Augmented Reality Annotation for Social Video Sharing

Alaeddin Nassani*[1], Hyungon Kim[1], Gun Lee[2], Mark Billinghurst[2], Tobias Langlotz[3], and Robert W. Lindeman[1]

[1]HIT Lab NZ, University of Canterbury
[2]University of South Australia
[3]University of Otago

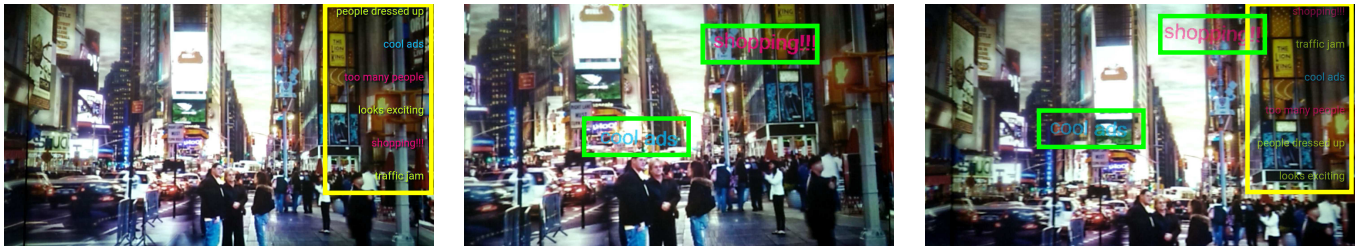

Figure 1: *Overview on the investigated interfaces showing screenshots of the different interface conditions. (L) List: Comments displayed as a list on the side. (AR) AR: Comments displayed over the background video. (L+AR) List+AR: Comments displayed both as a list on the right and over the background video.*

Abstract

This paper explores different visual interfaces for sharing comments on a social live video streaming platforms. So far, comments are displayed separately from the video making it hard to relate the comments to event in the video. In this work we investigate an Augmented Reality (AR) interface displaying comments directly on the streamed live video. Our described prototype allows remote spectators to perceive the streamed live video with different interfaces for displaying the comments. We conducted a user study to compare different ways of visualising comments and found that users prefer having comments in the AR view rather than on a separate list. We discuss the implications of this research and directions for future work.

Keywords: Augmented Reality; Live video streaming, Annotations

Concepts: •**Human-centered computing** → **Mixed / augmented reality;**

1 Introduction

Advancements in mobile phone hardware and increased network connectivity made live video streaming apps popular among smartphone users. Live video streaming apps have been used for sharing social experiences in various contexts. For instance, a person attending a conference or a concert could use her mobile phone to stream the event to her friends and family who could not be there. Similarly, live video streaming apps have also been used for Social journalism turning laypersons into live reporters. Consequently, these apps are now available from different sources with applications such as Periscope[1] and Facebook Live[2] among the most popular apps.

They all share common features such as using the phones' camera which can be either pointed outward (recording what the user sees) or inward (where the user appears in the video) and allowing users to send a live video stream of what they are doing to hundreds or even thousands of viewers. The purpose of sharing the video is social, so the experience is improved if the viewer can also provide feedback. Applications like Periscope allow the users who are sharing to receive comments on the video they are sharing as well as they can receive simple graphical feedback.

In these applications, 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. For example, a viewer might send the comment I really like that picture, but by the time the comment appears, the view might already have changed from the picture being commented on.

In this work, we investigate how comments can be displayed for a live video sharing experience using a mobile device, and especially focus on using Augmented Reality (AR). We implemented three different interfaces to display comments: (1) List, (2) Augmented Reality (AR), and (3) List + AR (see Figure 1). In the rest of the paper, we first describe earlier research, then our prototype implementations, and finally a user evaluation comparing the three different methods.

*email: alaeddin.nassani@pg.canterbury.ac.nz

SA '16 Symp on Mobile Graphics and Interactive Applications ,, December 05-08, 2016, , Macao
ISBN: 978-1-4503-4551-4/16/12
DOI: http://dx.doi.org/10.1145/2999508.2999529

[1]https://www.periscope.tv/
[2]https://live.fb.com/

2 Related Work

Our work extends earlier work on live video sharing on mobile devices and different types of interfaces for showing feedback from the viewers.

Current popular live video sharing apps, such as Periscope, or popular live streaming websites such as Douyu[3] and Ingkee[4] use a single way of displaying comments from other users. The common method is to display comments as a list either beside or below the shared video or sometimes floating on top of the video from left to right. This approach is a good extension of standard chat applications. However, it may not be the best for sharing a video from a hand-held device; the sharing person is moving the device and therefore the camera view could be different when the comments arrive.

