HIAR/VR FOR VARIOUS VIEWING STYLES IN THE FUTURE OF BROADCASTING

H. Kawakita, K. Yoshino, D. Koide, K. Hisatomi, Y. Kawamura, and K. Imamura

NHK, Japan

ABSTRACT

This paper shows that our prospect of various viewing styles, broadcasting service image, and R&D cases in AR/VR era. It is important for R&D to adopt both deductive approaches and inductive approaches to improve one's own technology and to discover a new problem respectively. Describing our three works by the former and our four works by the latter in the past and present, we give multiple viewpoints of AR/VR R&D in the field of broadcasting.

By an inductive approach, we expect that AR/VR media will connect the viewers with each other even beyond time and that handling various forms of 2D video will become important in AR glasses era. To handle them easily, we propose a presenting method to use equirectangular images with alpha channel (transparency), which can include almost all types of 2D video. To increase presence of an object in the image, we extend the format for depth sense in a binocular head-mounted display or AR glasses.

INTRODUCTION

In this paper, we discuss how viewing styles will be and what the technology and the R&D strategy is required in the future broadcasting. AR/VR is expected to be one of key factors of future media in the living room next to 4K/8K. To develop AR/VR technologies for broadcasting, we have been adopting two kinds of approach: deductive one and inductive one.

In the deductive approach, the future is drawn on the extension of R&D which we have been engaging in. It could also be said the one improving performance by taking advantage of our technical strength for sustaining innovation. Since this approach is aimed at realizing AR/VR, we call it "For AR/VR".

On the other hand, in the inductive approach, at first, we expect potential needs in the present or the future from the trend of technology and society. Next, we consider what service image we should provide. After that, we set to work on R&D of a necessary technology for the service. We may not take advantage of our technical strength, but it might generate something inducing disruptive innovation for new viewers. The target as a media company is to provide value through broadcasting, or communication between family members with broadcasting content, etc. It is in the upper layer than the one in which the goal of "For AR/VR" like presence is. Therefore, since we regard AR/VR technology as one of the means, we call the approach "By AR/VR".

In this paper, we introduce our R&D cases of "For AR/VR" and "By AR/VR" in the past and present, draw viewing styles in the future broadcasting, and discuss the role of media in the AR/VR age.

PAST WORKS

To consider about the future AR/VR technology and strategy for broadcasting, we introduce four R&D trials in the past as case studies.

"For AR/VR" case #1: VR viewing system using 8K display

Head-mounted displays (HMDs) for 360 VR has left room for improvement from various viewpoints. Particularly, it is a well-known fact that the resolution is not sufficient for comfortable viewing that gives high sense of immersiveness and reality. Here is a case utilizing our leading research on 8K as presented by '(1)'.

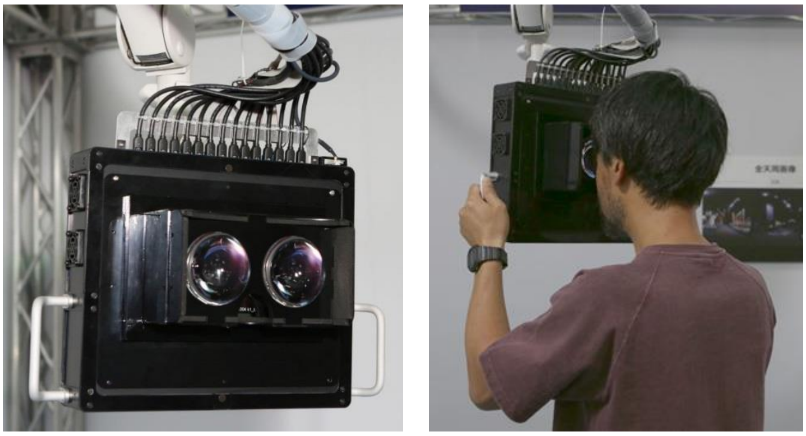


Figure 1 – HMD for 8K VR

When viewing part of 360 images by an HMD with a wide field of view, in order for users not to perceive the pixel structure, significantly high spatial resolution is required for an HMD. This leads to the necessity of much more spatial resolution of 360 images. To cope with the problems, as shown in Figure 1, we developed a prototype HMD using an OLED with a spatial resolution of 8K (1,058ppi) and captured a 360 image with a spatial resolution of 30K x 15K. This result was referred to as an example in the discussion for the Recommendation ITU-R BT.2123-0.

