

멀티 마이크론 기반 원격지 간 화상회의 시선 일치 기법

이대성 · 조동식*

울산대학교

Gaze Matching Based on Multi-microphone for Remote Tele-conference

Daeseong Lee · Dongsik Jo*

University of Ulsan

E-mail : dongsikjo@ulsan.ac.kr

요 약

최근, 대면 회의를 대체할 수 있는 수단으로 원격지 간 화상 회의 시스템의 활용도가 높아졌다. 기존 화상회의 시스템은 원격지 사용자의 시선 불일치로 인해 몰입감 있는 화상회의 환경을 제공하기에는 한계를 가지고 있다. 따라서, 서로 다른 원격지 간 참여자의 시선 정보를 일치하여 화상 회의의 몰입도를 높일 수 있는 기술 적용이 필요하다. 본 논문에서는 원격지 간 화상회의 참여자의 시선을 일치하기 위해 로컬에 설치된 멀티 마이크론 기반으로 화자의 위치를 추정하여 시선이 일치된 화상회의 영상을 구현하는 기술을 제안한다. 본 연구논문의 방법을 이용하면 원격지 간 화상회의 뿐 아니라 로봇 상호작용, 가상 휴먼 인터페이스 등 다양한 분야에 확대 적용될 수 있을 것으로 본다.

ABSTRACT

Recently, as an alternative to replace face-to-face meetings, video conferencing systems between remote locations has increased. However, video conferencing systems have limitations in terms of mismatch of the eyes of remote users. Therefore, it is necessary to apply a technology that can increase the level of immersion in video conferences by matching the gaze information of participants between different remote locations. In this paper, we propose a novel technique to realize video conferencing with the same gaze by estimating the speaker's location based on a multi-microphone. Using our method, it can be applied to various fields such as robot interaction and virtual human interface as well as video conferencing between remote locations.

키워드

Gaze, Tele-conference, Remote, Inconsistency

1. 서 론

화상 회의 시스템은 원격지 간 떨어져 있는 사용자들을 위한 커뮤니케이션 수단으로 널리 적용되고 있다. 메라비언의 법칙에 따르면 커뮤니케이션에 있어 상대방으로부터 받는 이미지는 시각이 55%, 청각이 38%, 언어가 7%에 이른다고 한다 [1]. 이는 언어적 소통의 비중은 7%에 불과하고 나머지 93%는 비언어적 소통으로 채워짐을 의미한다. 즉, 화상회의에서 시선 맞춤, 제스처 등의 시각적 요소가 중요함을 알 수 있다. 하지만, 영상 기반의 화상회의 시스템의 경우 참여자 상반신을 기반으로 표현을 하고 있어 시각적 회의 몰입도

요소가 부족한 상황이다. 예를 들면, 기존의 화상회의 시스템에서 사용자는 카메라가 아닌 상대방의 얼굴이 나오는 모니터를 주시하게 되므로 시선이 주로 아래로 향하고, 이로 인해 커뮤니케이션에 중요한 요소 중 하나인 참여자 간 시선 맞춤이 줄어들게 되고 몰입감 있는 화상 회의 환경을 제공하기에는 한계가 있다[2-4].

본 논문에서는 기존의 화상 회의의 시선 불일치 문제점을 해결하기 위해 로컬에 설치된 멀티 마이크론 기반으로 적용화된 화상 회의 영상을 구현하는 기술을 제안한다. 이를 위해 사용자의 2D 입력 영상을 3D warping을 통하여 3D 모델로 재구성하고, 멀티 마이크론으로 화자의 위치를 추정하였다. 최종적으로 획득한 화자의 위치를 기준으로 3D 모델을 조정하여 시선을 일치하였다. 그림

* corresponding author

1은 멀티 마이크론을 이용한 사용자 시선 일치 하드웨어 구성도이다. 기존 화상회의 시스템처럼 카메라 및 디스플레이 구성은 동일하고, 추가적으로 멀티마이크론을 구성하여 화자의 위치를 추정할 수 있도록 하였다.

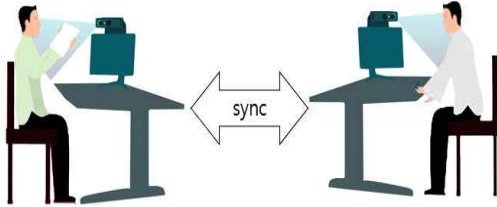


그림 1. 멀티 마이크론을 이용한 사용자 시선 일치 하드웨어 구성도

II. 멀티 마이크론 기반 시선 일치

본 논문에서 제안하는 화상 회의 시선 일치 기법의 구현을 위해서 2D 이미지를 3D 모델로 만드는 이미지 프로세싱, 멀티 마이크론을 통한 화자의 추정 위치, 3D warping을 통한 시선 보정 3단계 과정으로 수행하였다.

