Anon-Emoji: An Optical See-Through Augmented Reality System for Reducing Appearance Bias in Social Interactions

Ran Sun[1] Harald Haraldsson[1] Yuhang Zhao[1,2] Serge Belongie[1,2]

[1]Cornell Tech [2]Cornell University

Abstract

Facial expressions are an important part of human communication. At the same time, faces also reveal aspects of our identities – ethnicity, gender, or even sexual orientation – that can surface biases held by our counterparts and lead to downstream inequalities in our social interactions. However, we do not have the option to entirely change what we look like, or at least hide the identities that our faces might expose. As Optical See-Through (OST) Augmented Reality (AR) headsets possess the advantages of near-eyes display and better depth alignment between virtual renderings and the environment, we decided on an OST AR approach for the solution. In this paper, we present a system designed for OST AR headsets that occludes the subject's facial features with an emotion-presenting emoji model in 3D space.

1. Introduction

1.1. Background and Motivation

Physical appearance generates many biases in different social scenarios. Research in social psychology has shown that face appearance [13], for example attractiveness, often mediates hiring decisions [9] [8]. Biases caused by ethnicity, gender and other identity-related information have also resulted in workplace discrimination [5]. Recent work has shown the problem is especially acute for African-American ride-sharing or hospitality service providers [7].

Much of our identities can be easily obtained from our faces. Currently it is impossible to occlude our looks entirely in situations where identity information is unnecessary or even detrimental. At the same time, face-to-face social interaction provides important context beyond personal identity: for example up to 70 percent of communication during in-person interactions happen non-verbally [14]. Therefore, concealing people's faces during interactions could cause great deficiency. Since human faces are extremely expressive, we would want to preserve such information when wiping out people's identity from their appearance.

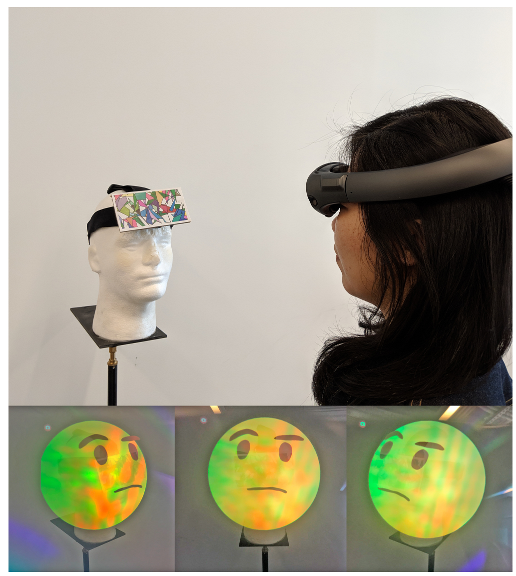

Figure 1. A viewer wearing the Magic Leap One headset facing a subject with a wearable image marker on their head. Below are the views captured through the AR headset - an emoji model rendered around the subject's head. Color variations are caused by the device and lighting, and are not noticeable to users.

Emojis are widely used in text-based communications, mainly to compensate for the loss of facial expression usage during virtual communication [6]. With its increasing usage in social media to hide people's faces in photos, we see a natural utility of emojis in creating anonymity. This inspired us to use emojis to bring back people's facial expression while concealing their faces for anonymity purposes.

Motivated by this, we explored a technological solution that could occlude people's physical appearances while preserving the other important elements of a social interaction. Our work could serve as a low-fidelity version of generative deep learning-based approaches to content manipulation, such as Deepfake or GAN [12], in AR settings. In addition to hiding people's identity, we have the intention to replace their faces with actual human faces of another generated identity. Therefore, in the viewer's perspective, they are still interacting with a real person that preserves full facial expressions and facial muscle movements. However, all unnecessary identity information could be left out. Emoji models are a generic, easily-understood face models that we use to demonstrate the idea.

1.2. Advantages of OST AR headsets

The advantages of OST AR displays make them ideal for this solution. Compared with handheld AR displays, head-mounted displays is a type of near-eyes display that imposes more control over what the users can see. With AR headsets, users no longer have to consume the virtual contents through handheld digital screens with limited field of view. In addition, virtual contents are better embedded in reality, enhancing the immersiveness of the experience.

Among AR Head-mounted devices, there are generally two different display technologies that superimpose virtual renderings with the environment in reality [4], which resulted in the following two types of display methods: Video See-Through (VST) display and Optical See-Through (OST) display. Compared to VST head-mounted display, OST display allows users to look into the reality through optical lenses while making use of optical combiners to present contents [4]. This guarantees a better alignment of virtual contents and reality not only on a pixel-precise basis but also on depth-precise basis [4]. However, the VST display does not possess such advantages.

