A Google Glass App to Help the Blind in Small Talk

M. Iftekhar Tanveer
ROC HCI
Computer Science
University of Rochester
itanveer@cs.rochester.edu

Mohammed E. Hoque
ROC HCI
Computer Science
University of Rochester
mehoque@cs.rochester.edu

ABSTRACT
In this paper, we present a wearable prototype that can automatically recognize affective cues such as number of people present, their age and gender distributions given an image. We customize the prototype in the context of helping people with visual impairments to better navigate social scenarios. Running an experiment to validate this technology in real life situations remains part of our future work.

Categories and Subject Descriptors
K.4.2 [**Social Issues**]: *Assistive Technologies for persons with disabilities*
H.5.1 [**Multimedia Information Systems**]: *Artificial, augmented, and virtual realities.*

General Terms
Algorithms, Design, Human Factors

Keywords
Google Glass; Face; Age; Gender; Accessibility; Blind

1. INTRODUCTION
Ever faced a situation in which you had to initiate small talk with someone you barely knew? How did you start? In a one-on-one conversation, many would commonly begin by inquiring about the weather. However, starting a conversation can be tricky when multiple people are present, as attempts to initiate small talk may be viewed as impolite or even intrusive. People with sufficient social skills are able to pick up on social cues (e.g. number of people, their age and gender distribution, choice of clothes, locations, etc.) to initiate conversation within a group. The ability to successfully launch small talk is an important social skill that may lead to meaningful discussions and new social relationships.

Now imagine someone with visual impairment initiating small talk with a random individual or a group of people. How would this impairment limit that person's social skills and his/her ability to form new relationships? In this paper, we demonstrate a wearable prototype (Google Glass) that is able to automatically sense and synthesize information about people within conversational distance using computer vision and machine learning techniques. Visually impaired individuals can take pictures of a scene by pressing a button or using the double-tap gesture. The picture is then uploaded to the cloud or to a nearby local computer to automatically analyze its content. Our current implementation of the framework allows for automated analysis of number of people present, their approximate age and gender given an image. Our algorithm then retrieves and synthesizes the information to the user using a text-to-speech engine (Figure 1). For example, given the information that four people are present — two males and two females with respective ages 15, 17, 22, and 19, our app can synthesize the following speech feedback:

``Group of 4. Age range 15 to 22. Average age 18. 50% male''

ASSETS '14, Oct 20-22 2014, Rochester, NY, USA
ACM 978-1-4503-2720-6/14/10.
http://dx.doi.org/10.1145/2661334.2661338

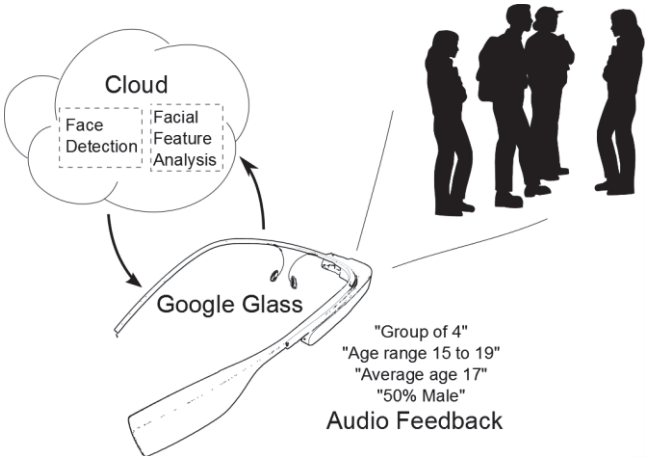

Figure 1. Main idea of the prototype: Picture taken by Google Glass is transmitted and analyzed in the cloud. The Glass gives speech feedback based on the analysis.

2. RELATED WORKS
In the past, accessible technologies have often been designed as self-contained, separate devices, such as money detectors, color detectors, water level detectors, etc. As a result, users had to carry all these devices with them which limited their ubiquity and accessible functionalities. The notion of social stigma also accompanied the usage of these devices. Kane et al. [1] showed that the use of devices for a more general user group (e.g. smartphone) attracts users with reduced cost, increased mobility, improved usability, and less social stigma. As a result, we opted to use the publicly available Google Glass[1] as the medium of our prototype.

Given the difficulty of automated labeling of objects, systems such as VizWiz [2] were introduced to use crowdsourcing techniques for accurate labeling of information in an image to assist blind users. While this solution of using crowd workers to label images is effective, it has several limitations. For one, relying on a crowd worker to label an image may not consistently yield nearly real-time solutions. Also, given that the system requires competent workers who desire compensation in return for their services, large-scale deployment could be expensive. Moreover, privacy concerns may arise, given that the pictures are being shared with random Turkers. In our prototype, we addressed

[1] https://www.google.com/glass/start/

these limitations by relying on automated computer vision techniques.

