The Thirty-Second AAAI Conference
on Artificial Intelligence (AAAI-18)

Lookine: Let the Blind Hear a Smile

Yaohua Bu,[12] **Jia Jia,**[1*] **Yuhan Tang,**[1] **Xuan Zhang,**[1] **Tianyu Gao**[1]
[1]Department of Computer Science and Technology, Tsinghua University, Beijing, China
Tsinghua National Laboratory for Information Science and Technology (TNList)
Key Laboratory of Pervasive Computing, Ministry of Education
[2]Academy of Arts & Design, Tsinghua University, Beijing, China
byh15@mails.tsinghua.edu.cn, jjia@mail.tsinghua.edu.cn

Abstract

It is believed that nonverbal visual information including facial expressions, facial micro-actions and head movements plays a significant role in fundamental social communication. Unfortunately it is regretful that the blind can not achieve such necessary information. Therefore, we propose a social-assistant system, Lookine, to help them to go beyond this limitation. For Lookine, we apply the novel techniques including facial expression recognition, facial action recognition and head pose estimation, and obey "barrier-free" principles in our design. In experiments, the algorithm evaluation and user study prove that our system has promising accuracy, good real-time performance, and great user experience.

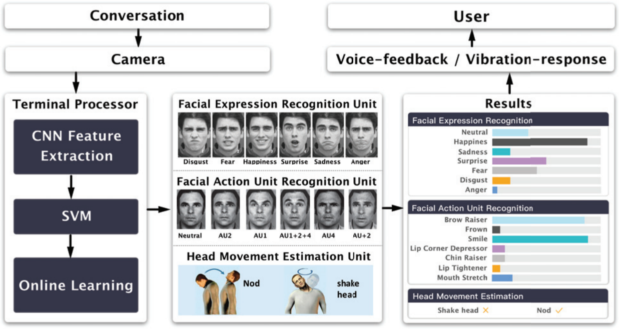

Figure 1: The working mechanism of Lookine

Introduction

A series of studies have suggested that nonverbal information plays a significant role in many cases (Watanabe et al. 2013). For example, it is nonverbal communication referring to facial micro-expression and body gestures that serves as a primary means of not only organizing interpersonal interactions, but also conveying cultural values and building friendship. Unfortunately, for people who have lost their sight, they can only depend on their hearings to get very limited information, which leads to much more misunderstanding or biased judgment in social communication.

According to our survey, only a few institutes start to pay some attention to social-assistant for the blind, e.g., a newest project called "Seeing AI" established by Microsoft. However, what they realize is just to take a photo of somebody and then recognize several basic facial expressions, which is not synchronously communication-required. Consequently, we propose, "Lookine", which is a social-assistant system specifically for the blind (see Fig.1).

We combine novel techniques including facial expression recognition (Valstar et al. 2017), facial action units recognition (Baltrušaitis, Robinson, and Morency 2016), and head movement estimation (Yan et al. 2016) to make Lookine can truly help users to obtain necessary nonverbal visual information in communication in real time. Furthermore, in the design, we obey "barrier-free" principles to make sure that Lookine can provide users with flexible interaction styles, e.g., people can freely select Voice-Feedback or Vibration-Feedback when in use.

Main contributions of this work are summarized as follows: (1) Lookine is a tool which not only can help the blind obtain necessary nonverbal visual information in social communication, but also is easy for users to use and adapt with to develop nonverbal social skills in daily life. (2) In technical implementation, we establish Lookine as a online system to guarantee promising real-time performance.

Technical Implementation

Feature Extraction

First, we use a simple CNN to detect a valid face in a given picture. Once a valid face is successfully detected, the network could also make the face aligned for the feature extraction. To reduce the dimensionality of features, we use a PCA model, which leads to a reduced basis of 1391 dimensions. And these feature vectors will be used in facial action unit recognition. We use Conditional Local Neural Fields (CLNF) to implement facial landmark detection, which uses advanced point distribution model to capture landmark shape variations.