"For AR/VR" case #2: AR broadcasting synchronization system

There are various works and standards as presented by '(2)' in which the broadcast program can be enriched with the additional content through Internet. Synchronization between the broadcast program and broadband content is one of the main topics. Here is the AR application case utilizing our leading research about MPEG media transport (MMT) by 'Kawamura and Otsuki (3)'.

In this case, as shown in Figure 2, the viewer will be able to enjoy a new viewing experience in which broadcast program on a TV and 3DCG content, including AR content, on a mobile device as a second screen are synchronized according to coordinated universal time (UTC). One of the features is that we utilized a regulation that the video frames over MMT have UTC-based presentation timestamps (PTSs).

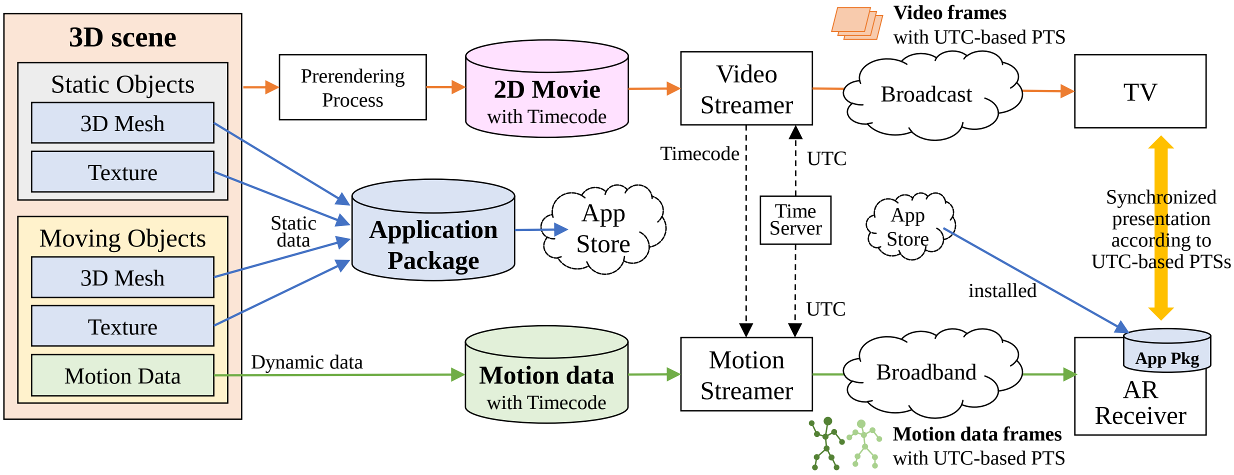
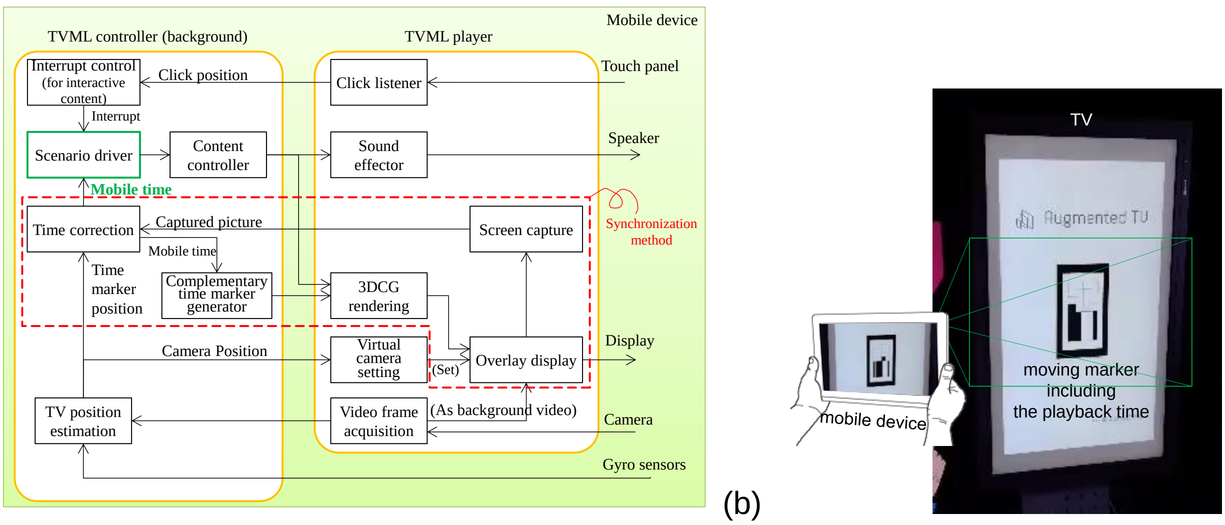


