HOLOSUITE: A HOLOGRAPHIC CLOUD-BASED VIDEO EDITING SUITE FOR MICROSOFT HOLOLENS

A. H. K. Gulbrandsen, F. H. Jensen, J. M. C. Ågotnes, J. Vindenes

University of Bergen, Norway

ABSTRACT

Professional video editing suites of today are resource-demanding. A video editor needs great machine power in addition to multiple screens to tackle the varying formats of today's media landscape. Effectively, this results in reduced freedom of mobility for video editors; they are dependent on their stationary office space to work. In addition to reducing flexibility, this lack of freedom may slow down the turnaround process for news agencies.

In this paper, we describe an Augmented Reality (AR) cloud-based video editing suite, where up to five virtual screens are presented to the user through a Microsoft Hololens Head-Mounted Display. By employing cloud computing, the prototype can access machine power remotely through the cloud, which has benefits in terms of mobility. Effectively, the AR application is a prototype of an office for video editors that can be carried in a backpack, and utilized wherever there is network connectivity.

INTRODUCTION

Video editing require many resources. The rendering and exporting of video files require heavy machine power, and many video editors prefer to use multiple computer monitors in their workflow, with separate displays for video assets, timelines and video previews. Nowadays, we also see more and more devices connected to the Internet, and video editors need to take into consideration the varying screen sizes connected to the Web. Furthermore, we see the media landscape changing even more drastically with the introduction of Virtual Reality (VR) technologies, such as holographic videos and 360° stereoscopic (3D) videos. Video editors need to preview this content through an appropriate display, and to perceive 360° 3D content require Head-Mounted Displays capable of displaying content in AR and VR.

These complex equipment requirements effectively result in that video editors are stationary, and "chained" to their offices. Photographers and journalists have to distribute the material to video editors before the content is published because it is not possible for the journalists to do the edit in the field themselves, at least not in the same manner. Although video "MoJos" (mobile journalists) are reaping the benefit of mobility, this is notably with a compromise in quality. Using smartphone cameras and editing on the smartphones or laptops is not equal quality-wize to using high-end cameras and powerful machines. What MoJos do have, however, is the possibility to do the edit then-and-there, resulting in a quicker turnaround for the news agencies and a simpler workflow over all. Although it is possible to organise video editing suites on wheels, this is expensive and resource-demanding. The ideal solution would provide both the flexibility of the MoJos, and the power, speed and quality of traditional video editing.

Cloud Computing

Although enabling traditional video editors to become mobile is a complex challenge, there is a technology that can provide hope to its potential realization: cloud computing. Cloud computing is similar, and often intertwined with, the concept of cloud storage – only instead of providing storage, the service provides computational power. The unique benefits of cloud computing services is that their functions can be remotely accessed: one's own computing tasks are executed by a remote computer instead. Effectively then, the connected laptop controls the operations of the remote computer, and display the process and results. When utilizing cloud computing in video editing applications, the files on the laptop often act as so called "proxy" files representing the actual files, while the true files originate on the remote computer where they are processed.

Augmented Reality

Conceptually, the technology of cloud computing is only able to partially provide the ideal solution. What it fails to take into consideration is the workflow of the video editors: a laptop screen can not replace the multiple screens of the office, and stuffing three or more screens into a backpack also does not work very well. In this paper we will discuss a cloud computing solution that utilizes virtual screens through AR with the Microsoft Hololens. With AR technology, multiple screens can be placed virtually throughout the room, augmenting the reality in to a mobile office. In addition to providing several screens, this can simultaneously offer a way to view holographic video content or 360° 3D video content, as the Microsoft Hololens is a Head-Mounted Display capable of displaying spherical graphics.

In this paper we present related work before discussing the development of the "HoloSuite" application. Further, we present the application and an evaluation by a focus group of professional video editors from different media organizations. Finally, we discuss the implications and limitations of the technology, and discuss future work before concluding the paper.

RELATED WORK

In this section, we discuss related applications and research. The section will present and discuss works related to the concepts of applied AR, cloud computing and video editing in the cloud. The section is also used to thoroughly define the technologies employed in our solution.