The most relevant work on giving blind users feedback on affective cues was performed by Tanveer et al. [3], Rèhman et al. [4], and Rahman et al. [5]. Tanveer et al. [3] reported that along with facial expressions, visually impaired individuals would also benefit from information related to the identity of a person. For example, such individuals would prefer to receive information about the interlocutor's age, gender, ethnicity, and height. In our prototype, we provide feedback on the number of people present, their approximate age and gender distribution as an early proof of concept. Providing feedback on other cues (clothes, ethnicity, height etc.) remains part of our future work.

3. DESIGN CONSIDERATIONS

In the following section, we describe some design considerations and issues we faced while designing the prototype.

3.1 Vision Framework

For computer vision methods of detecting age and gender, we used a library named SHORE — Sophisticated High-speed Object Recognition Engine [6]. SHORE detects and analyzes the pictures of faces present within an image. The framework extracts features related to age and gender, for example, and provides a statistical summary of pixel intensities associated with beard, mustaches, and wrinkles on the face. A boosting algorithm [7] is used to train and use a bunch of classifiers for detecting age and gender from ground truth information based off these pictures. More details as well as limitations and benchmark performances are available in [6].

Accuracy of SHORE for detecting facial features is within the range of 92% to 94% on BioID[2] and CMU+MIT[3] databases. SHORE achieves a gender classification accuracy of 94.3% on BioID and 92.4% on FERET dataset [8]. Age recognition rate is 95.3% on JAFFE database. SHORE is also faster in comparison to its competitors. It takes only 9ms to detect a face in a 384 x 286 pixel image on an Intel Core 2 Duo 6420 CPU. As SHORE is trained on a European dataset, it is possible bias may exist on some specific cultural components. With the introduction of more datasets, this bias could be addressed in future.

3.2 Heating Issue with Google Glass

Google Glass has an inherent limitation of overheating in the case of continuous processing. As it is purposed as a low powered device, it does not have any heat dissipation mechanism. As a result, executing too many computations can render the app nonresponsive [9]. Also, overheating causes the device to become uncomfortable to use.

Due to this limitation, our framework uses a light CPU load only to establish connection with a remote computer. Once the user double taps or presses a button to take a picture, the Glass starts transmitting the image to the remote computer for further processing. The ability to process a continuous video stream may be possible in the future with further enhancement of the Google Glass framework.

[2] http://www.bioid.com/downloads.html

[3] http://vasc.ri.cmu.edu/idb/html/face/frontal_images/

4. FEEDBACK DESIGN

We are currently using a Text-to-Speech feedback scheme to describe the faces within the picture. Running an informed user study to iterate through possible feedbacks remains part of our future work. The feedback structure is dynamically based on the number of faces it detects. When it detects only a few faces in the picture, it describes the primary features using Android's Text-to-Speech functionality. Otherwise, when the system spots more than three faces in the input, it describes the summary statistics associated with the group.

5. FUTURE WORK

In our future study, we plan to conduct interviews with both blind and sighted people to gain a thorough understanding on the specific cues that people rely on for initiating small talk. Based on these insights, we will add more features to the prototype. Additionally, we seek to run a participatory design technique with a small focus group of blind users to design a more comprehensive feedback system. Finally, we will run a qualitative study by freely distributing the application to a community of blind people. Following this, we would collect feedback on their use.

6. CONCLUSION

This paper presents a wearable prototype that can sense and synthesize affective information about people in a group (number of people, age and gender distribution). We envision validating this prototype in the context of helping individuals with visual impairments in order to help initiate small talk and make them more comfortable in groups. In the future, we plan to expand the framework so that visually impaired individuals may navigate conversations given real time feedback.

7. REFERENCES

[1] S. Kane et al. 2009 Freedom to roam: a study of mobile device adoption and accessibility for people with visual and motor disabilities In *Proceedings of ASSETS*

[2] Bigham, J., et al. 2010 VizWiz: nearly real-time answers to visual questions." *Proceedings of the 23rd UIST.*

[3] M. Tanveer et al. 2013 Do you see what I see? Designing a sensory substitution device to access non-verbal modes of communication. In *Proceedings of ASSETS.*

[4] S. Rèhman et al. 2010 ifeeling: Vibrotactile rendering of human emotions on mobile phones. *Mobile Multimedia Processing*

[5] Rahman, AKM et al. 2012 IMAPS: A smart phone based real-time framework for prediction of affect in natural dyadic conversation. *Visual Comm. & Image Processing (VCIP)*

[6] Ruf, T. et al. 2011 Face detection with the sophisticated high-speed object recognition engine (SHORE). *Microelectronic Systems.*

[7] Freund Y et al. 1999 A Short Introduction to Boosting, In *Journal of Japanese Society for Artificial Intelligence*

[8] Phillips, P et al. 1998 The FERET database and evaluation procedure for face-recognition algorithms. *Image and vision computing* 295-306

[9] Chen et al, 2014. Towards wearable cognitive assistance. *Mobile systems, applications, and services MobiSys*