TEllipsoid: Ellipsoidal Display for Videoconference System Transmitting Accurate Gaze Direction

Taro Ichii *
Tokyo Institute of Technology

Hironori Mitake†
Tokyo Institute of Technology

Shoichi Hasegawa‡
Tokyo Institute of Technology

ABSTRACT

We propose "Tellipsoid", an ellipsoidal display for the video conference system, that can realize not only accurate eye gaze transmission but also practicality in conferences, namely the convenience and the identity of the displayed face. The display consists of an ellipsoidal screen, small projector and convex mirror, where the bottom-installed projector projects the facial image of a remote participant onto the screen via the convex mirror.

Index Terms: Human-centered computing—Human computer interaction (HCI)—Interaction devices—Displays and imagers; Hardware—Communication hardware, interfaces and storage—Displays and imagers;

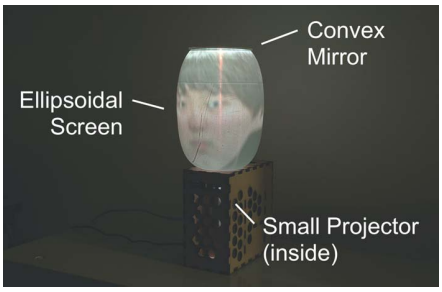

Figure 1: System Overview of TEllipsoid.

1 INTRODUCTION

In spite of the development of many kinds of information and communication technology, face-to-face conversation is the most standard way of communication. The technologies that make remote communication realistic as if face-to-face have huge advantages such as the reduction of labor/travel cost, which is needed to hold important decision-making face-to-face.

However, the planar display, which is commonly used on the current video conference system, cannot accurately transmit the non-verbal information (e.g. eye gaze or head movement), which plays an important role in communication, thus smooth communication is not possible. It is also reported that turn-taking on the multi-party conversation is adjusted by eye gaze or gesture [2, 7]. The difficulty of accurate gaze transmission is due to the Mona Lisa Effect [11], which makes it difficult to understand who is spoken to by the remote participant. Many studies have tried to eliminate this effect by using a volumetric display, such as a multi-view 3D display [6] or face-shaped display [9]. However, the devices require actuated mechanisms that cause mechanical noises and maintenance issues. As a more reasonable method, it is reported that a curved display can transmit eye gaze more accurately than a planar display [10].

Furthermore, we believe that practical videoconference system requires not only transmitting accurate gaze direction, but also the convenience of the system and identity between real/displayed face.

For practical use, high versatility is needed because unspecified many people have the possibility to participate in the conference as remote participants. The necessity of individual equipment, such as the tailor-made facial screen which is used on face-shaped display [9], brings inconvenience for the users. Moreover, lower manufacturing costs and higher reliability are also needed for practical application. It is obvious that the fewer the driving mechanisms of the system is, the smaller failure risk and manufacturing cost is. However, in the case of table meeting, the mechanisms are normally necessary because the face will turn by at most 90 degrees for left and right to look at the participants. The mechanisms are omittable on the vertical-axisymmetric surface of revolution shape display, which can express horizontal head rotation only with changes of image.

Identity between real/displayed face is also important for practical video conference system because we still cannot complete whole communication without face-to-face in today's society. In other words, the interlocutor of the video conference used to/will talk with the partner face-to-face at least once. In that case, if the displayed facial image is inconsistent with the real face of an interlocutor, it can be difficult for the user to recognize the image as the interlocutor [4]. It has been reported in human cognition studies, that we identify interlocutors by the overall arrangement of face parts and contour [1]. Hence the display that can show the arrangements of the facial parts naturally from any angle is necessary for the practical video conference systems.

Therefore we propose "TEllipsoid", an ellipsoidal display for the video conference system, which has curved, vertical-axisymmetric surface and the affinity with a human face, to realize not only accurate eye gaze transmission but also practicality in conferences, namely the convenience and the identity of the displayed face. This display is designed for the small-scale video conference in which the participants are 3 to 15 people.

2 PROPOSED METHOD

We proposed the ellipsoidal display for the video conference, which is named as "TEllipsoid" (Tele + Ellipsoid). Tellipsoid consists of an ellipsoidal screen, small projector and convex mirror (Fig. 1). The bottom-installed small projector projects the facial image onto the screen via the convex mirror. The artifacts on the face are due to the surface quality of the convex mirror. (We have a demo movie on YouTube.)

Screen shape was designed from the average face dimensions of Japanese young males [8]. We generated a 3D model of the ellipsoid of revolution, which was created by rotating simplified face shape along the vertical axis. Subsequently, we enlarged the lower part to improve the geometries of the nose, mouth and the naturalness of the profile of the facial image described later.

The facial image for the projection was firstly generated as the panorama composite photo from the photos shot 360 degrees around the person, which is equivalent to the cylindrical projection. The

*e-mail: ichitaro@haselab.net
†e-mail: mitake@haselab.net
‡e-mail: hase@haselab.net

978-1-7281-6532-5/20/$31.00 ©2020 IEEE
DOI 10.1109/VRW50115.2020.00272

Figure 2: The modified facial image for the projection.