Augmented Reality

Augmented Reality is best defined on a point of the "Reality-Virtuality Continuum" (see Fig. 1). While in Immersive VR, a user is encompassed in a wholly virtual environment, AR is less immersive in the sense that it features the real world as well as the virtual elements. In this sense, the world is *augmented* with virtual phenomena, such as holograms of graphical objects. AR is usually experienced with see-through Head-Mounted Displays such as the Google Glass or Microsoft Hololens (see Fig.2). VR and AR has different uses, as there are different benefits following different levels of immersion. AR, for instance, has the benefit of providing content while still being present in the real world. When entering a wholly immersive virtual environment through an Oculus Rift for instance, the whole world disappears, and navigating or orienting the real world is very hard due to no relevant sensory impressions. In this respect, AR is more useful for working purposes or other everyday scenarios that involves the real world.

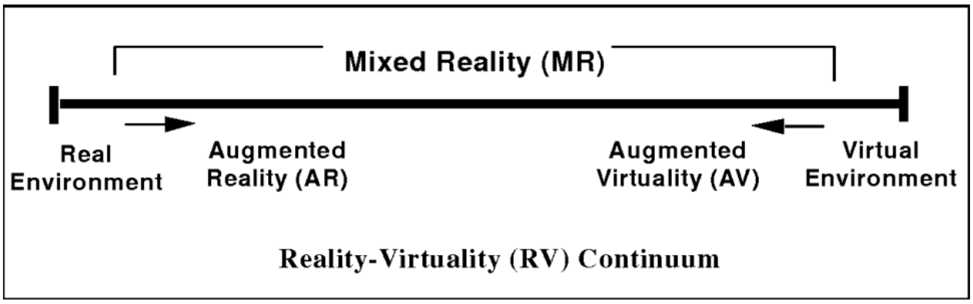

Figure 1 - The Reality-Virtuality Continuum (5)

Figure 2 - The Microsoft Hololens

This is represented within application development and research on how to apply AR technology. Surveying the field of AR, Krevelen and Poelman (2) mention industrial-, military-, medical-, entertainment-, and education sectors as sectors that benefit from use of AR technology. They also mention Personal Information Systems as another application area: for instance we can imagine the usefulness of augmenting our reality through GPS navigation embedded into the world, etc. Within this category, they also describe potential office solutions, which is the application area of our own work (2).

Cloud Computing

The idea of connecting AR to Cloud Computing is not a new idea. Luo (3) describes the idea of “augmented computing” on mobile devices, as a part of compensating for the computational limitations of mobile devices. The concept of augmenting the computation of less powerful devices has therefore been experimented with before on smartphones, which suffer the same computational limitations as the Microsoft Hololens. Within media production, Iligner-Fehns et. al (1) predicts a strong future role of cloud computing in the broadcast industry. They write that media production can be “extremely demanding, which requires to build very powerful processing resources” (1). Further, they write that employing cloud computing may aid in this challenge, and in addition make media production more flexible. They also write that “the flexibility of cloud-based applications will certainly impact today’s workflows” (1).

The need and utility of cloud computing in general is also seen represented in the commercial market. Amazon offers their “EC2”, Microsoft offers the “Azure” platform, and

Google offers their “AppEngine”. Products specifically targeted towards the broadcast industry is also increasing. In (1) several of these are listed: Avid’s “Media Composer Cloud”, Adobe’s “Anywhere”, “QCloud” by Textronix, “Zencoder” by BrightCove and many others. The authors of this paper also know of Viz Story by “Vizrt” and “Vimond IO” by Vimond, which both are professional cloud-based video editing applications. Works on mobile video editing through cloud computing is also apparent in research, for instance both Tsai and Yen (7) and Yu and Liao (8) describe different works on cloud-based video editing systems. In addition, the concepts behind cloud computing for video editing solutions are discussed and outlined in several other works as well (4, 6).

To summarize, cloud computing is already on its way into the broadcasting industry. Cloud-based video editing has benefits of mobility while still retaining the high machine power. In addition to this there are benefits of empowering mobile AR through cloud computing – and AR is increasingly used within many work settings to augment and aid the user in its operations. Albeit these results, we have not found any works that combine AR with remotely cloud computed video editing, as is the concept and prototype we will present in this paper.

APPLICATION

The HoloSuite prototype was inspired by an already-existing cloud-based video editing application for traditional web browsers. The initiative of the authors was to find use cases in the media industry for AR. During the development, the authors were situated at the company whose video editing application the prototype was inspired by. The prototype was built as a Universal Windows Platform (UWP) app, using Unity and the APIs available in Unity for implementing features such as gaze, gestures, voice input, etc. The application was then exported to Visual Studio where the app was built and deployed to the HoloLens.

