Conversation Aid for People with Low Vision Using Head Mounted Display and Computer Vision Emotion Detection

Rafael Zuniga and John Magee(✉)

Clark University, Worcester, MA 01610, USA
{RZuniga,jmagee}@clarku.edu

Abstract. People with central vision loss may become unable to see facial expressions during face-to-face conversations. Nuances in interpersonal communication conveyed by those expressions can be captured by computer vision systems and conveyed via an alternative input to the user. We present a low-cost system using commodity hardware that serves as a communication aid for people with central vision loss. Our system uses a smartphone in a head-mounted-display to capture images of a conversation partner's face, analyze the face expressions, and display a detected emotion in the corner of the display where the user retains vision senses. In contrast to purpose-built devices that are expensive, this system could be made available to a wider group of potential users, including those in developing countries.

Keywords: Accessibility · Computer vision · Low-vision
Head-Mounted-Display

1 Introduction

It is estimated that 285 million people are visually impaired worldwide according to the World Health Organization. Many existing interfaces that aim to help these individuals do not take into consideration that 246 million of them are categorized as having low vision. Macular degeneration is the leading cause of vision loss. This disease is caused by the deterioration of the central portion of the retina resulting in central vision loss but leaving a certain amount of peripheral vision intact. This means that we can take advantage of this little vision they have left to communicate information to them instead of immediately resorting to sound or tactile feedback. One common problem that arises from being visually impaired is not being able to see people's facial expressions as they are having a conversation with someone; leaving out many visual cues that out brain uses to have a more complete interpretation of what a person says. A rough simulation of the challenge posed by macular degeneration is shown in Fig. 1.

Here, we present one approach in which we could take advantage of the little vision a visually impaired person has left to mitigate the common problem described above. Performing Emotion Recognition Analysis with computer

© Springer International Publishing AG, part of Springer Nature 2018
K. Miesenberger and G. Kouroupetroglou (Eds.): ICCHP 2018, LNCS 10897, pp. 44–50, 2018.
https://doi.org/10.1007/978-3-319-94274-2_7

Fig. 1. Two images showing a rough simulation of how macular degeneration could obscure a large portion of the center of a person's vision, while still providing some visual information on the peripheral edges. In the top image, taken from a popular internet meme, a person is showing a strong emotion of joy. In the bottom image, the person's face is entirely obscured. Image credit: memegenerator.net, New Line Cinema

vision, we can computationally analyze the facial expressions of the person the visually impaired user is talking to and categorize them into emotions such as joy, sadness, anger, etc. We use a Head-Mounted-Display (HMD) to display the information right in the part of the eye where the user can still see. Depending on the specific user, we could use preset colors or symbols that map to a specific feeling so that the user knows what expression or emotion the person they are talking to is conveying as they say something.

An example of one of the few ventures that have tried taking advantage of this peripheral view is oxSight (Fig. 2 Left). They created their own customized Head-Mounted-Displays which performs adequately but is too expensive (Approximately $15,000). Another device, the OrCam MyEye 2.0 has a retail cost of approximately $3500 and produces audio output. 90% of the people that are visually impaired live in developing countries where this technology is completely unattainable because of financial reasons. Thanks to the rise of augmented and virtual reality, there now exists cheap Head-Mounted-Display that can have any mobile phone attached to it (Fig. 2 Right). Even in developing countries, android phone devices are widely available at prices that are affordable to the people. The phone then runs an application that uses an Emotion Recognition Analysis API that processes the frames captured from the phone's camera. This allows the phone to become a facial expression recognizer that categorizes them into its appropriate emotion. Once the phone is attached into the Head-Mounted-Display, we have the whole screen available to give information to the user directly into their available vision.

Several wearable and head-mounted technologies have been proposed as assistive devices blind or visually impaired individuals [4]. There are related work with conversation aids for people with autism [2] or aphasia [3]. Adoption of head-mounted assistive technologies nevertheless faces challenges due to the highly intrusive nature and perceived stigma [1]. Nevertheless, we hope that by developing a commodity system that is affordable to people in developing countries, this work has potential to positively impact people's lives.

2 Systems Overview

A flowchart of our system is shown in Fig. 3. A smartphone device is placed within a head-mounted display that contains a head strap and some lenses to hold the phone a short distance in front of the user's eyes. The HMD device does not contain any electronics of its own, so it is relatively inexpensive. The input to the system is the camera built into the back of the smartphone. The camera send a video stream to the computer-vision based face analysis program. Our current system is using Affectiva [5] to detect facial pose and emotion prediction scores. The output from Affectiva is then mapped to output values that the system wants to communicate to the user. These output values are sent to the user-interface generation module which produces video with the emotion and engagement encoded into a user-defined area of the display. This video is displayed on the screen of the smartphone which, as described before, is attached into the head-mounted display device.

Fig. 2. Top: OxSight's smart specs. Image credit: OxSight. Bottom: A $20 Head-Mounted-Display unit for a smartphone.

