Journal of Korea Multimedia Society Vol. 23, No. 3, March 2020(pp. 483-489)
https://doi.org/10.9717/kmms.2020.23.3.483

Augmented Reality Annotation for Real-Time Collaboration System

Dongxing Cao[†], Sangwook Kim[††]

ABSTRACT

Advancements in mobile phone hardware and network connectivity made communication becoming more and more convenient. Compared to pictures or texts, people prefer to share videos to convey information. For intentions clearer, the way to annotating comments directly on the video are quite important issues. Recently there have been many attempts to make annotations on video. These previous works have many limitations that do not support user-defined handwritten annotations or annotating on local video. In this sense, we propose an augmented reality based real-time video annotation system which allowed users to make any annotations directly on the video freely. The contribution of this work is the development of a real-time video annotation system based on recent augmented reality platforms that not only enables annotating drawing geometry shape on video in real-time but also drastically reduces the production costs. For practical use, we proposed a real-time collaboration system based on the proposed annotation method. Experimental results show that the proposed annotation method meets the requirements of real-time, accuracy and robustness of the collaboration system.

Key words: Augmented reality, Real-time, Video Annotation, Collaboration

1. INTRODUCTION

Advancements in mobile phone hardware and increased network connectivity made video's popularity continues to grow. Some aspects are urged to be improved. For example, live video streaming apps have been used for sharing experiences with other viewers. On the contrary, the viewer can also provide feedback such as make comments. The feedback comments usually appear in a list below or beside the video being shared, separate from the visual context of what the viewer is commenting on. This may cause problems when the person sending the video changes his or her viewpoint. Compared to static media, people now prefer to take a video with their phone to convey some informations, but they cannot do such highlight important parts as same as static medias.

The best solution is to provide a way to comment directly on the specific frame of the video. We call this process as video annotation. Annotation is an important learning strategy. After reading some document or textbook, the learner write their ideas and record important content summary by drawing line or circle in their reading content for invoking their memory or offering hints for reading in the future. In explainer video, the video producer use a variety of annotations and virtual

※ Corresponding Author : Sangwook Kim, Address: (702-701) IT No.4-401, Kyungpook National University, Daehakno 80, Bukgu, Daegu, Korea, TEL : +82-53-940-8881, E-mail : kimsw@knu.ac.kr

Receipt date : Aug. 29, 2019, Revision date : Jan. 13, 2020
Approval date : Feb. 3, 2020

[†] Dept. of Computer Science Engineering., Graduate School, Kyungpook National University
(E-mail : dxcao@media.knu.ac.kr)

[††] Dept. of Computer Science Engineering., Graduate School, Kyungpook National University

※ This research was supported by the BK21 Plus project of the Ministry of Education and the Korea Research Foundation (SW Human Resources Development Team for the realization of Smart Life at Kyungpook National University) (21A20131600005).

contents to help people understand better. At present, adding annotations to videos requires professional video editing software. The editor has to set the position of the virtual contents in each frame directly in order to fix 3D contents and produce animations. This takes enormous cost and time, therefore it's not easy for the public to use.

There have many real-time annotation system base on static medias [1,2,3], but when the static image is change to sequence of images, the original intention of the annotations will be misinterpreted from novel perspectives. Therefore it is greatly significant to research on the lightweight real-time video annotation system which can adapt to everyday environment. There are two difficulties in real-time annotation system. One have to make sure annotations are fixed in the corresponding scene of the video. The other is the annotation processing should meets the requirements of real-time.

In this work, we investigate how annotations can be displayed directly on the streamed video in real-time using a mobile device, and especially focus on using Augmented Reality (AR) technic. ARCore is a recent AR platform for mobile device. It combines the advanced feature matching algorithm ORB with the smartphone's IMU to achieve the following functions on mobile phones: motion tracking, environment understanding, and light estimation [4]. We use ARCore to tracking the motion of object in video and associate this movement with annotations and adjust the shape of the annotations in real time with OpenGL.

This paper consists of the following parts. Some earlier studies are been discussed in Section 2. Section 3 presents our real-time video annotation method. A real-time collaboration system base on the propose method will be shown in Sections 4. Finally, we summarize and evaluate our work in Section 5.