Figure 2 – System model of synchronization between broadcasting and AR content

"By AR/VR" case #1: Augmented TV

We developed Augmented TV 'Kawakita, et al (4)', to realize such a dream "If only characters on TV could jump out into our living rooms…" As shown in Figure 3, Augmented TV is an AR technology for the service in which the viewer can see as if TV world is connected to the real world in front of the TV. It has two main functions: synchronization of the frame presentation time between TV image and 3DCG content on mobile device as AR content and estimation of the relative position and orientation between TV and mobile device. The feature of synchronization is that the original moving marker including the playback time of TV program is displayed on the TV as a part of the TV program. By scanning the marker with the mobile device, the viewer can enjoy the synchronized AR content.

This case is similar to "For AR/VR case #2" regarding synchronization between TV and mobile device. However, the approach is reversed. In "For AR/VR" case #2, the broadcast



(a) (b)

Figure 3 – (a) System architecture in mobile device, (b) Moving marker for synchronization

program on TV does not need to be changed for the synchronization. On the other hand, in this case, since we developed technology specialized for the service, it can be applied not only for live-broadcasting but also for time shift viewing, package media, or digital signage, etc.

"By AR/VR" case #2: 360 video delivering system to see outside the frame

Here is another augmentation system of a TV program. As shown in Figure 4, NHK provided a VR service in which the viewer can see what is outside the frame of the broadcast program on TV with a mobile device, as presented by 'Yamaguchi (5)'. Using the developed delivery system for 360 video for a mobile device, the viewers could see the video outside of the frame in conjunction with the broadcast program.

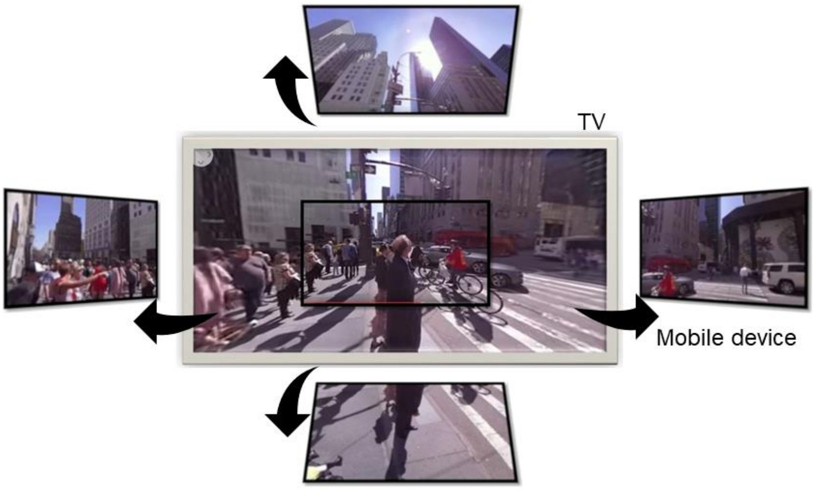


Figure 4 – VR service for seeing outside the frame

PRESENT WORKS TOWARD THE FUTURE

We describe three of our present R&D cases toward the future including expectation of the future and our service image, which are important especially for "By AR/VR" approach.

"For AR/VR" case #3: High-immersive VR production using 8K cameras

On VR video shooting, it is an ideal to shoot objects with a single camera without stitching to avoid artefacts from misalignment. However, there is no device to shoot 360 degrees. To reduce stitches, high-resolution cameras are frequently demanded to achieve high-

	Selection of lens and arrangement of 8K cameras	(1) "wide-field" (original) Fish eye lens	(2) "360" Fish eye lens 180°	(3) "stereoscopic 180" Fish eye lens interpupillary distance (fixed)
Feature	Resolution	Excellent	Good	Good
	Immersive	Good	Excellent	Good
	Sense of depth	Fair	Fair	Excellent
	Compatibility	Good (even only center camera)	Fair	Fair
	Interoperability	Good	Fair (cameraman is always visible)	Good
	Data format	Fair (needs to standardize)	Good	Good

Table 1 – Shooting trial using 8K cameras

resolution VR video. We have tried to shoot the VR video using 8K cameras. As shown in Table 1, we shot several camera configurations selecting different lens and camera alignment of 8K cameras and evaluated the video of various scenes. Although we cannot ignore the fact that the results depend on the scene, we expected technical issues in producing high-resolution VR and will build a technical shooting plan for comfortable video.

