The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

VR エージェントによるインタビュー対話の視線解析

Fuhui Tian [*1] Shogo Okada [*2] Katsumi Nitta [*3]

[*1] Tokyo Institute of Technology [*2] Japan Advanced Institute of Science and Technology [*3] Tokyo Institute of Technology

In this paper, we present a computational framework to analyze human eye movement using virtual reality system in job interview scene. First, we developed a remote interview system using virtual agents, and implemented the system into a virtual reality headset. Second, by tracking eye movement and collecting other multimodal data, the system can better analyze human personal traits in interview communication with virtual agents, and can better support training in people's communicating skills.

1. Introduction

There exist a large number of researches in non-verbal and verbal behavior fields to create the framework in automated performance prediction. I. Naim et al. [Naim 2016] collected prosody, facial and lexical features into regression models for training. Eye tracking data were not collected in the research above, so the score for eye contact get lowest regression coefficients due to their research. Lei Chen et al. [Chen 2016] created an automated scoring system for interview videos, for lack of direct gaze feature extraction, gaze directions are approximated through the Visage tracker SDK using head pose and eye rotation. There is still lack of research about eye movement for multimodal approach.

The purpose of this paper is to create a virtual reality based interview system and analyze human eye movement using this system. In human-agent interaction, human emotions while interacting with agents are high-level personally traits that quantifies human attitudes, intentions, motivations, and behaviors. In addition, the Virtual-reality space provides a chance to interact with virtual agents in a more immersive way. In the field of user interface, the focus depth from gaze based interaction technique as input modality showed significant difference in performance in VR game [Yun 2016]. So in this research, we devoted to combine social behavior assessment system using present modeling methods with an eye tracking system in VR. The intention is to explore the relationship between human gazing modality and social behavioral traits.

2. Eye Tracking System

2.1 The virtual environment built by Unity

To create the interview scenario, we use Unity 5.5 engine to build the virtual environment. Three 3D characters were set into an office room representing interviewers. One interviewer sits in front of main camera as chief interviewer, other two sit beside him. Under these three 3D characters, a Mecanim animation system was attached to each of their body to perform necessary movements required in interview scenario. Animations were controlled by state machines. All three 3D characters' models have a rigid body for gaze detection. Two cameras were set to render both left and right eye's view in the experiment.

Contact: Fuhui Tian, Tokyo Institute of Technology,
TEL: 07010679189, Email: tianfuhui@gmail.com

2.2 FOVE Eye Tracking System

Unlike previous VR headsets that require customization with eye trackers, FOVE HMD comes with native eye tracking system that is free to customization. In virtual reality space, two cameras render images for people's eyes. Gaze detection is calculated by matching the black pupil ellipse in each video frame and computing its center. Mapping of gaze to screen coordinates is performed by fitting a polynomial equation in 2D space [Lutz 2017].

2.3 Collision Detection for Eye Movement Feature Acquirement

To understand what people are looking at, we use Unity's collision detection method to find gazed objects in interview scene. People's sights can be established by calculating the vector originates form camera.

For each frame rendered by engine, if the ray of gaze directions hit 3D characters' rigid body, an OnCollisionEnter message will be generated. Our system catches this message and computes those into gaze features. In our 3D scene, two spheres in red and green indicates left eye and right eye sights respectively. These rendered images were also captured by top of cameras in scene and stored in order to do data annotation.

3. Interview Setup and Data Collection

10 graduate students in job seeking were invited to the experiment. Participant sit in front of the desk and interacts with virtual agents, 11 questions of general question in job hunting were asked by agent through the scenario, and their eye movement data is recorded. Overall interview videos taken aside are also recorded for data annotation.

These 11 questions covered the aspect from general, personality to visions. At last part of interview we added three frequently asked questions in stress interviews, for further investigation of people's behavior. A corpus of data organized as [True/False, True/False] of each frame is logged to help us better understand participants' behavioral traits. For example, if the head of agent were being looked at, and body is not, the record would be [True, False]. We are convinced that whether an appropriate gazing at interviewer is a valuable clue to comprehend human gaze, also in interview evaluation.