2. RELATED WORKS

There have been several earlier studies that explored the real-time video annotation system for different purposes. Lai et al. [5] developed a web-based video annotation system for online classroom. Students can enter answers in the pop-up dialog box, and the teacher can browse the student's answers in the corresponding video position. The system is completely developed based on html5 features. So the annotations just simply floating on top of the video. Cho et al. [6] proposed a real-time interactive AR system for broadcasting. The system support the real-time interaction between the augmented virtual contents and the casts, the system perceives the indoor space using a RGB-D camera and the area of each object is separated through clustering then replaced with virtual content. Venerella et al. [7] proposed a lightweight annotation system for collaboration. The system is using CS architecture, client side generate 3D mesh using 6D.AI, According to the textured mesh, server side can annotate virtual objects and add virtual landmarks in the client's environment, These previous works have many limitations that do not support free-hand drawings or lightweight.

With the rapid increase in mobile device performance and network technology in recent years. Recently there have been many attempts to make video annotation system totally base on mobile phone. Choi et al. [8] proposed a collaboration system, which allows multi-user to create and situate contents on live video streaming. This system uses an image-based AR platform called Vuforia to implement such features belows. Such systems only support annotation on live video streams [9]. Nassan et al. [10] proposed a system that can comments displayed over the background video. But the system only support text type annotations.

3. REAL-TIME VIDEO ANNOTATION

3.1 System design

The proposed system using ARCore SDK and OpenGL library to rendering annotations directly

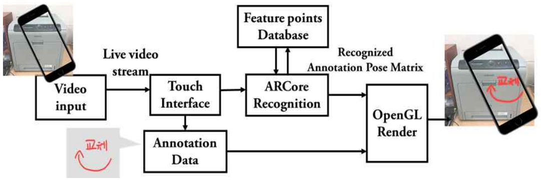


Fig. 1. AR-based real-time video annotation processing.

on the video in real-time. ARCore can recognize the natural image as a target object with advanced registered dictionary data. We can create a set of dictionary data from the frame of the video. When drawing annotations on the video, at the same time, the feature points in the video frame will be saved into a dictionary. By tracking the feature points, a motion matrix can be calculated. According to the motion matrix, OpenGL is able to render specified 3D models displayed on the screen (Fig. 1). The feature matching process will be introduced in Section 3.1, and the calculation process of the pose matrix will be presented in Section 3.2.

3.2 Feature matching

Feature descriptors are a part of computer vision and image processing eds. In this paper, a corresponding binary keypoint descriptor algorithm ORB [11] is used for feature detection. First, find the feature points of two photos using FAST algorithm, then describe the attributes of these feature points using BRIEF algorithm, finally compare the attributes of the feature points of the two pictures. If there are enough feature points with the same attributes, then the match is successful. The feature matching result is shown in Fig. 2.

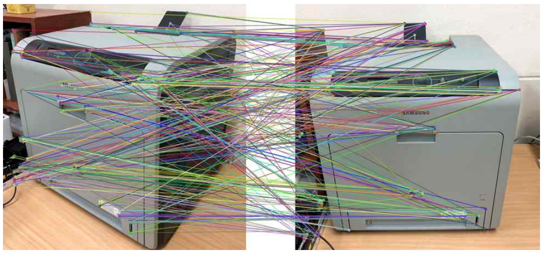


Fig. 2. Feature matching result using ORB algorithm.

Research has shown that the ARCore 's performance and ability to detect feature points could prove to be dependent on lighting conditions, the texture of a surface and the angle the device is in [12].

3.3 Estimates the pose of virtual object in real-time

A real-time video stream generated by device camera is processed in ARCore recognition module for feature point detection. Together with the device's orientation and its accelerometer sensors, ARCore can real-time estimates the pose of the M_{recog} recognized sparse point cloud in a physical objects. M_{recog} represents the motion process of recognizing the spatial position of feature points during the change of the viewpoint. M_{recog} is a 4 × 4 matrix consisting of a translation and a rotation. The derivation process of the M_{recog} matrix as follows:

In Fig. 3, Since C_0, C_1, and p are coplanar, C_0, C_1, p_0, and p_1 are coplanar. The coplanarity of vectors can be established by the following equation:

$$\overrightarrow{C_0p_0} \bullet \left(\overrightarrow{C_0C_1} \times \overrightarrow{C_1p_1}\right) = 0 \tag{1}$$

$$\because p_0 = \begin{bmatrix} x_0 \\ y_0 \\ 1 \end{bmatrix}_{C_0}, p_1 = \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix}_{C_1} \quad \therefore \overrightarrow{p_0} \bullet \left(\vec{t} \times \overrightarrow{Rp_1}\right) = 0 \tag{2}$$

$$\because \vec{a} \times \vec{b} = [a]_X \vec{b} \quad \therefore p_0^T [t]_X R p_1 = 0 \tag{3}$$

$$E = [t]_X R \tag{4}$$

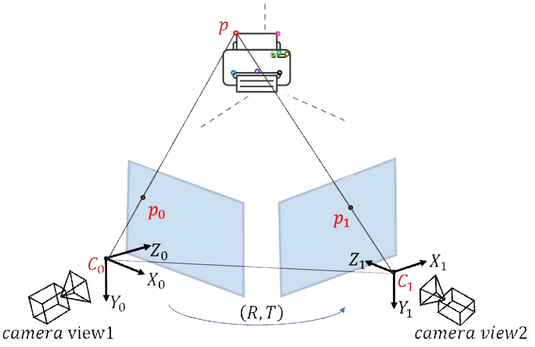


Fig. 3. Polar geometry model for matrix M_{recog} estimation.

Here E is a 3×3 matrix :

$$\therefore \begin{bmatrix} x_0 \; y_0 \; 1 \end{bmatrix} \begin{bmatrix} E_{11} & E_{12} & E_{14} \\ E_{21} & E_{22} & E_{23} \\ E_{31} & E_{32} & E_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} = 0 \tag{5}$$

$$\begin{bmatrix} x_0 x_1 \; x_0 y_1 \; x_0 \; y_0 x_1 \; y_0 y_1 \; y_0 \; x_1 \; y_1 \; 1 \end{bmatrix} \begin{bmatrix} E_{11} \\ E_{12} \\ \vdots \\ E_{33} \end{bmatrix} = 0 \tag{6}$$

$$\begin{bmatrix} x_0 x_1 \; x_0 y_1 \; x_0 \; y_0 x_1 \; y_0 y_1 \; y_0 \; x_1 \; y_1 \; 1 \end{bmatrix} E_{33} \begin{bmatrix} \frac{E_{11}}{E_{33}} \\ \frac{E_{12}}{E_{33}} \\ \vdots \\ 1 \end{bmatrix} = 0 \tag{7}$$

Since there are 8 unknown variables $\left\{ \frac{E_{11}}{E_{33}}, \frac{E_{12}}{E_{33}}, \ldots, \frac{E_{32}}{E_{33}} \right\}$, so we need to require at least 8 feature points to solve Eq. 7. Fig. 2 shows the feature matching result using the ORB algorithm. The result shows there have more than 50 matching pairs. And in most of the scenarios we have tested there have more than 8 pairs of feature points, so AR drawings can be kept in place steadily.

By using M_{recog} we can calculate out the pose of device M_{device} and virtual canvas $M_{virtualcanvas}$ easily (Fig. 4).

$$M_{device} = M_{recog} \cdot T_{device\,center} \tag{8}$$

$$M_{virtualcanvas} = M_{recog} \cdot T_{offset} \tag{9}$$

Here, T_{offset} is a translation matrix from the camera coordinate system to the virtual object coordinate system, which is mainly defined by a device's camera parameter. $T_{device\,center}$ is a translation matrix from the camera coordinate system to device coordinate system, which is mainly defined by the physical position of the build-in camera. By dynamically updated parameter $M_{virtualcanvas}$, Open GL is able to render specified 3D models during movement.

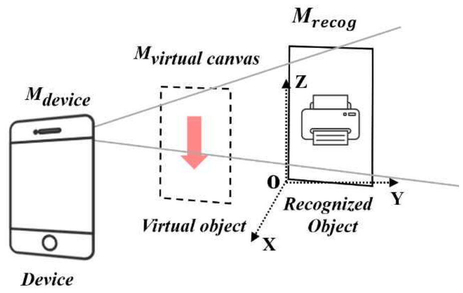


Fig. 4. Matrix definition.

4. REAL-TIME COLLABORATIVE AR SYSTEM

4.1 System overview

For the practical use, we proposed a novel real-time collaboration system based on the proposed annotation method. We think the integration of remote collaboration and a co-located collaborative way is one of the novelty points of the proposed system. Remote collaboration allows multiple users simultaneously to view and annotate three-dimensional virtual information among shared real-time video streams. Co-located collaborative service allowed a user to persist virtual objects in the same place as before (aka Relocalization) so that they can share AR experiences with other users in the same environment. The System overview is shown in Fig. 5.