첫 번째 단계인 사용자 시선 일치를 위한 이미지 프로세싱 과정은 3D reconstruction과 3D warping 과정으로 구성하였다. 3D reconstruction은 여러 개의 2D image로부터 3D 모델로 얻어내어 실제 물체의 모습이나 모양을 복원하는 과정이고, 참여자의 3D 모델을 추출하였다. 두 번째 단계인 화자의 위치를 추정하기 위해서 본 논문에서는 멀티 마이크론이 내장된 카메라를 이용하여 소리의 발생지를 추정하고, 멀티 마이크론을 통하여 전방위 화자의 위치를 추정하도록 하였다. 마지막으로 추정된 화자의 위치를 기반으로 3D warping 이미지 보정을 통해 시선을 일치하도록 하였다. 요약하자면, 입력 영상인 사용자 2D 이미지를 이미지 프로세싱 과정을 통하여 3D 모델로 얻어내고, 추정된 화자의 위치를 기준으로 시선 일치된 2D 이미지를 참여자에게 제공하도록 하였다.

III. 결 과

그림 2는 본 논문에서 제안한 화상 회의 시스템의 시선 일치 결과이다. 시선일치를 수행하지 않는다면 발표자의 시선이 모니터를 향하므로 시선이 아래를 향할 수 있어 참여자들의 화상 회의 몰입감 저하로 이어질 수 있을 것이다. 이러한 시선 불일치의 문제를 해결하기 위해 본 논문의 방법을 적용한 3D warping을 통해 시선을 처리하였다[5]. 결과적으로 시선 일치(Gaze Correction)를 적용하

여 참여자의 시선이 교정이 되면 참여자의 몰입감이 높아질 것이다.

기존의 시선일치 기술은 사용자가 정면을 바라보는 경우에만 시선을 교정할 수 있고, 사용자가 카메라 측면에 위치하거나, 측면을 바라보는 경우에는 시선 맞춤을 할 수 없다는 한계가 있지만, 본 논문 제안 방법을 원격지 간 화상 회의 시선 일치 기법을 이용한다면 사용자의 위치, 각도와 관계없이 시선 일치가 가능할 것이다.



(a) (b)

그림 2. 시선 일치 결과: (a) 알고리즘 적용 전, (b) 알고리즘 적용 후

IV. 결론 및 추후 연구

본 논문에서는 원격지 화상 회의 참여자의 시선 불일치 문제를 해결하기 위해 멀티 마이크론을 이용하여 화자의 위치를 추정하고, 이를 기준으로 화자의 영상 입력에 대해 시선을 일치시키는 방법을 제시하였다. 본 연구 결과에 따르면 참여자의 위치와 각도와 관계없이 시선을 일치하는 것이 가능하다. 추후 다수의 화상회의 참여자에 대해 동시에 시선을 일치할 수 있는 방법에 대한 연구가 필요하고, 또한, 3D 복원 데이터를 구성하는 정밀도를 높일 수 있는 추가 연구가 필요하다.

Acknowledgement

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2021R111A3060198)

References

[1] Albert Mehrabian, "Some referents and measures of nonverbal behavior," *Behavior Research Methods & Instrumentation*, 1968,
 [2] S. Firestone, T. Ramalingam, and S. Fry, "Voice and video conference fundamentals," *Cisco Press*,

- 2007.
- [3] K. Otsuka, "MMSpace: Kinetically-augmented telepresence for small group-to-group conversations," *Proceedings of the IEEE VR*, pp. 19-28, 2016.
 - [4] A. Maimone, X. Yang, N. Dierk, A. State, M. Dou, and H. Fuchs, "General-purpose telepresence with head-worn optical see-through displays and projector-based lighting," *Proceedings of the IEEE VR*, pp. 23-26, 2013.
 - [5] Hsu, Chih-Fan and Wang, Yu-Shuen and Lei, Chin-Laung and Chen, Kuan-Ta, "Look at Me! Correcting Eye Gaze in Live Video Communication," *ACM Trans. Multimedia Comput. Commun. Appl.* Vol. 1, No. 1, Article 1. January 2016.