Viewing styles with AR glasses in the future

AR glasses have recently received broad attention because they are regarded as one of the most likely alternative prospects to the smartphone. We discuss how viewing styles will be in AR glasses era. Figure 5 shows the expected main functions of AR glasses in the future. The most significant difference from the smartphone is the high degree of freedom in spatial representation. For example, AR glasses can substitute for a general display to view 2D video image. As presented by 'Kawakita (6)', we call 2D video image arranged in 3D real space by AR glasses "Virtual TV". Table 2 shows the summary by classifying various viewing experiences including from a conventional one to a new one expected in Virtual TV era. Figure 6 shows the illustrated coordinate systems in Table 2. As can be seen from Table 2 and Figure 6, the feature is the coordinate system with the origin located at the viewer. Therefore, we call the system VCS (viewer coordinate system), focusing on VCS in the following, we

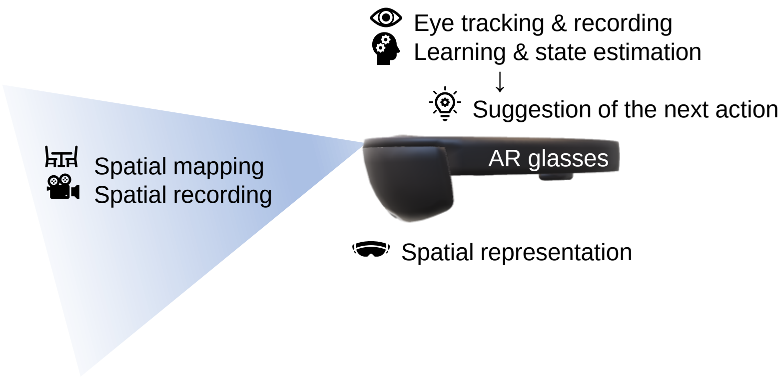


Figure 5 – Expected functions of AR glasses

Example of viewing experience	Category	Ideal coordinate system	Screen size*
TV program, playing game	concentration on viewing	World, Environment, Head	S – L – 360
searching, referring something	reference	World, Environment, Object, Head, Wrist	S
SNS, E-mail	private information	Vehicle, Head, Wrist	S
video chat, communication in virtual space	expression of face/body	World, Environment, Head, Wrist	S – L – 360
signage, advertisement	signage	World, Object, Environment	S – L
virtual pet, virtual partner	character	Body, Head (tag along)	S

↓ new experience

*S: smartphone - tablet, L: TV, 360: 360 spherical panorama

Table 2 – Summary by classifying various viewing experiences

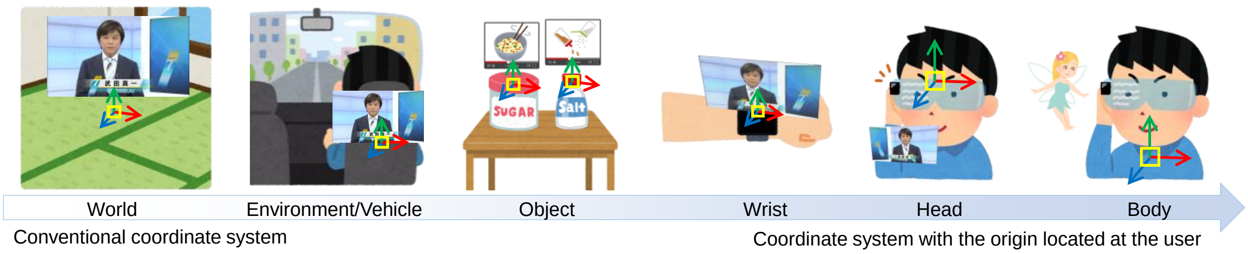


Figure 6 – Examples of coordinate systems

discuss service image in the future as a media company.