Other research has looked into adding comments on video by analyzing the video content [Laiola Guimarães et al. 2012]. Kim *et al.* [Seungwon Kim et al. 2013] have explored how spatially aligned drawing on a live AR view can improve remote collaboration. However this did not include live text comments. Some researchers have explored how to easily add text and graphic annotations to recorded videos on mobile devices. For example MoVia [Cunha et al. 2013] allows people to draw on or add text tags to recorded video that can then be shared asynchronously. However the focus of applications like this is on annotation and not real time social sharing or live streaming.

One of the few examples of previous research into annotation on live video streamed from a mobile platform is the work of Huang [Huang and Fox 2012], who developed a system for adding text or drawing onto a live camera view and sharing it with a remote user. However in this case their research was focusing on the system performance and not an evaluation of the interface usability. The interface also did not support real time comment feedback and was not focusing on social networking.

In our research we want to place comments in a spatially aligned AR view on top of the live video feed. Using spatially aligned AR to add content to the real world is not a novel idea. For example, [Langlotz et al. 2013] used GPS coordinates to determine the position of a sound and positioned them spatially around the user. Similarly the AR browser applications Junaio[5] and Sekai Camera[6] allowed users to add AR comments in the real world. However, to our knowledge, no previous research on methods for commenting on live video has been done. Spatially aligned comments or annotations can benefit from understanding the surrounding 3D environment. For example, [Nassani et al. 2015] implemented AR tagging using Google Tango to track from the environment, and Google Glass to display AR comments, however this did not support real time video sharing.

Although previous work has demonstrated live video sharing on a mobile platform and support for viewer feedback, there has been little evaluation of different methods for providing feedback. In this paper, we report on investigations into different user interface (UI) options for viewing comments left by multiple users on a shared live video stream. Thus, the main contribution of the work is investigating if comment placement on live video sharing improves the user experience. In the next section we describe the prototype developed to explore this question.

[3] http://www.douyu.com/
[4] http://www.ingkee.com/
[5] junaio.com, unavailable since 2015. Acquired by Apple
[6] sekaicamera.com, unavailable since 2014. Evolved to http://tab.do/

3 System Design

We developed a prototype that enables a user to share a live video stream with others and receive comments from multiple users watching. Our system consists of a WebRTC[7] application running on AppEngine[8] on Google Cloud servers, which offers a fast peer-to-peer connection between devices. Being built on a web platform, this solution can run on multiple hardware specifications including desktop, hand-held, and wearable devices. Figure 2 shows the overall design of the prototype system.

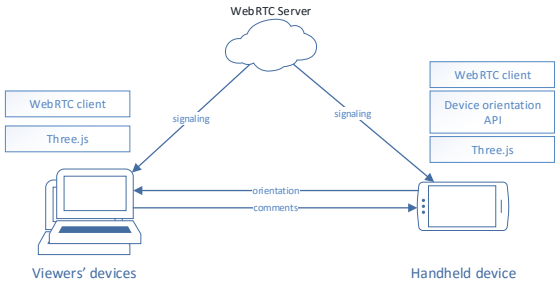

Figure 2: *System architecture based on WebRTC*

The prototype was built on top of AppRTC[9] which hosts a website that enables people to start a video conferencing session on the web. To support AR visualization of comments, we utilized the AppRTC code to track device orientation by listening to the device sensors. The AppRTC application is written in Python for the backend and Javascript for the front-end. It takes advantage of being hosted on AppSpot so that it complies with the WebRTC requirements for HTTPS. The AppRTC system allows users to communicate with each other over the Internet. In addition to the video stream, we modified the code to transfer the device orientation data to the receivers' devices via DataChannel. To render comments in the AR visualization, we used the Three.js library[10]. The AR visualization is implemented with two graphical layers. The background layer shows the video stream captured by the camera on the mobile device. On top of the background, comments are drawn on the front layer using orientation tracking information to show them in a body-stabilized manner [Billinghurst et al. 1998].

4 Implementation

The application starts by turning on the back camera on the mobile device. It then asks the user to enter a room number to start the connection. Once this is entered, the application will start a call mode, waiting for other participants to enter the same the room number. Once the call is established, the mobile device will start streaming video and device orientation data to the viewing PC.

