Annotating on Live Video for Collaboration

Dongxing Cao*, Heeju Choi*, Seungdeok Seo*, Junghun Hwang**, Sangwook Kim*
* School of Computer Science and Engineering, Kyungpook National University
** School of Convergence & Fusion System Engineering, Kyungpook National University

협업을 위한 비디오의 주석 방법 연구

차오동싱*, 최희주*, 서승덕*, 황정훈**, 김상욱*
*경북대학교 컴퓨터학부
**경북대학교 융복합시스템공학부
e-mail : dxcao@media.knu.ac.kr, ujst02@gmail.com, sudang17,anthony9307@naver.com, kimsw@knu.ac.kr

Abstract

Graph is much easier than word to be understood for people. When explaining complex problems, people prefer using a graph to word to demonstrate and analyze it. But Prior work base on a static picture with 2D annotation, it provides a simple way to collaborative with remote users. Therefore, this paper explores using Augmented reality (AR) in a remote collaboration system. Augmented reality (AR) allows the annotations to be overlaid directly on an environment. This can improve readability, thus increasing work efficiency.

1. Introduction

Previous research has shown that real-time annotation can help to improve the work efficiency [1]. But when collaborate on the live video, a 2D annotations can easily lose their meaning when shown from novel viewpoints due to perspective effects. In this paper, we introduce AR technology into my existing collaboration system, make it possible to render the 2D annotations in 3D in a way that conforms more to the original intention of the user than with traditional methods.

The fast-technical advancement of handheld devices has increased the interest of HAR (Handheld Augmented Reality) among researchers and developers [2,3]. This paper research and implements the mapping of 2D drawings in 3D space using HAR, and explore ways to improve the accuracy of 3D annotations in space.

2. Related Works

There have been a number of earlier studies that explored different annotation methods using AR. However, most of these systems require special equipment like (Head-mount-displays) and cannot adapt to everyday environment. Sebastian et al. [4] designed a multi-user AR system based on HAR, it can place 2D annotations onto 3D planes and share to other users, but the main purpose of it is for entertainment, cannot collaborative with geographically dispersed users.

3. Render the 2D annotations in 3D

The system uses image-based AR recognition technology (ARcore), which recognizes a natural image as a target object with advanced registered dictionary data. We can create a set of dictionary data from the picture of the surface from the everyday object. When drawing annotations on the live video, at the same time, the video frame will be saved as a recognized object. These dictionary data contain the image feature points for recognition. This feature is like a fingerprint that indicates every surface as a unique object (Fig 1).

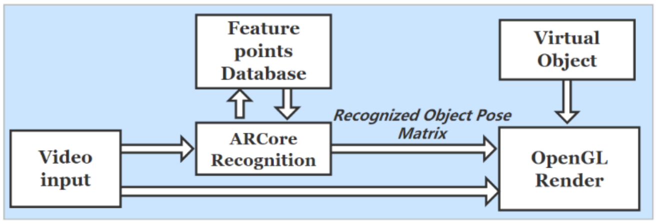

Figure 1. AR recognition process

A real-time video stream from the built-in camera is used for AR recognition and the system estimates the pose M_{recog} of the recognized object in relation to the camera focusing on it (Fig 2). M_{recog} represents a 4 x 4 matrix consisting of the pose includes a translation and a rotation. M_{recog} is the reference which affects other matrices.

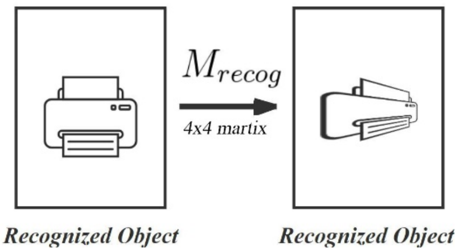

Figure 2. Generation process of matrix M

By using M_{recog} we can calculate out the pose of device and virtual canvas easily(Fig 3).

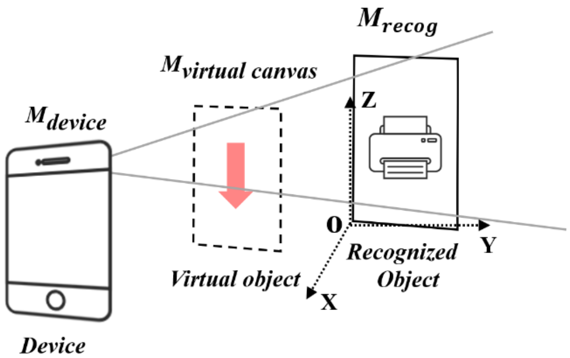

Figure 3. Matrix definition

$$\mathrm{M}_{device} = \mathrm{M}_{recog} \cdot \mathrm{T}_{device\ center} \quad (1)$$

$$\mathrm{M}_{virtual\ canvas} = \mathrm{M}_{recog} \cdot \mathrm{T}_{offset} \quad (2)$$

Here, T_{offset} represents a translation matrix from the camera to virtual canvas, which is defined based on a device' s camera parameter $\mathrm{T}_{device\ center}$ represents a translation matrix from the camera to center of the device, which is defined based on the physical placement of device' s camera. Because of the change of M_{recog} , the pose of virtual object Will follow the change (Fig 4,5). Afterward, these virtual objects are rendered on top of the video stream.


Figure 4. Pose of annotation from two angles(a)


Figure 5. Pose of annotation from two angles(b)

6. Conclusions and Future Work

In this paper, we researched the method of rendering 2D drawing annotations in 3D. Compared to 2D annotations, the semantics of 3D annotations are more explicit and can greatly improve the efficiency of collaboration systems. There are two main focuses on future work :(1) automatically inferring depth for 2D drawings in 3D space, and (2) multiple users can drawing annotations on shared live video accurately.

Acknowledgements

This research was supported by the Korean MSIT (Ministry of Scienceand ICT), under the National Program for Excellence in SW)(2015-0-00912) supervised by the IITP(Institute of Information & communications Technology Planning & Evaluation)"(2015-0-00912)

References

[1] D. Kirk, T. Rodden, D. S. Fraser, “Turn it this way: grounding collaborative action with remote gestures,” SIGCHI conference on Human Factors in Computing Systems, pp. 1039-1048, 2007.

[2] E. Kruijff, J. E. Swan, S. Feiner, “Perceptual issues in augmented reality revisited,” IEEE International Symposium on Mixed and Augmented Reality, pp. 3-12, 2010.

[3] J. Polvi, T. Taketomi, G. Yamamoto, A. Dey, C. Sandor, “SlidAR: A 3D positioning method for SLAM-based handheld augmented reality,” Computers & Graphics, pp. 33-43, 2016.

[4] S. Boring, S. Gehring, A. Wiethoff, A. M. Blöckner, “Multi-user interaction on media facades through live video on mobile devices,” SIGCHI Conference on Human Factors in Computing Systems, pp. 2721-2724, 2011