Service image of viewing together with sharing space

One of the values of media has been said for a long time to be provision of common topics. Especially, TV media has been gathering people in the living room, and they communicate with each other in the same space. It is important for a media company to provide not only information but also service in which viewers share the same space even in Virtual TV era.

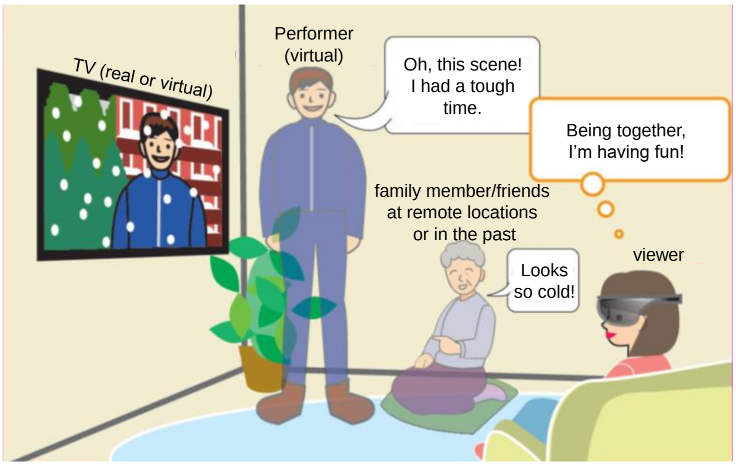


Figure 7 – Service image of viewing together with sharing space

Figure 7 shows a service image in such future. In the service, a broadcasting company broadcast the program for TV and additionally provides the 3D models or the 2D human bodies, such as a performer, a reporter, and an economic analyst, in the program for AR glasses. The viewer can enjoy their performance in life-size with AR glasses. The viewer can also enjoy them with someone else using similar technology, who can be the family members/friends at remote locations, the family members/the viewer him/herself in the past, or a stranger interested in the program. Then, the company may provide the metadata allowing the viewer to invite a suitable person easily.

Content display control

As the necessity of OS (VROS) dedicated to VR/AR is discussed in 'Kapoor (7)', new methods of content display control in consideration of the high degree of freedom in spatial representation will be required in the Virtual TV era. Figure 8 shows the purpose and role relating to content display control as a hierarchical structure. In the VROS/app layer, the display is initially optimized with the default coordinate system, and functions for replacement or resizing of content are offered to the upper layer. In the content layer, according to the meta data with content, the display is controlled to maximize the production intention and/or to restrict viewer's operation. In the user layer, the display is

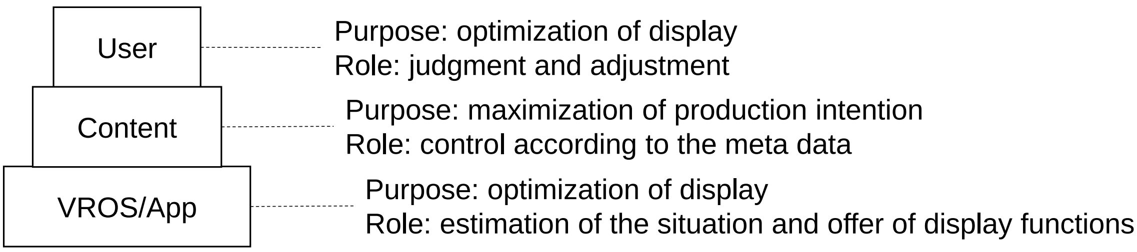


Figure 8 – Purpose and role relating to content display control

finally adjusted such as replacement or resize of content or switching the coordinate systems as needed. The manual adjustments would shift from the user layer to the VROS/app layer by AI technology with the times. Considering VCS like “Head” or “Body” in Figure 6, tag-along in ‘(8)’, a movement to follow the viewer’s movement with a slight delay, would be included as an example of display functions of VROS. In the next section, focusing on VROS/app and content layer, we, as a content production company, discuss an inclusive format of 2D content in VCS.

“By AR/VR” case #3: Utilizing 360 video with alpha-channel

As written above implicitly, various types of 2D content would be mixed in the Virtual TV era. There are, for example, rectangular 2D image and 360 video like “By AR/VR” case #2 or rectangular 2D image and human body without background like Figure 7. Taking diversity of types of content or device into account, NHK call a new concept of broadcasting service corresponding to such era “diverse vision” ‘(9)’.