Both users can send comments to each other by clicking on any part of the shared video. The system then calculates the 3D position of the comment in the AR space and waits for the comment text to be entered. Once the user enters the message, the text is displayed on both the sender's and receivers screens. The motion data of the senders device is also shared so that the receiver will see the comment appearing at the same place as the sender turns his/her device.

Three different ways of showing comments on the live video stream were implemented. A list view where the comments are listed on

[7] https://webrtc.org
[8] https://appengine.google.com/
[9] https://apprtc.appspot.com/
[10] http://threejs.org/

the side of the camera feed view. An AR implementation (AR) where the comments were overlaid on top of the video feed and rotated around the user based on phone orientation, so the comments appear fixed at the location on the video where they were first entered. Finally, an AR + list implementation combined the list view with the AR view. In the next section, we report on a user study exploring these three implementations.

5 User Study

We conducted a controlled within-subjects user experiment to test the different user interfaces for displaying comments. There were three conditions: L) comments in a list, AR) comments on the video with AR visualization, and L+AR) comments on both. The experiment started with the participants giving consent and answering questions about demographic information. Then they went through a training session to get familiar with the application and the experimental procedures.

To simulate different environments for the user, we used 180-degree panoramic images projected around the user on large screens to simulate different real spaces (see Figure 3). We selected four different images where the user might be interested in sharing his or her surroundings, varying in terms of indoors/outdoors and busy/quietness. A different background was randomly assigned for each condition between subjects.

Figure 3: *Participant during the experiment*

Each participant was asked to sit in the middle of the projection screens showing the background image, hold a smart phone, and aim its camera at the background to share it with remote users. The experimenter simulated multiple users sending comments on the shared video in a Wizard of Oz style setup. There were six predefined comments for each background. The comments appeared on the screen in three different styles depending on the experimental condition. The order of the conditions was counterbalanced using a balanced Latin square design. While watching the comments, the participant was asked to remember which part of the background each comment was talking about and who made the comment, which could be identified by the colour of the comment. There were up to four colours (commenters) in the experiment. The comments faded away one minute after being displayed. This was to simulate the user receiving multiple comments while having limited time to read them all.

After completing a condition, participants were asked to place a printed version of each comment on a background image, at the correct location, and with the correct colour, testing their knowledge of where each comment appeared. The participants were also requested to answer a questionnaire on system usability [Brooke 1996] and social presence [Harms and Biocca 2004]. The questions were slightly modified to fit the scenario being tested and only focused on one-way communication. Table 1 shows the social presence questions that were answered on a seven-level Likert scale rating (1: strongly disagree - 7; strongly agree).

Table 1: *Social presence questionnaire. Negative questions marked with (-)*

Q1	Comments from others were clear to me.
Q2	It was easy to understand comments from others.
Q3 (-)	Understanding others comments was difficult.
Q4	I could tell how others felt by my video sharing.
Q5 (-)	Others emotions were not clear to me.
Q6	I could describe others feelings accurately.

After finishing all three conditions, participants answered a post-experiment questionnaire that asked them to rank and compare all three conditions in terms of strengths and weaknesses. Finally, the experiment ended with a debriefing and the opportunity for participants to provide open-ended comments.

6 Results

We recruited 20 participants (11 female, aged between 19 and 35 years old, Median=27.5, SD=4.55). Most (95%) of them had experience with live video streaming a few times a week to a few times per month and 80% were familiar with AR applications. We used a non-parametric Friedman test for all the results with =0.05, and post-hoc tests using Wilcoxon signed-rank tests with the Bonferroni correction (=0.017)

The statistical result for SUS (see Figure 4) showed that there was a statistically significant difference between conditions ($X^2(2) = 9.658, p = 0.008$). Post-hoc analysis showed significant differences between L and AR (Z=-2.638, p=0.008) and between L and L+AR (Z=-2.559, p=0.010). However, there was no statistically significant differences between AR and L+AR (Z=-0.197, p=0.844). This shows that the list condition on its own was considered considerably less usable than the other two conditions.

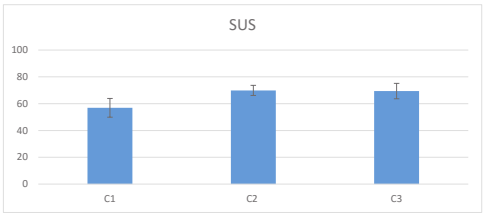

Figure 4: *SUS score*

As for the social presence questions (see Figure 5), we inverted the responses on the negative questions Q3 and Q5, to allow all questions to be aggregated, combining the answers for both perceived message understanding and affective understanding. There was a statistically significant difference in perceived social presence ($X^2(2) = 16.892, p < 0.001$). Post-hoc analysis found there were significant differences between L and AR (Z=-3.459, p=0.001) and between L and L+AR (Z=-3.311, p=0.001) while there was no statistically significant difference between AR and L+AR (Z=-0.427, p=0.670). This shows that the list condition (L) was perceived as being less easy to understand, and that viewer comments in this condition were less clear.