When a user enters the HoloLens, the “HoloSuite” is seen installed as a native application on the HoloLens, represented by a logo. When the application is selected from the menu and is launched, the user is presented with an asset collection (see Fig 3).

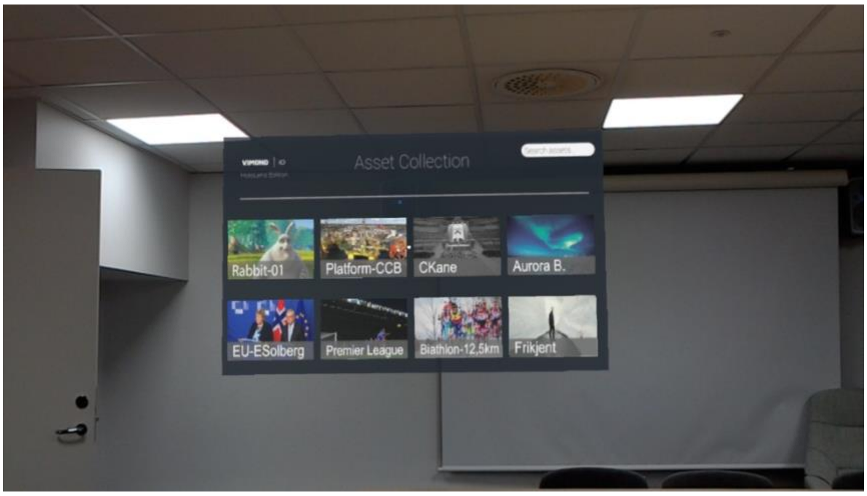

Figure 3 - The Asset Collection

When the user has chosen an asset, in this case the “Rabbit-01”, a new project is created and opened. The program also opens a video timeline, in addition to two preview windows of the video content (see Fig. 4).

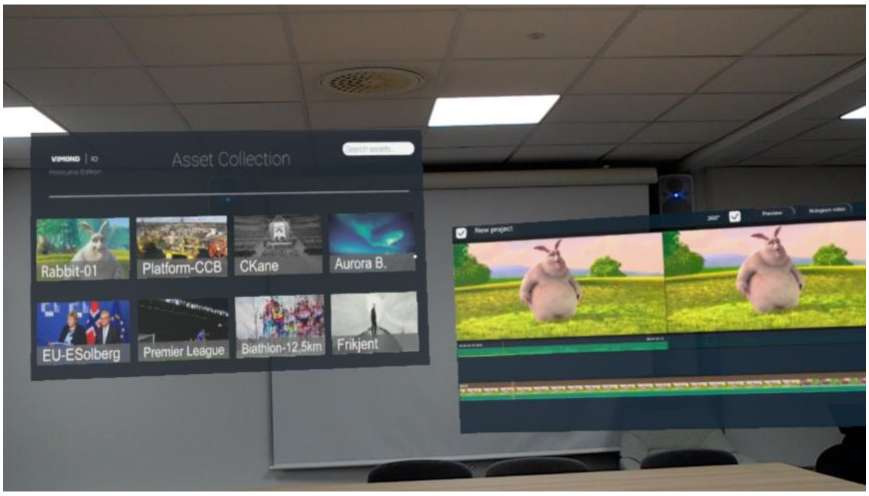

Figure 4 - Project window with two preview windows and a timeline

In the project window, the user can also click “Preview”, which brings forth three new preview windows - in total five - for previewing the video (see Fig. 5). These are in dimensions 9:16, 1:1, and 16:9. It should be noted that the “Preview” button is not very visible in the Figures of this paper, however, for the users this button is very visible – as the real HoloLens image is much larger when viewed embedded in the real world. The screenshot function in the HoloLens also provide images in poorer resolution compared to when viewed through the HoloLens.

Figure 5 - Three new preview screens, issuing 5 in total

INTERACTION

Gestures

In terms of application interaction, we employed two of the Microsoft Hololens Standard gestures into the application: AirTap and Bloom (see Fig. 6). The HoloLens “Airtap” is what traditionally amounts to a click. In the HoloSuite, to choose an asset or project, or enable more preview screens, an airtap is used. The bloom gesture is used to open and close the application.