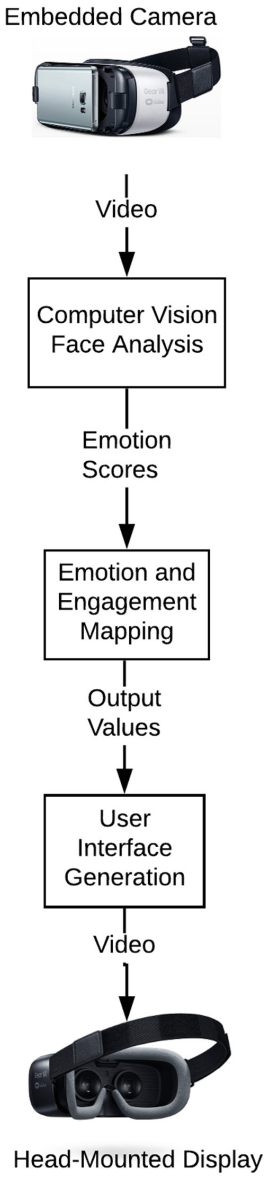

Fig. 3. System Flowchart. The core of the system is a smartphone placed within a head-mounted display. The phone's camera captures live video of the conversation partner. The video is processed by the computer vision system to produce emotion scores, which are then mapped to output values. The user interface generates video to be displayed on the screen in a location that the user still has visual acuity.

The current system recognizes smiles, anger, joy, and sadness. The head pose information is mapped to an attention score. The attention score is highest if the face in the video is looking toward the camera, whereas the score decreases if the face turns away from the camera. This score is intended to convey if the person is looking at their conversation partner. Smiling, Anger, Joy, and Sadness can be mapped to user-defined colors. The intensity of the color reflects the attention score. A screenshot of the prototype display is shown in Fig. 4.

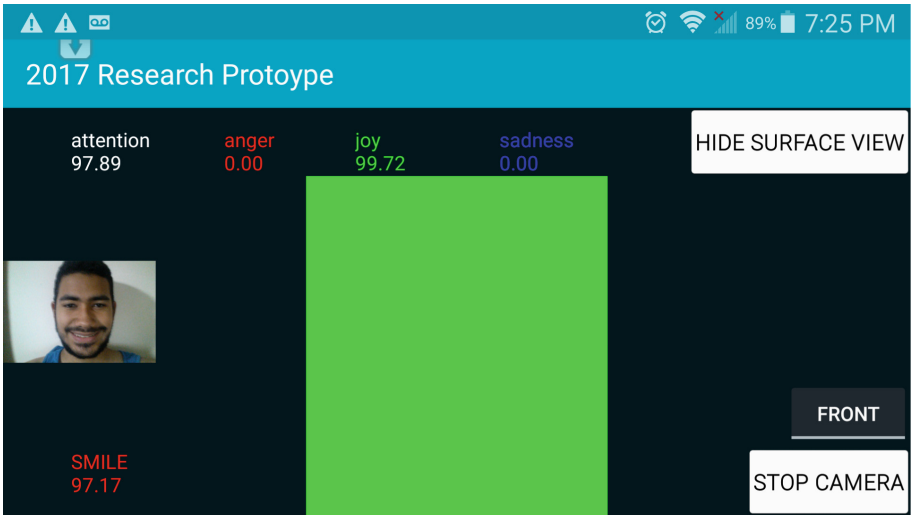

Fig. 4. Prototype system user interface displayed on a smartphone screen. (Color figure online)

3 Preliminary Evaluation

We conducted a preliminary evaluation of the system with two users who do not have visual impairments. The goal of the preliminary evaluation was to determine if the system functioned as intended. The participants took turns wearing the head-mounted display running our system. The display had the live-video feed disabled so that the person could not see the face of their conversation partner. The conversation partner would attempt to convey facial expressions of Smile, Joy, Anger, and Sadness. They would also turn their head away from the conversation at various times. At this stage, we did not yet perform a quantitative evaluation, however, in the opinion of the participants, the system was able to deliver a color-encoded emotion and attention score most of the time.

As a preliminary measure of success, the participants were able to discern the face expression emotion of their conversation partner while using the device.

4 Conclusions and Future Direction

Our proposed system successfully combines together existing techniques in a commodity device that can be deployed inexpensively to a wide range of users, including those in developing countries. Purpose-built devices may be expensive and therefore unattainable to those with economic disadvantages. Even in developing countries, relatively powerful smartphones are obtainable and affordable to the general population. By basing our system on such commodity hardware, a software-based assistive technology solution as a conversation aid becomes practical.

The most significant downside of our approach is the overly intrusive nature of the head-mounted-display with smartphones. The purpose-built devices can be made to look more like glasses, and therefore potentially reduce the stigma of using the technology.

We plan to conduct an evaluation of the system with several participants that will provide empirical evaluation and quantitative results as well as a more robust qualitative evaluation for a future paper. We also hope that we will be able to include an evaluation and feedback by users with macular degeneration that would benefit from the technology.

Acknowledgements. The authors wish to thank the Clark University undergraduate summer research program for funding and mentoring this research project. We also wish to thank the program's anonymous sponsor.

References

1. Profita, H., Albaghli, R., Findlater, L., Jaeger, P., Kane, S.K.: The AT effect: how disability affects the perceived social acceptability of head-mounted display use. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pp. 4884–4895. ACM (2016)
2. Boyd, L.E., Rangel, A., Tomimbang, H., Conejo-Toledo, A., Patel, K., Tentori, M., Hayes, G.R.: SayWAT: augmenting face-to-face conversations for adults with autism. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pp. 4872–4883. ACM (2016)
3. Williams, K., Moffatt, K., McCall, D., Findlater, L.: Designing conversation cues on a head-worn display to support persons with aphasia. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pp. 231–240. ACM (2015)
4. Rector, K., Milne, L., Ladner, R.E., Friedman, B., Kientz, J.A.: Exploring the opportunities and challenges with exercise technologies for people who are blind or low-vision. In: Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility, pp. 203–214. ACM (2015)
5. Affectiva. https://www.affectiva.com/. Accessed 1 Feb 2018