To deal with various types of 2D content easily, we propose utilizing 360 (equirectangular video) image with alpha channel (transparency) in VCS. As shown in Figure 8, a 360 image with alpha channel in VCS can include almost all types of 2D video from the perspective of shape and representation for the viewer, such as a conventional rectangular image, a 360 image, and even a free form image. Since we can use some existing codecs for the 360 image with alpha channel, such as VP8/WebM, a new file format is not required. Furthermore, if most of the image is transparent, the image data can be compressed at a high compression rate. For example, in the case that there is only one human body, file size is about 1-2Mbps in VP8/WebM, equirectangular 4K, and 30fps.

To increase presence of an object like a human body in a 360 image with alpha channel in VCS, just introducing one parameter, we propose a new format for distance sense. As also shown in Figure 8, the parameter *d* is the distance from position of the viewer (same as

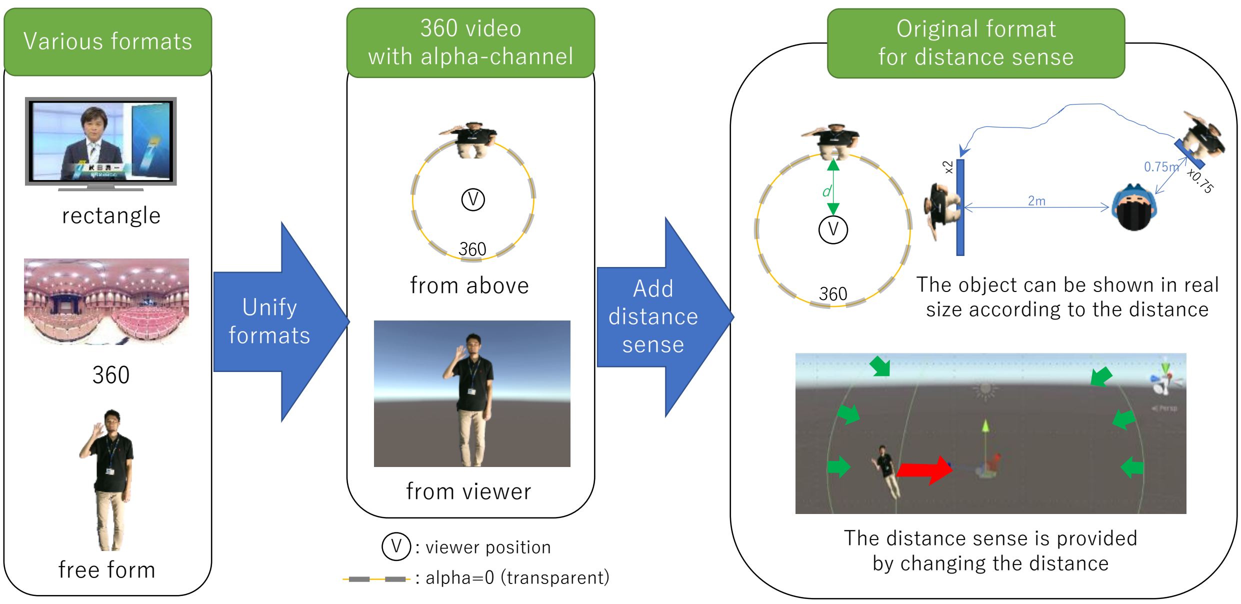


Figure 9 – Utilization of 360 video with alpha channel and the format extension

the origin) to the spherical image (as a virtual screen) and defined as a function of time. The content provider specifies the value of *d*, that the viewer will enjoy the content using distance sense in VCS. Concretely, there are two ways for the viewer to feel distance sense as follows: with binocular HMD, the viewer perceives the parallax between the object on stereoscopic 360 image with alpha channel and the background in virtual space, or with AR glasses, there is an occlusion between the object on 360 image and a real object.

Providing 10m as maximum value of *d* and 1mm as resolution unit of *d*, the quantifying bit depth is 14, and the amount of increase of the bit rate is 1.3 kbps under 90 fps, which is negligible compared to the video.