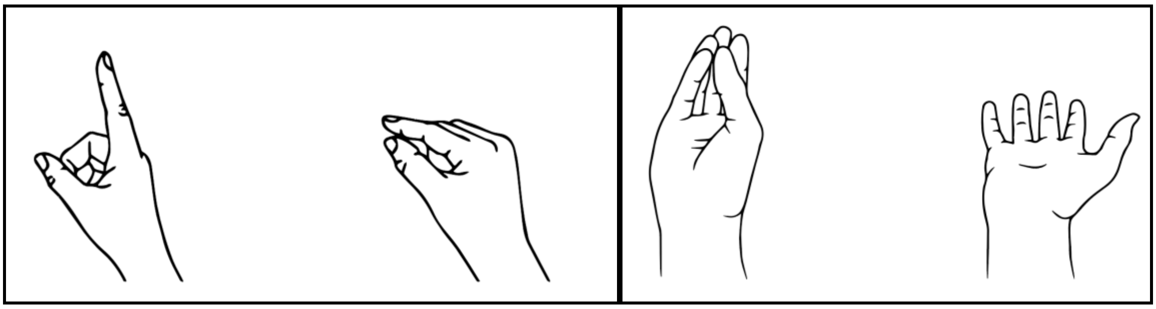

Figure 6 - AirTap Gesture (left) and Bloom Gesture (right)

Voice Recognition

Moreover, in addition to employing gestures for interaction, voice commands were also used. To play, pause and stop the preview of the video, a user can issue the three corresponding voice commands: “play”, “pause” and “stop”.

Physical Navigation and Orientation

Another means of interacting with the application that should be mentioned out of its significance, is physical navigation and orientation. As the HoloLens is a Head-Mounted Display, the user can orient the head in all 360° (horizontal and vertical) to see all possible virtual objects that are augmenting the reality. It should be noted, however, that these are absolutely positioned in the augmented environment, they are not positioned relatively to the eyes of the user. This means that the user can physically move around the virtual screens, and even move and lean in closer to get a better view. Examples of this can be seen in Fig. 4 and Fig. 5, which show the same virtual screen from the two different physical angles of the user.

Traditional Keyboard & Mouse

In addition to the natural interaction features, we wanted it to be possible use more traditional ways of interaction. To this purpose, we also connected a wireless keyboard & mouse to the HoloLens by BlueTooth. In this case, both mouseclicks and the space-button on the keyboard function as a “Play/Pause” toggle.

FEATURES

As of today, the “HoloSuite” prototype is still a relatively early work in progress, and is limited in its functionality compared to the cloud-based video editing application which it is based upon. As illustrated in Fig. 3 - 5, it offers an asset collection, creation of video projects, an illustrated timeline and preview of up to five virtual screens. The video files with sound that can be played and paused in the application, originate on a web server, and is thus accessed remotely through the cloud. As part of our investigation of video rendering through cloud computing, we can also access the corresponding video editing application it is based upon, when opened on a remotely located computer. In this case, the HoloLens can also be used to control such features as render, export, etc.

EVALUATION

To evaluate the prototype, we gathered a focus group comprising two video editors from different media organizations in Bergen, Norway. The aim of the evaluation was to discuss the user experience and features of the prototype, but also to use the prototype as a means to discuss the potential of the technology in the future. The editors were asked about how they perceived the interaction, the graphical quality in the HoloLens, and the general utility and usefulness of the application. The editors were asked to be critical and to suggest potential areas of improvement, both in terms of software and hardware.

Generally, both of the video editors stated that this was something they would have found useful to incorporate in their workflow already as it was. They liked the fact that it would free up space, and that they would be more free in terms of choice of working environment. The video editors also had many suggestions for improvement. One of the users pinpointed that the narrow field of view of the HoloLens made it somewhat difficult to keep an overview of the whole suite. This was alleviated by stepping back from the screens, which gathered them in the field of view of the HoloLens, but this was not ideal as he preferred to stay close to the timeline. Both of the users missed the ability to move the virtual screens around and change the crop, in other words they missed the ability to customize their workflow (they both also had different preferences on where they would have placed the different screens, which emphasized this need). In terms of interaction, both the users were enthusiastic about the use of gestures, and even suggested how the Bloom gesture in the future could be used to alternate between a 360° degree view and a flat equirectangular view of a 360° video. They also wanted the ability to move the virtual screens using gestures in this way as well. Moreover, the video editors would like have a darker background behind the virtual screens to make the office more immersive. When asked about the possibility for cooperation through AR, both the video editors were enthusiastic. They both agreed after discussion that it would be useful if this functionality could assign different roles to the two video editors, so that one had the editor role and another had the role of viewer.