Although the general intention of AR is to add objects to a real environment, it can also remove objects [10] [3]. Derived from AR, the concept of Diminished Reality (DR) describes the process of concealing, altering, and seeing-through objects in reality [11]. However, occluding real world objects is not one of the strengths of OST Augmented Reality displays. Compared to OST AR displays, VST Augmented displays possess much more control over the display of both virtual contents being rendered and the real world objects in the view. Due to this, almost all DR systems are built with Video See-Through displays [11]. We investigated the performance of occlusion of an OST AR headset, Magic Leap One, and proposed a system that leverages the limited concealing effect of the headset to provide people with an option of anonymity during interactions.

2. Related Work

Bimber and Raskar [4] discussed disadvantages of OST AR head-mounted displays with regard to rendering. They state that OST AR devices are generally incapable of providing consistent occlusion effects between virtual and real objects. As users see through the optical lenses into the real world, light of the illuminated real environment goes through the optical lenses and interferes with the virtual renderings displayed on the optical combiners (essentially half-silvered mirrors)[4].

Unlike the rendering of virtual contents in VST AR displays, where renderings can either be opaque or semi-transparent, the nature of OST displays means that all renderings are semi-transparent. Transparency is further increased outdoors or in the presence of a strong light source. Therefore, occluding real-world objects with virtual renderings in OST AR displays is challenging. Although LCD panels are being used by Kiyokawa et al. to selectively block the incoming light with respect to the rendered graphics[4][1], this approach is not widely adapted by modern OST AR headset manufacturers. This lays the ground for the challenges of the problem that we are seeking to tackle.

Overlaying avatar models on top of human faces is becoming a new trend in applications for OST AR headsets. Weta Gameshop announced a multiplayer mode at GDC 2019[2] [1] which overlays comic style 3D avatar models on top of other player's head. As a multiplayer mode which requires each participant to wear the headset, they could get the position and orientation of other player's head to facilitate the rendering. In the system we present, one user wears the headset and other users - subjects - do not. Our focus is on altering the perception of the viewer with the headset to preserve the anonymity of the subject. Therefore, we do not require whoever is in view wearing the device. On the other hand, with the subject not wearing anything that would occlude their face, this opens up opportunities to perform facial feature detection and analysis which could enhance the rendering with more information we obtain from the subjects' faces, for example their facial expressions.

3. Limitations

As discussed, concealing objects in reality by virtual renderings in OST headsets is generally challenging because of the nature of the display. Most OST headsets are designed for indoor purposes, favoring the visibility of the renderings in lower ambient light. Similarly, our design would work better in indoor settings without strong light coming through the headset. In addition, users with different pupil distances might experience differences in the depths of the rendering in 3D space, which is a challenge introduced by the device. This resulted in some of the users experiencing misalignment between the rendered model and the subject's head.

4. Method

An informal survey of members of the research team provided preliminary anecdotal evidence of how renderings in OST AR displays alter people's perception of color and edges. To understand how color renderings affect how people perceive colors in the real world, we overlay single-colored squares on top of single-colored squares displaying on computer screens and shuffle the colors of both renderings and display. Viewers reported colors they perceive on the display and the results confirmed that bright colors such as white, yellow and light green are best at changing people's perception of color, and thus have the best concealing effect.

In a second survey, we observed how different renderings influence people's perception of edge information. We combined bright colored background with pixelated faces and high frequency facial features separately and overlay the renderings on top of images of human faces. Based on the viewers' feedback, high frequency facial features easily stand out from renderings with only bright color background or with low frequency details. Therefore we decided on a combination of high frequency details with a bright color background, which would be a better approach in obscuring human faces.

Based on our observations, we designed the system of Anon-Emoji. Our system consists of an image marker that the subject wears on the head and a system running on the Magic Leap One headset. When the headset picks up the image marker, our system tracks the position and orientation of the person's head and renders an emotion-presenting 3D emoji around it so as to occlude his/her face.

We chose emojis for three main reasons: its nature in expressing non-verbal information; its usage in social media and casual portfolios for anonymity purposes; its strength in concealing effects under OST AR settings. The emotional information they bear is widely recognized and easy to comprehend. As they are in the rough shape of a head, they make sense to the viewers when overlayed on the head of a subject. People sometimes conceal their faces when posting photos on social media when they want to preserve their anonymity. Finally, emojis are generally bright yellow in color, which our experience tells us is ideal for occluding objects in OST AR displays.