Totally 10 participants were invited to the experiment, the average length of the interview section is 6 minutes 36 seconds. To explore the relationship between people's eye movement and interview performance, we extract gaze features from each frame. All data set were divided into 11 parts by questions, so we gain 110 data clips. Duration of clips ranges from 12 seconds to 1 minute 35 seconds. We computed 27 gazing features. Gazing object change is computed by gazing state transformation, which means gaze target was changed between head, body and outside of the agent's rigid body within two frames. For example, if gaze flag changes form [True, False] to [False, True], this means interviewer's gaze looked down once between two frames. We call this feature TF FT for short.

4. Results of Interview Evaluation

First, we analyze the quality and reliability of ratings by observing how well raters agree with each other. In our result, two raters have high agreement in rating field such as Total and Eye contact, and low agreement in engagement. Next, we present the prediction accuracies for the trained regression models based on automatically extracted features. Finally, we analyze the weights of the features in our trained models, and quantify the relative importance of the behavioral features to overall interview performance.

4.1 Predictions using Gaze Features

We trained all eye movement features into four regression models and compute its accuracies in each rating category. We used the same set of features, and the model automatically learned the feature weights. Table 1 shows the prediction accuracy of each category of scores under unimodal data collection.

As observed in Table 1, unimodal eye movement features gained highest score of 73.6 among other rating categories. From this clue, we can draw the conclusion that eye movement features plays an important role in predicting people's social traits in eye contact. Our modal showed acceptable accuracy value for single modality of eye movement in prediction of eye contact and total performance in job interviews. We also noticed that the accuracy of Total is still at medium stage at 59.3, which indicates that eye movement plays a role in concluding people's total social traits.

4.2 Feature analysis

To get importance of each input values in the regression model, we further calculated predictor importance using IBM SPSS Modeler. IBM SPSS uses the sum of squares for residual (error variance) as a signal regarding predictor importance. To the extent that a predictor is important in the model, remove this feature would produce a considerable increase in the residual sum of squares.

In our system, Feature FF FF gets highest score of 0.5, so we can conclude that if participants do not gaze at interviewer for a while, that shows participants are not good at using eye contact to communicate, and their interview score of eye contact will be low.TT FT: this feature is logged if interviewee's sight moved to body form gazing at interviewer's head. This would mean gaze goes down by once. The feature has a negative correlation with eye contact score. So we can conclude that the more one person looked down, his/her score is going down.

Field	**Accuracy in percentage**
Total	59.3
Engagement	38.7
Eye Contact	73.6
Friendliness	56.6
Logical, clear presentation	47.5

Table 1. Prediction Accuracy using gaze features

5. Conclusion

The key contributions of this paper can be summarized as follows. Firstly, we report on a new virtual reality based interview system that supports the use of Social Signal Processing in interview performance evaluation. To our knowledge, this is the first research effort collecting eye movement features from a virtual reality space to structured interview questions. Secondly, automated assessment of interview performance using these eye movement features was introduced to the multimodal research area for the first time. Our model showed acceptable accuracy value for single modality of eye movement in prediction of eye contact and total performance in job interviews. Our experimental results suggest that this method is not only conceptually easy to understand, but also shows promising empirical results.

References

[Naim 2016] Naim, I., Tanveer, M. I., Gildea, D. & Hoque, E. Automated Analysis and Prediction of Job Interview, Performance IEEE Transactions on Affective Computing, 2016, PP, 1-1. DOI: https://doi.org/10.1109/FG.2015.7163127

[Chen 2016] Chen, L., Feng, G., Leong, C. W.; Lehman, B., Martin-Raugh, M., Kell, H.; Lee, C. M. & Yoon, S.-Y. Automated Scoring of Interview Videos Using Doc2Vec Multimodal Feature Extraction Paradigm, Proceedings of the 18th ACM International Conference on Multimodal Interaction, ACM, 2016, 161-168. DOI: https://doi.org/10.1145/2993148.2993203

[Yun 2016] Yun Suen Pai, Benjamin Outram, Noriyasu Vontin, and Kai Kunze. 2016. Transparent Reality: Using Eye Gaze Focus Depth as Interaction Modality, In Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16 Adjunct). ACM, New York, NY, USA, 171-172. DOI: https://doi.org/10.1145/2984751.2984754

[Lutz 2017] Lutz, Otto Hans-Martin, Charlotte Burmeister, Luara Ferreira dos Santos, et al, Application of head-mounted devices with eye-tracking in virtual reality therapy, Current Directions in Biomedical Engineering, 3.1 (2017): 53-56. Retrieved 25 Jul. 2017, DOI: https://doi.org/10.1515/cdbme-2017-0012