4.2 System architecture

Recently ARCore team added native support for Firebase's real-time Database which will bring a

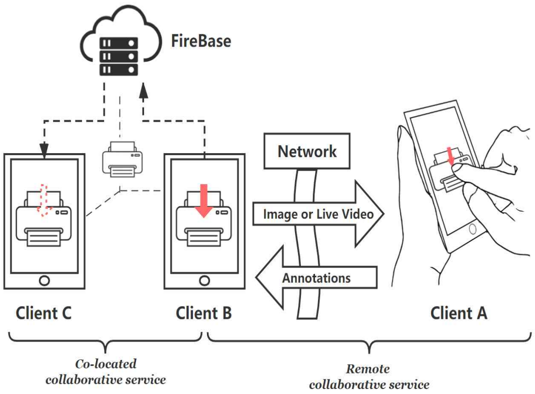


Fig. 5. System overview.

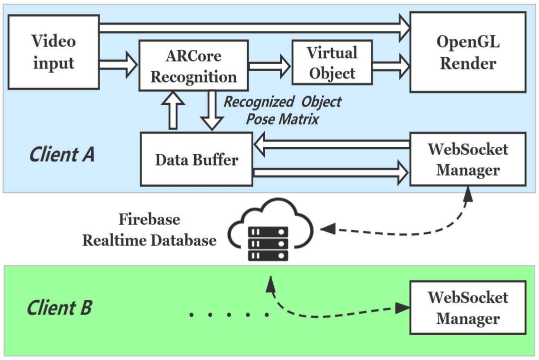
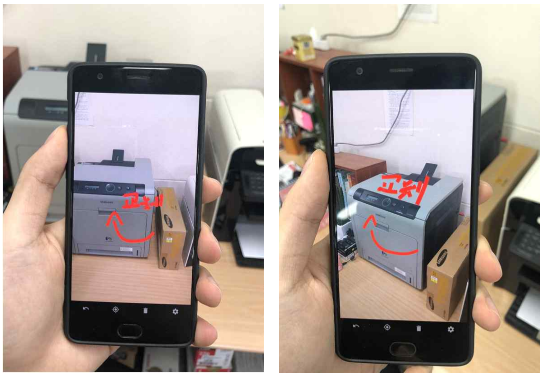


Fig. 6. System architecture.

stable shared AR experience [13]. The system architecture is shown in Fig. 6. Live video stream code by H.264 then shared by RTMP protocol on the internet. Guidance data created by the remote helper side are transmitted via the TCP protocol.

5. IMPLEMENTATION AND EVALUATION

The experiments in this paper were conducted on OnePlus 3T, which is one of ARCore's officially supported devices. First, we drew a 2D arrow pointing to the printer cartridges, we can see the 2D arrow is immediately been augmented to a 3D arrow in the environment (Fig. 7), which express the intention of replacing the printer cartridge clearly.



(a) (b)

Fig. 7. Implementation results of the proposed real-time video annotation method from two viewpoints. (a) Front view. (b) Side view.

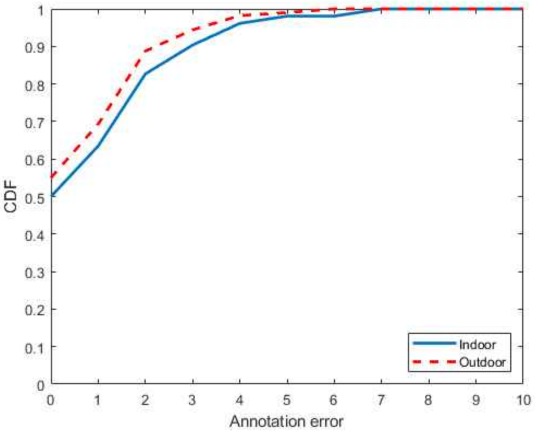


Fig. 8. Video annotation performance in various environments.

In order to qualitatively evaluate our work, especially the success rate. We performed the same annotation task 100 times under indoors and outdoors environment, each task will produce ten virtual annotations. Record the number of failures in each task. As evaluation criteria, if the augmented annotation has a large displacement or suddenly disappears, then it will be thinking as annotation error, vice versa. Fig. 8 plots the CDF of annotation error comparisons for two environments. This result demonstrates the proposed system can complete the annotation task after about 4 failures in different environments.