DISCUSSION

The prototype described in this paper is a good example on the potential usage of technologies such as cloud computing and AR to solve challenges within the media industry. Although in its early developmental stages in terms of finished implemented features, the prototype works as a proof-of-concept on how mobile video editors or journalists can bring a virtual office with them in the field in their backpacks. By employing AR with cloud computing, traditional video editing can be performed with the same freedom that MoJos enjoy in their journalistic work. The technology may also have other implications: economically it may be possible that virtual offices may be even cheaper than traditional offices as they employ virtual rather than physical screens, and rents instead of buying computational power. Right now, however, it should be noted that Microsoft Hololens is only available for purchase in a Developer Edition at the price of 3000 USD - so this is not yet reality.

FUTURE WORK

When it comes to areas of improvement and future work on the prototype, there are several categories. The first one of these involve adding more editing features to the prototype, such as special effects, captions, sound editing, etc. There are also other features that could more innovatively exploit the features of the medium of AR, for instance collaboration. It can be imagined that another user, situated in for instance another city or country, could be represented virtually by an avatar in the room alongside the same workspace. This way of sharing holograms, and representing the user by avatar, may be an appreciated feature that can enable collaboration across distances. Further, by implementing the preview and editing of 360° stereoscopic videos and holographic videos, the utility of the AR Head-Mounted Display could be maximized. In such a case, the application could work as a novel way of editing content designed specifically for the more immersive mediums.

CONCLUSION

This paper presented “HoloSuite”, a prototype of a holographic cloud-based video editing suite for Microsoft Hololens. The application was developed as a means to explore solutions to offer video editors more mobility and flexibility in their work. With HoloSuite, a video editor can preview up to five virtual screens in different aspect ratios and sizes as a part of the video workflow. The application was created for the purpose of being connected to a cloud computing video editing suite. In this way, the lack of computer processing power on the Microsoft Hololens is not so much of an issue, as computer power is distributed over the cloud. An evaluation of the prototype highlighted the benefits of the AR app as mobility and freeing up space. The evaluation also isolated several possibilities of improvement, such as increasing the FOV (hardware of HoloLens), and being able to customize the setup of the virtual screens (software of the prototype).

REFERENCES

1. Iligner-Fehns, K., Schäfer, R., Keltsch, M., Altendorf, P., Polanec-Kutija, G., Vogl, A. 2018. The Forecast Is Cloud – Aspects of Cloud Computing in the Broadcast Industry. Digital Marketplaces Unleashed, Springer-Verlag, GmbH Germany 2018.

2. Krevelen, D.W.F., Poelman, R. 2010. A Survey of Augmented Reality Technologies, Applications and Limitations. The International Journal of Virtual Reality. 9(2). pp. 1-20.

3. Luo, X. 2009. From Augmented Reality to Augmented Computing: a Look at Cloud-Mobile Convergence. International Symposium on Ubiquitous Virtual Reality. pp. 29 - 32.

4. Malkin, J. 2014. Look to the Cloud: Enabling Seamless Video in a Multidevice World. SMTPE Motion Imaging Journal. 123 (7). pp. 41-45.

5. Milgram, P., Kishino, F. 1994. A Taxonomy of Mixed Reality Visual Displays. IEICE Transactions on Information Systems. 77 (12).

6. Sinclair, A. 2013. Deploying Video Platforms in the Cloud. Annual Technical Conference and Exhibition, SMTPE.

7. Tsai, C. J., Yen, Y.N. 2014. Cloud-Based Video Editing System. Advances in Computer Science and Its Applications, Lecture Notes in Electrical Engineering 279. pp. 1423-1428.

8. Yu, L., Liao, X. 2013. Cloud based Mobile Video Editing System. IEEE 9th International Conference on Mobile Ad-hoc and Sensor Networks. pp. 378 to 382.