interaction input information. In this scenario, the social robot uses Intel RealSense to capture multimodal information from the user, including both non-verbal and verbal information. In our method, we focus on the analysis of the user's non-verbal information input, that is, the potential affective information carried by the states of head pose and eye movement. Through the eye movement and head pose based affective computing module, the user's emotional state of non-verbal information is calculated and combined with the emotional state of verbal information to analyze the overall emotional state and cognition state of the user. Fig. 7 shows the affective and cognitive computing procedure of the social robot.

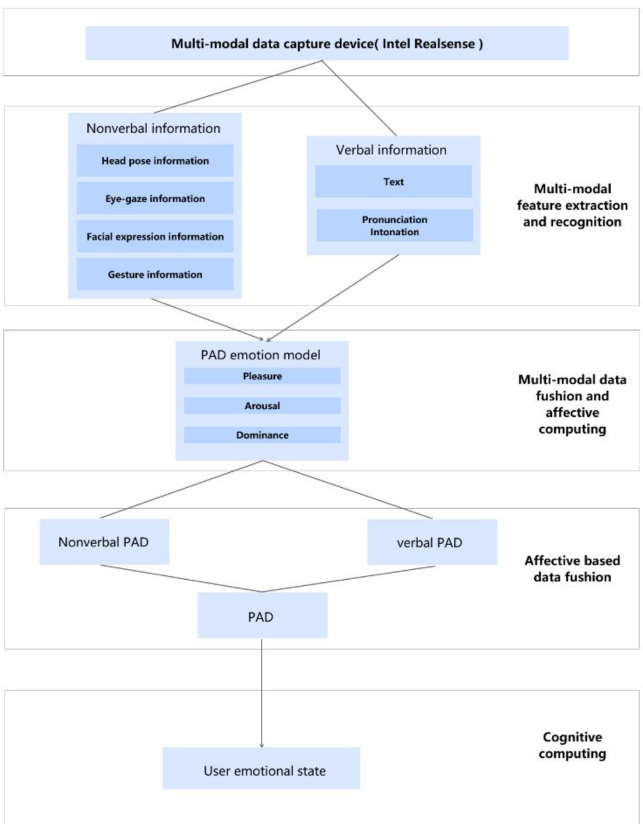


Figure 7: Human-robot social interaction through nonverbal affective computing.

5 RESULTS

In the experiment, we acquire the eye-gaze data of the user on the 2D screen, combined with the object moving data and the object characterization features with the machine learning algorithms to construct the eye-gaze based object attention prediction module. In addition, we extract the eye movement and head pose features, through Naive Bayesian classification algorithm, develop the eye movement category and head pose classification algorithms. With the support of PAD emotion model, the eye movement and head pose information are mapped with human's emotional state, as a result, the eye movement and head pose based affective computing module is generated. To test the accuracies and practicalities of these modules in real-time, four application scenarios are designed.

In the multi-user eye-gaze attention content sharing scenario, we propose an interactive method based on the combination of eye-gaze attention and head pose to help different speakers in a multi-user conference quickly share the content.

In the cursor-aided shooting game scenario, the prediction results of eye-gaze based object attention prediction module are satisfactory. Table 13 shows the prediction accuracies of the assistive aiming.

Table 13: The prediction accuracy of cursor-assisted aiming game

Object categories	Times	Accuracies
Still	200	90.4%
Linear	200	91.3%
Curvilinear	200	88.3%
Random	200	89.1%
Shape	200	82.1%
Color	200	92.1%
Movement category	200	86.3%

According to the results of the assistive aiming game test, we can draw a conclusion that use attention prediction module to predict human's attention state and preferences has high accuracy and great practicability. In addition, it can be concluded that different people have different preferences for objects with different characterization features. In the same situation, different people may have different attention intention, and the corresponding behavior intention also varies.

In the s hand-free interaction for social robot scenario, users can interact with the robot by their eye-gaze behaviors while their hands are busy. Compared with the traditional HCI pattern, the idea of predicting the human's behavior intention through their eyes is far more effective. According to the comparison results, the average operating time of our method is 5 to 6 times faster than the average operating time of the touch operation. Besides, the robot's interactive feedback has a 76% satisfaction according to the tester's judgements. From those results, we can conclude that use the eye-gaze based interaction as an assistive method for the traditional HCI patterns can significantly improve the HCI and HRI efficiencies.

In the nonverbal affective computing for social robot scenario, we mainly analyze the emotion content of the non-verbal information from human during the HRI process. With the support of the eye movement and head pose based affective computing module, nonverbal information can be combined with the verbal information to understand the user's emotional state in a better way. And can be further used to analyze the user's cognitive state.

6 CONCLUSION AND FUTURE WORKS

The method we introduce in this paper fully reflect the research and application potential for eye-gaze based cognitive and affective computing. These models not only can be used to predict the objects of concerns for people, but also can analyze the user's personal preferences in the psychological level. Furthermore, it can predict the behavioral intention for people during real-time interaction. At the same time, the eye-gaze and head motion patterns also carry emotional information which can be combined with other multimodal information in the affective computing for people. Our approach can be extended to the development of different interaction systems in order to optimize the level of human computer interaction. It can highly improve the efficiency of communication between human and robots. Furthermore, it can enhance the user experience from the level of mechanical interaction to the level of cognitive interaction. Also, our approach of intention prediction and emotion analysis extend the spectrum of eye-gaze involved interaction to the potential development in NUI systems. As an important factor in the study of the cognitive computing, the eye-gaze based learning and prediction models can improve computer-mediated interpersonal communication and exhibit high flexibility in uncertain or complex HCI/HRI environments.