"By AR/VR" case #3.1: Prototyping an app moving 2D video spatially

As shown in the above case #3, to examine what effect 2D video moving 3D-spatially has, we prototyped an AR app with Microsoft HoloLens. As shown in Figure 10, we configured the app with recording and playback mode on the HoloLens. In the recording mode, the app not only shoots forward with an RGB camera on the front of the HoloLens but also logs the position of the object from the HoloLens with the infrared sensors, in which the position is associated with each frame of the shot video. In the playback mode, the app plays the video at the respective position on the HoloLens. As the object looks correct, the size of the video is increased proportional to the distance like Figure 9. Perceiving the parallax of the video, the viewer can feel the distance sense on the HoloLens. As above, the app can record and play the object with the position, which is the same as the viewer saw.

Figure 11 shows screenshots of the app. In the recording mode, the child went back and forth from a person wearing the HoloLens, who saw her. The app recorded the video in which she is captured as the object and the position. In the playback mode, a person

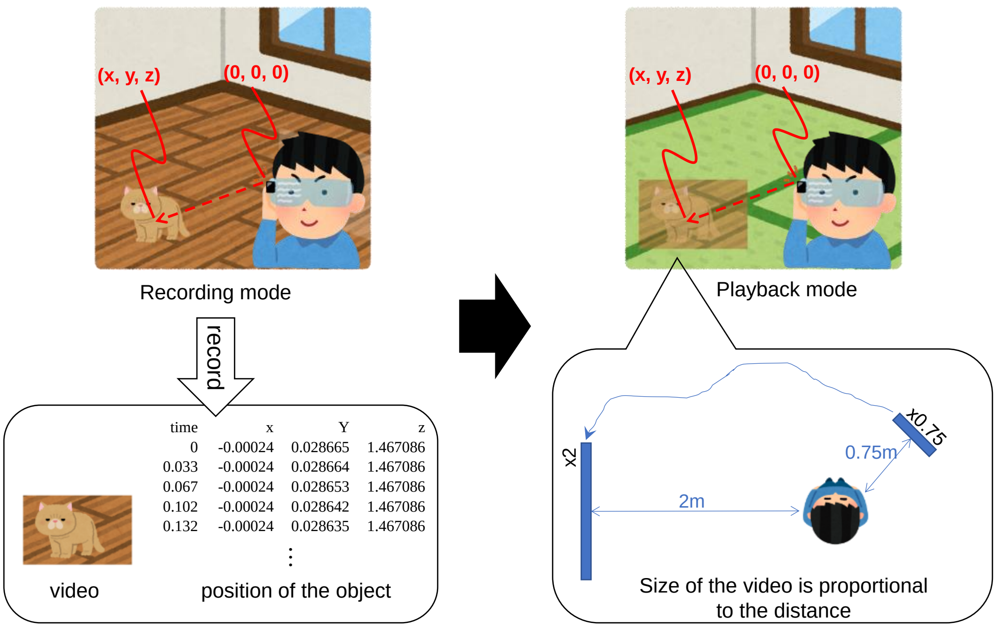
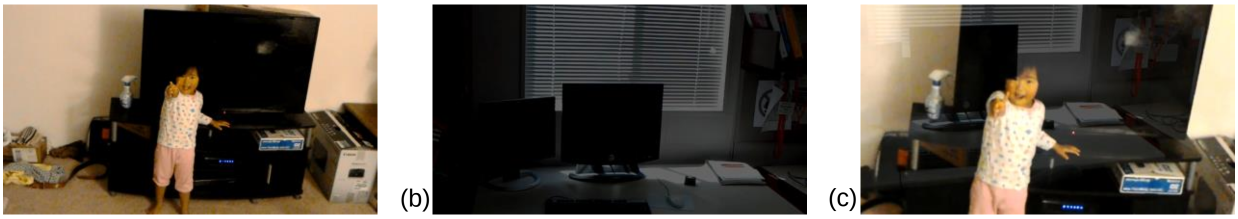


Figure 10 – Prototyped app to examine what effect 2D video moving 3D-spatially



(a) (b) (c)

Figure 11 – Execution environment (a) Captured video (b) A room in the playback mode (c) Field of view through HoloLens in the playback mode

wearing the HoloLens confirmed that she was reproduced on the video with the movement in a dimly lit room. The observer looked down and up according to her movement, where the movement of the head seemed to contribute to the increase of her presence.