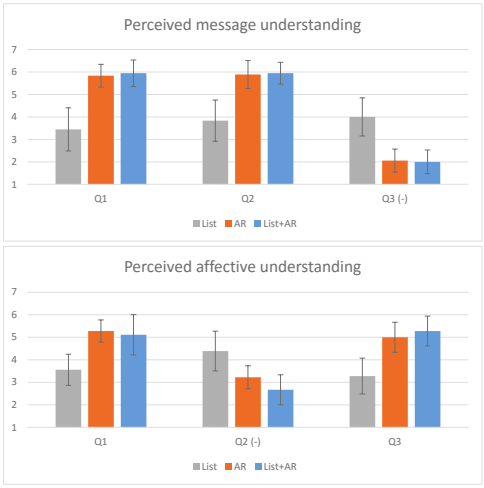

Figure 5: *Results for the social presence questions "perceived message understanding" and "perceived affective understanding"*

As for the ranking results (see Figure 6), we calculated the average of the answers (where 3=highest ranking, 1=lowest ranking). The results show a statistically significant difference between conditions ($X^2(2) = 9.100, p = 0.011$). Post-hoc analysis showed a significance level set at $alpha$=0.017. There were significant differences between L and AR (Z=-2.766, p=0.006) and between L and L+AR (Z=-2.502, p=0.012). However, there was no statistically significant difference between AR and L+AR (Z=-0.039, p=0.969). This shows that the list condition (L) was ranked the worst out of the three conditions.

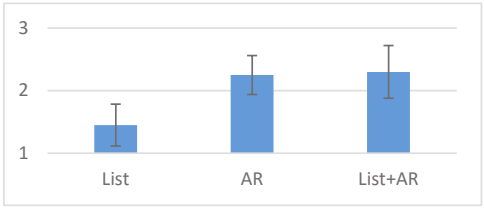

Figure 6: *Results for conditions ranking questions*

For the task of matching the position and colour of the comments (see Figure 7), The results show that there was a statistically significant difference ($X2(2) = 22.030, p < 0.001$). Post-hoc analysis showed that there was no significant difference between the L and AR conditions (Z=-1.016, p=0.310). However, there was a statistically significant difference between L and L+AR (Z=-3.628, p<.001) and between AR and L+AR (Z=-3.447, p=0.001).

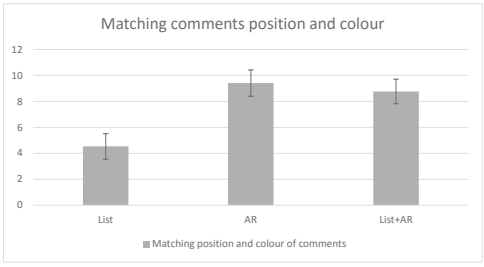

Figure 7: *Results for correctly matching comments with background and colour.*

Participants were asked free-form questions to comment on their experience in terms of the strengths and the weaknesses of each condition. Approximately 80% of feedback from the participants noted that in the list condition (L) it was more difficult to identify the area of the comments comparing to the AR conditions. Eight participants (40%) found it more challenging to remember the comment colours as a means to identify the person who sent the comment.

In the AR and L+AR conditions participants felt that the comments were contextual and relevant to the background. For example, *"It's easier to remember comments on video (AR) because the comments acts as cues on the video you can directly see what the people are commenting on which I think makes me feel more connected to them"*. One of the strengths of the L+AR condition commented on included having an overview of the list of comments even if they are outside the current viewpoint of the user.

However users felt that comments in the L+AR condition could clutter the UI and partially block the background. One participant said *"The screen just became too busy with comments that I don't have the time to actually sort out the comments and associate them on the video"*. Some suggested this could be resolved by making the comments not in the center of view more transparent.

We asked the participants what would they like to improve. Most reported that they would like to use a head-mounted display to view comments in the AR mode. It was also suggested that we use a profile image instead of colours on comments to distinguish remote users.

Overall users felt that the AR and L+AR conditions were fun and cool to use, providing comments such as *"It's pretty awesome. I love the experience and I would really like to use this app with my social network."*.