ACKNOWLEDGMENTS

This work was supported by the Projects "The research of multimodal situated cognitive computing models based on eye-gaze interaction" by State Key Laboratory of Virtual Reality and Systems (No. BUAA-VR-15KF-09) and XIAMEN University President Foundation (No. 20720150081), and was partially supported by National Natural Science Foundation of China(Grant No.61502402) and State Key Laboratory of Virtual Reality and Systems(No.BUAA-VR-16KF-22).

REFERENCES

[1] Langton S R H. How the eyes affect the I: gaze perception, cognition, and the robot-human interface." IEEE International Workshop on Robot and Human Interactive Communication, 2001. Proceedings. IEEE, 2001:359-365.
[2] Maria Guarnera , Zira Hichy ,Maura I. Cascio ,Stefano Carrubba "Facial Expressions and Ability to Recognize Emotions From Eyes or Mouth in Children", Europe's Journal of Psychology, Vol. 11(2), 2015
[3] Lance B, Marsella S C. The Relation between Gaze Behavior and the Attribution of Emotion: An Empirical Study[M]// Intelligent Virtual Agents. Springer Berlin Heidelberg, 2008:1-14.
[4] https://www.tobii.com/group/news-media/press-releases/2017/8/tobii-and-microsoft-collaborate-to-bring-eye-tracking-support-in-windows-10/
[5] Vicente F, Huang Z, Xiong X, et al. Driver Gaze Tracking and Eyes Off the Road Detection System[J]. IEEE Transactions on Intelligent Transportation Systems, 2015, 16(4):2014-2027.
[6] Sundstedt V. Gazing at games:using eye tracking to control virtual characters." 2010:1-160.
[7] Borys M, Barakate S, Hachmoud K, et al. Classification of user performance in the Ruff Figural Fluency Test based on eye-tracking features[J]. 2017, 15(380):02002.
[8] Reading performance Using Eye Tracking to Assess Reading performance in Patients with Glaucoma: A Within-Person Study, Nicholas D. Smith, Fiona C. Glen, Vera M. Mönter, and David P. Crabb, Hindawi Publishing Corporation Journal of Ophthalmology Volume 2014, Article ID 120528, 10 pages
[9] Hofler M, Wesiak G, Purcher P, et al. Modern education and its background in cognitive psychology: Automated question creation and eye movements." International Convention on Information and Communication Technology, Electronics and Microelectronics. 2017:619-623.
[10] Swearingen T, Ross A. A Label Propagation Approach for Predicting Missing Biographic Labels in Face-Based Biometric Records[J]. 2017.
[11] Li Q, Li P, Zhang Q, et al. Icing load prediction for overhead power lines based on SVM." International Conference on Modelling, Identification and Control. IEEE, 2011:104-108.
[12] Buschjäger S, Morik K. Decision Tree and Random Forest Implementations for Fast Filtering of Sensor Data[J]. IEEE Transactions on Circuits & Systems I Regular Papers, 2017, PP(99):1-14.
[13] https://en.wikipedia.org/wiki/Kinematics
[14] Wade J, Zhang L, Bian D, et al. A Gaze-Contingent Adaptive Virtual Reality Driving Environment for Intervention in Individuals with Autism Spectrum Disorders[J]. Acm Transactions on Interactive Intelligent Systems, 2016, 6(1):3.
[15] Ruhland K, Peters C E, Andrist S, et al. A Review of Eye Gaze in Virtual Agents, Social Robotics and HCI: Behaviour Generation, User Interaction and Perception[J]. Computer Graphics Forum, 2015, 34(6):299-326.
[16] Bâce M. Augmenting human interaction capabilities with proximity, natural gestures, and eye gaze." The, International Conference. 2017:1-3.
[17] Zhao Q, Yuan X, Tu D, et al. Eye moving behaviors identification for gaze tracking interaction[J]. Journal on Multimodal User Interfaces, 2015, 9(2):89-104.
[18] Zheng R, Nakano K, Ishiko H, et al. Eye-Gaze Tracking Analysis of Driver Behavior While Interacting With Navigation Systems in an Urban Area[J]. IEEE Transactions on Human-Machine Systems, 2016, 46(4):546-556.
[19] Kern D, Marshall P, Schmidt A. Gazemarks:gaze-based visual placeholders to ease attention switching." Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2010:2093-2102.
[20] Yun S P, Outram B I, Tag B, et al. GazeSphere: navigating 360-degree-video environments in VR using head rotation and eye gaze." ACM SIGGRAPH. ACM, 2017:1-2.
[21] Springer. Journal on Multimodal User Interfaces[J]. 2015.
[22] Xiahou J, He H, Wei K, et al. Integrated Approach of Dynamic Human Eye Movement Recognition and Tracking in Real-time." International Conference on Virtual Reality and Visualization. IEEE, 2017:94-101.
[23] Kao T C, Sun T Y. Head pose recognition in advanced Driver Assistance System." IEEE, Global Conference on Consumer Electronics. IEEE, 2017:1-3.
[24] C. Z. Liu, H. Aliamani and M. Kavakli, "Behavior-intention analysis and human-aware computing: Case study and discussion," 2017 12th IEEE Conference on Industrial Electronics and Applications (ICIEA), Siem Reap, Cambodia, 2017, pp. 516-521. doi: 10.1109/ICIEA.2017.8282899
[25] Albert Mehrabian. Pleasure-arousaldominance: A general framework for describing and measuring individual differences in temperament. Current Psychology, 14(4):261–292, 1996.
[26] Brent Lance and Stacy C Marsella. The relation between gaze behavior and the attribution of emotion: An empirical study. In International Workshop on Intelligent Virtual Agents, pages 1–14. Springer, 2008.
[27] Albert Mehrabian. Pleasure-arousaldominance: A general framework for describing and measuring individual differences in temperament. Current Psychology,14(4):261–292, 1996.