"By AR/VR" case #4: Sharing VR experience

In regards to viewing together using a sharing space, we gave the example using AR glasses above, also here is the case with an HMD for VR by 'Yoshino (10)'. There is a well-known problem in VR like 360 images, that it is difficult for the viewer to enjoy the content together with others in the same place. In the case, as shown in Figure 12, the viewers need to wear an HMD on which a stereo camera is mounted. By depth map from the stereo camera, the system displays the image of real objects close to the viewer like a video see-through AR and the 360 images in the remaining area, which is at a certain distance away from the viewer.

As shown in Figure 13, the viewer can see the viewer nearby, such as a friend, together with 360 images. For example, if the friend points at something, the viewer can see in the same direction at the same time. Therefore, the viewers can have an impressive sharing experience with the viewers nearby.

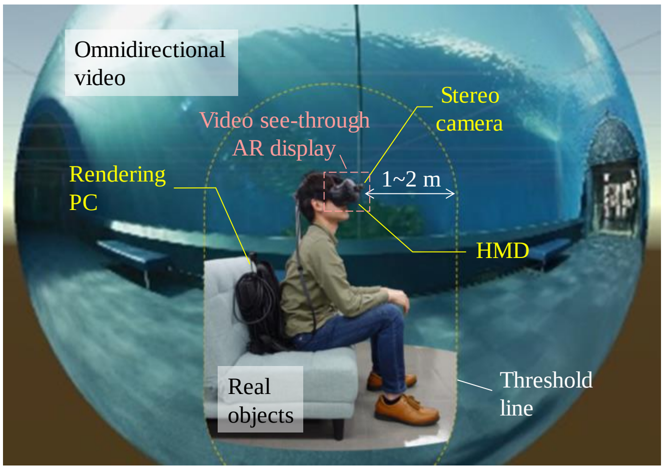


Figure 12 – VR sharing system

Figure 13 – Difference with/without HMD

CONCLUSIONS

We have discussed how viewing styles will be in the future with some of our AR/VR R&D cases. Since AR/VR media has a high freedom of representation that can include conventional media, AR/VR is sometimes called "metamedia". TV media have been connecting spatially between the points in front of the camera and a TV. AR/VR media will connect between the viewers each other in the same space and even beyond time as well.

Since there are many work elements in the AR/VR R&D field, both "For AR/VR" and “By AR/VR” are important. “For AR/VR” is the conventional approach, but we still have some clear remaining problems including resolution of HMD. “By AR/VR” needs us to consider not only within the technological field, but across fields.

REFERENCES

1. 2018. Collection of usage scenarios and current statuses of advanced immersive audio-visual systems. Rep. ITU-R BT.2420-0. pp. 9 to 11
2. 2019. Integrated broadcast-broadband system. Rec. ITU-R BT.2075-2.
3. Kawamura, Y. and Otsuki, K., 2018. 3D Free-viewpoint Augmented Reality on a Second Display Synchronized with a Broadcast Program. Proc. of the 4th ICCIP. pp. 273 to 277
4. Kawakita, H., Uehara, M., Nakagawa, T. and Sato, M. 2015. Development of a TV System Augmented Outside the TV Screen. Proceedings of the 22nd IDW. pp.1449 to 1452
5. Yamaguchi, M. 2017. The Meaning of VR: Delivering 360 Degree Videos by Public Service Broadcasters Towards 2020 and Beyond. https://www.nhk.or.jp/bunken/english/reports/pdf/report_17121201.pdf
6. Kawakita, H., Fujisawa, H. and Hollerer, T. 2017. Research Challenge of 2D Moving Picture on AR Glasses as a New TV. ITE Winter Annual Convention 2017. 21C-2
7. Kapoor, A. and Sharma, S. 2017. Implementation of a Virtual Reality Operating System (VROS) for the Next Generation of Computing. 2016 6th International Conference - Cloud System and Big Data Engineering. pp. 731 to 736
8. Windows Dev Center. Billboarding and tag-along. https://docs.microsoft.com/en-us/windows/mixed-reality/billboarding-and-tag-along
9. NHK STRL 3-Year R&D Plan (FY2018-2020) https://www.nhk.or.jp/strl/pdf/3kanen_e.pdf pp.3 to 4
10. Yoshino, K., Kawakita, H., Koide, D. and Hisatomi, K. 2019. Sharing immersive VR experience among local users by using a video see-through AR display. APMAR2019. (poster & demo)