In this paper, we proposed a novel AR-based collaboration system. For implementation, we researched a 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. During the experimental evaluation, the average success rate of annotation is about 90.21%, which shows the stability of system. There are two main focuses on future work :(1) Improve annotation method to achieve automatically inferring depth for 2D drawings in 3D space, and (2) Using Integrated IoT system to searching users for real-time collaboration [14].

REFERENCE

[1] H. Attiya, S. Burckhardt, A. Gotsman, A. Morrison, H. Yang, M. Zawirski, et al., "Specification and Complexity of Collaborative Text Editing," *Proceeding of ACM Symposium on Principles of Distributed Computing*, pp. 259-268, 2016.

[2] L. Gao, D. Gao, N. Xiong, and C. Lee, "CoWeb Draw: A Real-time Collaborative Graphical Editing System Supporting Multi-clients Based on HTML5," *Journal of Multimedia Tools and Applications*, Vol. 77, No. 4, pp. 5067-5082, 2018.

[3] X. Wang, J. Bu, and C. Chen, "Achieving Undo in Bitmap-based Collaborative Graphics Editing Systems," *Proceedings of the Conference on Computer Supported Cooperative Work*, pp. 68-76, 2015.

[4] Shared AR Experiences with Cloud Anchors, https://developers.google.com/ar/develop/java/cloud-anchors/overview-android (accessed August 20, 2019).

[5] A.F. Lai, W.H. Li, and H.Y. Lai, "A Study of Developing a Web-based Video Annotation System and Evaluating Its Suitability on Learning," *Proceeding of Second International Conference on Education and Multimedia Technology*, pp. 44-48, 2018.

[6] H. Cho, S.U. Jung, and H.K. Jee, "Real-time Interactive AR System for Broadcasting," *Proceeding of IEEE Virtual Reality*, pp. 353-354, 2017.

[7] J. Venerella, L. Sherpa, H. Tang, and Z. Zhui, "A Lightweight Mobile Remote Collaboration Using Mixed Reality," *Proceedings of Computer Vision and Pattern Recognition 2019*, pp. 1-4, 2019.

[8] S.H. Choi, M. Kim, and J.Y. Lee, "Situation-dependent Remote AR Collaborations: Image-based Collaboration Using a 3D Perspective Map and Live Video-based Collaboration with a Synchronized VR Mode," *Journal of Computers in Industry*, Vol. 101, pp. 51-66, 2018. https://doi.org/10.1016/j.compind.2018.06.006 (Issue Numbers Are Missing)

[9] S. Lukosch, M. Billinghurst, and L. Alem, "Collaboration in Augmented Reality," *Journal of Computer Supported Cooperative Work*, Vol. 24, No. 6, pp. 515-525, 2015.

[10] A. Nassani, H. Kim, G. Lee, M. Billinghurst, T. Langlotz, R.W. Lindeman, et al., "Augmented Reality Annotation for Social Video Sharing," *Proceeding of Special Interest Group on GRAPHics ASIA 2016 Mobile Graphics and Interactive Applications*, pp. 1-5, 2016.

[11] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An Efficient Alternative to SIFT or SURF," *Proceedings of IEEE International Conference on Computer Vision*, pp. 2564-2571, 2011.

[12] A. Eklind, *An Exploratory Research of ARCore's Feature Detection*, Master's Thesis of KTH Royal Institute of Technology, 2018.

[13] Working with Anchors, https://developers.google.com/ar/develop/developerguides/anchors (accessed December 26, 2019).

[14] S. Ryu and S. Kim, "Development of an Integrated IoT System for Searching Dependable Device Based on User Property," *Journal of Korea Multimedia Society*, Vol. 20, No. 5, pp. 791-799, 2017.

Dongxing Cao

received the B.S. degree in Digital Media Technology from Century College, Beijing University of Posts and Telecommunications, China in 2018. He joined the Department of Computer Engineering for pursuing his M.S. degree at Kyungpook National University, South Korea, in 2018. He is interested in Mobile Media Computing and Internet of Things.

Sangwook Kim

received Bachelor's Degree in Computer Engineering from Kyungpook National University in February, 1979. He received his master's degree in Computer Science from Seoul National University in February, 1981. He obtained his Ph.D. in Computer Science from Seoul National University in February, 1989. He has been a professor of Computer Science at Kyungpook National University since 1982. His interests include mobile media, social media, interaction between humans and computers, digital art and IoT (Internet of things).