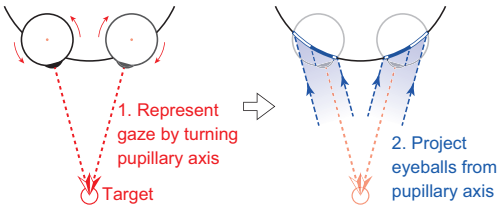

Figure 3: The method of gaze representation.

eye part was also cropped for the gaze representation, which is described later. Part of the neck was secondly filled with black, which is equivalent to no projection, to express the jaw contour (Fig. 2). This improvement makes the head contour more natural and has the potential to evoke Wollaston Illusion [12], in which the perceived gaze direction is influenced by face orientation. The shape of the black fill was carefully adjusted to accomplish both slender frontal face appearance and natural jaw contour appearance.

Thirdly, the facial image was inverse cylindrical-projected on the virtual screen surface as a texture. Trimming of the image was also done to adjust the geometry of facial parts. Fourthly, the transformed facial image for the projection was generated by the same optics setup assembled in Unity(2019.1.2f1). The projector in the device is replaced by a virtual camera in which the field of view is consistent with the throw ratio of the projector. We finally acquired a consistent stereoscopic image on the device by projecting the image captured by the virtual camera on the setup.

We also implemented the gaze representation because a participant gazes someone not only with the head turn but also with the eye rotation in the actual conference. The method of gaze representation is as shown in Fig. 3. The eyeball models are placed on the eye part of the virtual screen. The vertical/depth position and size are adjusted to be consistent with the real face. The lateral position and the pivot are aligned to the center of the orbit. They gaze the target position by turning the pupillary axis towards the target. Subsequently, the eyeballs are projected to the screen surface from the pupillary axis. This projection method is adopted because the geometry looks consistent with the eyeball models from the target point, which theoretically enables the perception of eye contact from the target position. This projection method is also similar to the previous study on the spherical display [3].

3 Exposition

We demonstrate the ellipsoidal display "TEllipsoid", which was also evaluated and submitted to the IEEE VR 2020 Conference Paper Track [5]. We provide the demo as the video conference system in which attendees can interact with the remote participant from any direction on the front side. The demo consists of TEllipsoid, a fish-eye camera, and a head-mounted display with face tracking, which enables the representation of eye gaze / the facial expression and immersive view for the remote participant.

4 Related Works

One-to-many 3D video teleconferencing system using an autostereoscopic horizontal-parallax 3D display [6] is a kind of solution to eliminate the Mona Lisa effect. The face of the remote participant is projected using a high-speed projector and rotating tent-shaped aluminum sheet metal.

A face-shaped display for the telepresence system [9], where the facial image is projected to the face-shaped screen made from the mold of an actual human face, is also a solution to eliminate the Mona Lisa effect. However, the versatility is not high because an individual screen is necessary for each user. Furthermore, some driving mechanism is needed to move or rotate the face.

On the evaluation of a cylindrical multiview videoconference system [10], it was reported that not only the gaze transmission accuracy of the proposed method was better than the planar display, but also a single facial image projected to the cylinder surface had more accuracy than planar display. However, the viewing angle that facial parts arrangements can be observed as consistent, which guarantees the identity, is limited.

Acknowledgments

This work was supported in part by a grant from JSPS KAKENHI 19K12284.

References

[1] R. Cabeza, V. Bruce, T. Kato, and M. Oda. The prototype effect in face recognition: Extension and limits. *Memory & Cognition*, 27(1):139–151, Jan 1999.

[2] S. Duncan. Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology*, 23(2):283–292, Aug. 1972.

[3] M. Hashimoto, H. Kondo, and Y. Tamatsu. Effect of emotional expression to gaze guidance using a face robot. In *Proc. RO-MAN*, pp. 95–100. IEEE, Munich, Germany, Aug 2008.

[4] M. Heike, K. Fujita, and T. Tanaka. Personalized and deformed avatars for realizing avatar-mediated real person-to-person communication. In *Proc. OCSC*, pp. 207–215. Springer-Verlag, Berlin, Heidelberg, 2009.

[5] T. Ichii, H. Mitake, and S. Hasegawa. Tellipsoid: Ellipsoidal display for videoconference system transmitting accurate gaze direction. In *Proc. IEEE VR*. (Under review as of Jan. 16, 2020).

[6] A. Jones, M. Lang, G. Fyffe, X. Yu, J. Busch, I. Mcdowall, M. Bolas, and P. Debevec. Achieving eye contact in a one-to-many 3d video teleconferencing system. *ACM Transactions on Graphics*, 28(3):64:1–8, Aug. 2009.

[7] A. Kendon. Some functions of gaze-direction in social interaction. *Acta Psychologica*, 26(1):22–63, Jan. 1967.

[8] M. Kouchi and M. Mochimaru. AIST anthropometric database. National Institute of Advanced Industrial Science and Technology, H16PRO 287.

[9] K. Misawa, Y. Ishiguro, and J. Rekimoto. Livemask: A telepresence surrogate system with a face-shaped screen for supporting nonverbal communication. In *Proc. AVI*, pp. 394–397. ACM, New York, NY, USA, 2012.

[10] Y. Pan and A. Steed. A gaze-preserving situated multiview telepresence system. In *Proc. CHI*, pp. 2173–2176. ACM, New York, NY, USA, 2014.

[11] S. Rogers, M. Lunsford, L. Strother, and M. Kubovy. *The Mona Lisa effect: Perception of gaze direction in real and pictured faces*, pp. 19–24. Lawrence Erlbaum, Jan. 2003.

[12] W. H. Wollaston. Xiii. on the apparent direction of eyes in a portrait. *Philosophical Transactions of the Royal Society of London*, 114(1):247–256, Jan. 1824.