Facial Action Unit Recognition

Facial Action Coding System(FACS) is a system to taxonomize human facial movements by their appearance on faces (Dimberg, Thunberg, and Elmehed 2000). In our work, we train the model based on this to recognize facial movements

[*]Corresponding author: J.Jia (jjia@mail.tsinghua.edu.cn)

Table 1: The recognition accuracy of our algorithm on some typical units

	Precision	Recall	F1-Measure
Happiness	0.90	0.79	0.84
Disgust	0.91	0.42	0.57
Surprise	0.92	0.96	0.94
Nodding	0.93	0.58	0.72
Head Shaking	0.92	0.50	0.65
AU4	0.70	0.79	0.75
AU15	0.92	0.50	0.65
AU26	0.56	0.75	0.64

Table 2: The delay of our algorithm on different tasks

Recognition Tasks	Head Movement	Action Units	Facial Expression
Delay (sec)	0.167	0.167	1.134

specified as 15 facial action units (AUs), e.g., brow raiser (AU1,AU2), cheek raiser (AU6), etc. With a combination of extracted features above, we use models based on SVM to recognize facial action units. In details, we take a strategy of learning online and use the median value of all feature vectors as the feature of normal face.

Head Movement Estimation

We extract head pose from the facial landmark feature. In CLNF model, the 3D representation of facial landmarks are projected to the image using orthographic camera projection. Thus we can accurately estimate the head pose including the translation and orientation by solving a PnP problem. Here, n is the number of facial landmarks, and head pose is stored in the following format: (x, y, z, rx, ry, rz), in which (x, y, z) is the translation with respect to camera centre, and (rx, ry, rz) is the rotation in radians around X, Y, Z axes. We mainly apply the FSM method, in which we define"vertical movement" as several consecutive frames in which rx changes in the same direction. And then a head nod formed by another three or more vertical movements with the style of "up and down" alternately, can be detected.

Experiments

Algorithm Performance We invited 30 volunteers to participate in the experiment of recognition accuracy. Each participant was asked to perform specified 15 AUs, 6 facial expressions and 2 head movements for 3 times. And the results of the recognition are as shown in Table 1, which proves that the performance of our algorithm is promising. For real-time capability, we record the time from taking images through the camera to obtaining recognition results. The results can be seen in Table 2, in which we can see that the delay of the recognition of AUs and head movement are both within 0.2 seconds.

User Study This part aims to get a comprehensive investigation of the function satisfaction and user experience of our product. And we invited 40 blind people which had an uniform distribution of age and includes congenitally blind and who became blind later in life from a professional blind-helping center. Then two groups of comparison experiments and relevant questionnaires were designed for qualitative and quantitative analysis. In the first group, testers were asked to take a five-minutes' dialogue with the observer, and a video was recorded at the same time. Then they re-heard the video with Lookine's support, in which Lookine analysed the visual concepts and delivered results in Voice-Feedback. In the second group, testers were asked to take two dialogues with observers, in which Voice-Feedback and Vibration-Response were taken as the interaction style respectively. As analysed, results were very promising because 36 participants said they really need this kind of product. And the average degree of satisfaction for Lookine's experience is 4.2 according to the computation (where 0 is the lowest and 5 is the highest). Furthermore, their preferences of different interactive modes give us a solid support to choose the most proper interaction methods, e.g., Voice-Feedback.

Conclusion

In this paper, we proposed Lookine which can effectively help the blind to obtain necessary nonverbal visual information including facial expressions, facial actions and head movements to achieve a better social communication. In the future, we will deploy it on suitable wearable devices.

Acknowledgments

This work is supported by National Key Research and Development Plan $(2016YFB1001200)$, the Innovation Method Fund of China $(2016IM010200)$, and the National Natural and Science Foundation of China $(61370023, 61602033)$.

References

Baltrušaitis, T.; Robinson, P.; and Morency, L.-P. 2016. Openface: an open source facial behavior analysis toolkit. In *Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on*, 1–10. IEEE.

Dimberg, U.; Thunberg, M.; and Elmehed, K. 2000. Unconscious facial reactions to emotional facial expressions. *Psychological science* 11(1):86–89.

Valstar, M. F.; Sánchez-Lozano, E.; Cohn, J. F.; Jeni, L. A.; Girard, J. M.; Zhang, Z.; Yin, L.; and Pantic, M. 2017. Fera 2017-addressing head pose in the third facial expression recognition and analysis challenge. *arXiv preprint arXiv:1702.04174*.

Watanabe, T.; Yahata, N.; Kawakubo, Y.; Inoue, H.; Takano, Y.; Iwashiro, N.; Natsubori, T.; Takao, H.; Sasaki, H.; Gonoi, W.; et al. 2013. Network structure underlying resolution of conflicting non-verbal and verbal social information. *Social cognitive and affective neuroscience* 9(6):767–775.

Yan, Y.; Ricci, E.; Subramanian, R.; Liu, G.; Lanz, O.; and Sebe, N. 2016. A multi-task learning framework for head pose estimation under target motion. *IEEE transactions on pattern analysis and machine intelligence* 38(6):1070–1083.