7 Discussion

From the user study, we found that subjects preferred the conditions that contained an AR view, compared to showing comments only displayed in a list format. They thought these conditions were more usable, provided a higher degree of social presence, and enabled them to better remember the comment layout. This is probably because the spatial association of comments increases the likelihood of the message being understood and being attended to.

We expected that one of the AR conditions (AR or L+AR) would have been more popular than the other, however this was not the case. Some users preferred L+AR over the AR as the former provided an overall list of comments even if they were not visible in the current user viewpoint; making the user more aware of new comments without needing to look around to find them. Other users preferred the AR only condition, as the screen is less crowded. One solution to this might be by hiding the comments on the list that are visible on the AR view, removing any duplication. Alternatively we could use a radar view that shows dots to represent comments.

We learned more about how to make the live streaming a better experience for the user. Some users found the one-minute timeout for the comments fading away to be too fast. Associating the comments with colour to represent different users may not be the best option. An alternative approach would be to use an avatar or name of the person to identify the comment source.

The study has a number of limitations that we will have to address in the future. The experiment was conducted in a simulated environment rather than outdoors. We also used a static background image to simplify the conditions. However in real life, things will be moving in the background (i.e. people walking, cars passing by). In

such scenarios, the comments in the AR condition may not stick to the moving objects. However, this could be solved by using image processing techniques to track objects that will allow the comment to be moved with them. Finally, the all of the comments were generated by an experimenter and were fixed, rather than coming from real people who could write whatever they liked.

8 Conclusions and Future Work

In this paper, we investigated AR annotations for social live video streaming. We conducted a user study testing three variations of the interface for showing comments: 1) a list, 2) an AR view and 3) both list and AR views. Participants felt that the AR and the List + AR conditions were significantly better than the List condition in terms of system usability and social presence. This was probably because the spatial alignment of comments increases the likelihood of them being understood and attended to.

In the future, we plan to investigate alternative mechanisms for communicating in a social live video sharing sessions, such as using sketching or emojis. We will also explore how depth cameras could be integrated into the system to enrich the social sharing experience by providing improved tracking and environment recognition. Finally, we would like to conduct more extensive user studies that test various user interface designs for using AR for sharing social experiences. This could include being able to navigate back in time to see comments before.

References

BERGSTRAND, F., AND LANDGREN, J. 2009. Information Sharing Using Live Video in Emergency Response Work.

BILLINGHURST, M., BOWSKILL, J., DYER, N., AND MORPHETT, J. 1998. An evaluation of wearable information spaces. *Proceedings. IEEE 1998 Virtual Reality Annual International Symposium (Cat. No.98CB36180)*, 20–27.

BROOKE, J. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry 189*, 194.

CUNHA, B. C. R., NETO, O. J. M., AND PIMENTEL, M. D. G. 2013. MoViA: A Mobile Video Annotation Tool. *Proceedings of the 2013 ACM symposium on Document engineering - DocEng '13*, 219–222.

HARMS, C., AND BIOCCA, F. 2004. Internal Consistency and Reliability of the Networked Minds Measure of Social Presence. *Seventh Annual International Workshop: Presence 2004*, 246–251.

HUANG, T., AND FOX, G. 2012. Collaborative annotation of real time streams on android-enabled devices. *Proceedings of the 2012 International Conference on Collaboration Technologies and Systems, CTS 2012*, 39–44.

LAIOLA GUIMARÃES, R., CESAR, P., AND BULTERMAN, D. C. 2012. "Let me comment on your video". In *Proceedings of the 18th Brazilian symposium on Multimedia and the web - WebMedia '12*, ACM Press, New York, New York, USA, 253.

LANGLOTZ, T., REGENBRECHT, H., ZOLLMANN, S., AND SCHMALSTIEG, D. 2013. Audio Stickies : Visually-guided Spatial Audio Annotations on a Mobile Augmented Reality Platform. 1–10.

NASSANI, A., BAI, H., LEE, G., AND BILLINGHURST, M. 2015. Tag it!: AR annotation using wearable sensors. In *SIGGRAPH ASIA 2015 Mobile Graphics and Interactive Applications on - SA '15*, ACM Press, New York, New York, USA, 1–4.

SEUNGWON KIM, LEE, G. A., AND SAKATA, N. 2013. Comparing pointing and drawing for remote collaboration. In *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, IEEE, no. October, 1–6.

X.CHENG, AND J.LIU. 2009. NetTube: Exploring Social Networks for Peer-to-Peer Short Video Sharing. *Infocom 2009, Ieee*, 1152–1160.