In the near future, we will implement a separate process runs in the background on the device which aims to perform facial landmark detection and sentimental analysis to extract the facial expression of the subject, map it to a certain emoji, and change the renderings accordingly in real-time. We foresee that the changes of the renderings based on facial expressions could compensate for the loss of non-verbal communications that the occlusion wiped away.

5. System Design

System requirements: Magic Leap One Headset, OS version 0.95.2, MLSDK 0.20.0, Unity 2019.1.1, OpenCV for Unity 2.3.4

Our system consists of an AR headset and an image marker. Viewers are required to wear the AR headsets, while subjects only wear image markers on their foreheads. The image marker helps identify the position and rotation of the subject's head. Once the marker is picked up by the headset, a virtual rendering of 3D emoji model will be rendered around the subject's head. The emoji model will obscure most visual information of the subject's face from all directions.

Figure 2. The OST AR headset being used in the system: Magic Leap One headset with a lightpack and a 6DoF controller

6. Future Work

As mentioned in previous sections, we will integrate a facial expression extracting process to obtain facial expression information and present that information through the changes of emoji models. Once the system implementation is done, we aim to use the system to conduct several further studies: (a) How could rendered emojis influence how users perceive other people's emotions? (b) If the rendered emoji doesn't match the real facial expression of the subject, how would that affect the viewer's perception of others' emotions? (c) Can we ease the emotional tension during social interactions by delaying the rendering of negative facial expressions (angry, etc.)?

7. Discussions

Augmented Reality headsets has become increasingly available to consumer market. We foresee a near future where more people would have access to AR headsets. At a point of time where some people own these type of head-mounted smart devices that enable them with more information and real time computation power while others don't, inequality and privacy issues might occur. Similar to the discussion where people start to concern whether their buildings and properties would become the free billboards of AR advertisements, for those who don't own headsets, we would want them to have the option of being anonymous in the AR world as well.

References

[1] Magic leap weta workshop preview multiplayer mode for dr. grordbort's invaders at gdc. https://magic-leap.reality.news/news/magic-leap-weta-workshop-preview-multiplayer-mode-for-dr-grordborts-invaders-gdc-0194957/. Accessed: 2019-05-08. 2

[2] Magic leap unveils latest multiplayer vr prototype at gdc 2019. https://variety.com/2019/gaming/news/magic-leap-multiplayer-vr-grordbattle-1203166465/. Accessed: 2019-05-08. 2

[3] Ronald T. Azuma. A survey of augmented reality, 1997. 2

[4] Oliver Bimber and Ramesh Raskar. Modern approaches to augmented reality. In *ACM SIGGRAPH 2006 Courses on - SIGGRAPH '06*, 2006. 2

[5] Elizabeth A. Deitch, Adam Barsky, Rebecca M. Butz, Suzanne Chan, Arthur P. Brief, and Jill C. Bradley. Subtle yet significant: The existence and impact of everyday racial discrimination in the workplace. *Human Relations*, 2003. 1

[6] Daantje Derks, Agneta H. Fischer, and Arjan E.R. Bos. The role of emotion in computer-mediated communication: A review, 2008. 1

[7] Benjamin Edelman, Michael Luca, and Dan Svirsky. Racial discrimination in the sharing economy: Evidence from a field experiment. *American Economic Journal: Applied Economics*, 2017. 1

[8] Madeline E. Heilman and Lois R. Saruwatari. When beauty is beastly: The effects of appearance and sex on evaluations of job applicants for managerial and nonmanagerial jobs. *Organizational Behavior and Human Performance*, 23(3):360–372, 1979. 1

[9] Stefanie Johnson, Kenneth Podratz, Robert Dipboye, and Ellie Gibbons. Physical attractiveness biases in ratings of employment suitability: Tracking down the "beauty is Beastly" effect. *Journal of Social Psychology*, 2010. 1

[10] Steve Mann. Mediated Reality. *Fundamentals of Wearable Computers and Augmented Reality, Lawrence Erlbaum Associates, Inc*, 1994. 2

[11] Shohei Mori, Sei Ikeda, and Hideo Saito. A survey of diminished reality: Techniques for visually concealing, eliminating, and seeing through real objects. *IPSJ Transactions on Computer Vision and Applications*, 9(1), 6 2017. 2

[12] Anand Rajaraman and Jeffrey David Ullman. Generative Adversarial Nets. *Mining of Massive Datasets*, 9781107015:1–315, 2011. 1

[13] Sarah V. Stevenage and Yolanda McKay. Model applicants: The effect of facial appearance on recruitment decisions. *British Journal of Psychology*, 1999. 1

[14] D. S. Sundaram and Cynthia Webster. The role of nonverbal communication in service encounters, 2000. 1