Diversifying Viewing Styles and Expansion of Media Services –Diverse Vision for Future Media–

Kohji Mitani

mitani.k-gw@nhk.or.jp
1-10-11 Kinuta, Setagaya, Tokyo 157-8510, Japan
Keywords: Diverse Vision, 8K, 3D TV, AR/VR, Space Sharing

ABSTRACT

NHK Science & Technology Research Laboratories (STRL) is working on R&D of a new media service, Diverse Vision. It will deliver a rich variety of content experiences via 2D or 3D TV, AR/VR, or any emerging device technology. This article introduces the concept and research initiatives of Diverse Vision.

1 INTRODUCTION

The novel coronavirus that appeared last year has caused a pandemic that instantly changed lifestyles around the world. Teleworking and working from home are being encouraged, with people being urged to avoid nonessential, nonurgent outings and to stay at home. Teleconferencing and SNS-enabled video communication are now used widely not only for work, but also for school, seminars, and even communication with family and friends. Friends are also sharing time together through PCs and smartphones, each preparing their own food and drinks, and having a lively time, just as though they were gathered around a table at a restaurant or pub. More than ever before, communication in virtual space has become important in addition to communication in real space and a part of the “new normal.”

To make such experiences and communication in virtual space richer, it is essential that we improve the performance and functionality of sensing technologies, which can obtain accurate information regarding the bodily senses, and device technologies, which can faithfully reproduce such information from digitized data.

NHK STRL has been promoting R&D on 8K UHDTV, which provides a sense of presence and reality as the ultimate 2D television technology. In December 2018, the first ever regular 8K broadcasts via satellite started in Japan. NHK STRL is now taking this opportunity to establish and begin R&D on our new “Diverse Vision” concept for a future broadcast media service as a successor to 8K Super Hi-Vision, which will enable anyone, at any time and any place, to view and experience various types of content on the device of their choice including 3D television and AR/VR.

This article will give the background and describe the concept of Diverse Vision, then introduce our current research on space-sharing technology as a core of this new paradigm.

2 DIVERSE VISION

NHK STRL is conducting research toward a media service called Diverse Vision [1], a future media service delivering a rich variety of content experiences via visual, audio, and even haptic devices. The time frame we envision for this service is around 2030 to 2040.

In the past, the evolution of broadcast technology has proceeded along two primary axes: improving the quality of images and sound and extending functionality via digital technology. In the future, we will also incorporate communications and networking technologies such as 5G and the internet, and advance in new directions, expanding the expressive space (Fig. 1). In particular, by advancing 3D information processing and user interactive functionality, images will evolve from 2D to 3D, and information will evolve from audiovisual to multimodal. Thus, we will diversify viewing styles and provide information that is better suited to users’ needs and viewing environments.

In the future society anticipated by Diverse Vision, receiver devices will go beyond conventional 2D televisions of HDTV and UHDTV to handle spatial information, as with 3D televisions, AR, and VR. People can look forward to completely new media experiences that overcome the constraints of time and space, experiencing worlds never seen before, enjoying programs together with the performers or other viewers

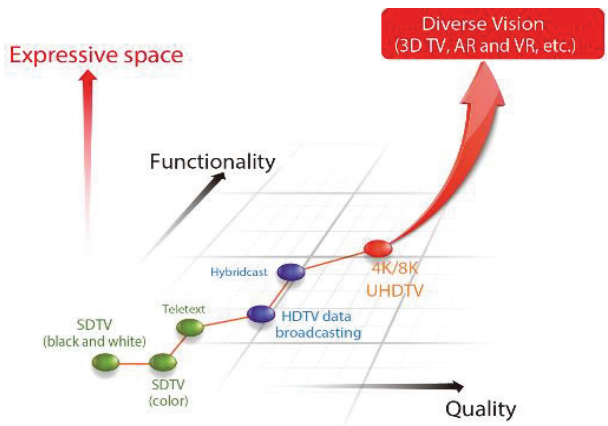

Fig. 1 Development of broadcast technology

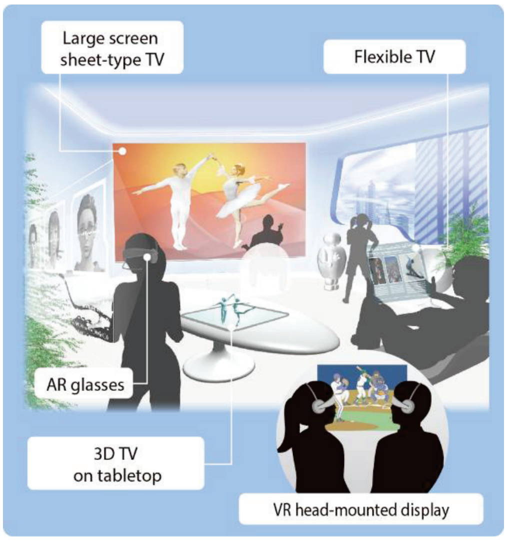

Fig. 2 Concept of Diverse Vision

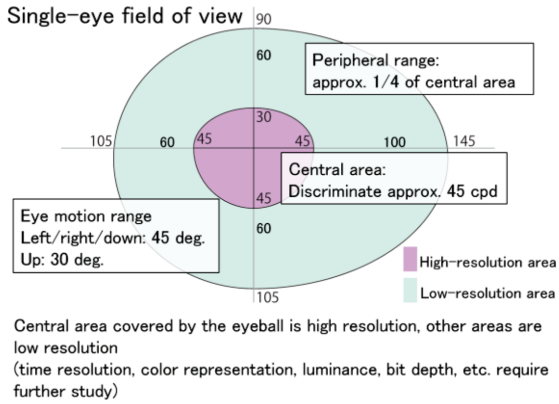

Fig. 3 Required specifications for "ultimate" HMD

in different locations, as though they are together in the same room, and receiving reliable information in the form that is needed, anytime and anywhere (Fig. 2).

3 RESEARCH AND DEVELOPMENT OF NEXT-GENERATION MEDIA DEVICES: SPATIAL INFORMATION SHARING SYSTEMS

In order to realize the viewing experiences expected from Diverse Vision, we will need systems that provide realism comparable to actually being there and that also enable multiple viewers to share the same experiences.

First, it will be necessary to build environments that produce such immersive realism. 8K has an increased sense of presence, reality, and immersivity through high-resolution images and a wider angle of view. An effective way to further increase this sense of immersive realism is to add an element of interactivity between the viewers and the displayed images. By presenting high-definition 8K images and 3D sound in addition to the feature of a head-mounted display (HMD), which matches the images to the orientation of the viewer's head and body, we should be able to evoke a sensation of being transported into the world of the video.

By utilizing results of experiments and discussions of visual perception and visual function, we have attempted to derive specifications for an HMD that can present an ultimate virtual space indistinguishable from reality by focusing on image resolution (Fig. 3).

As shown in Fig. 3, the human visual field of view of a single eye [2-4], considering the effective field of view and also the motion of the eyeball with the eye position fixed, covers 145° on the ear side, 105° on the nose side, 90° upward, and 105° downward (light green region of figure). However, the maximum spatial frequency that humans can perceive by a single eye is known to be 45 cycles/degree [cpd] in the center of the field of view, with the ability to discriminate decreasing to one-quarter of this value toward the periphery. On this basis, the pixel density required in the center and at the periphery of the visual field would be 90 pixels/degree [ppd] and 22.5 ppd, respectively. Furthermore, considering the range of motion of the eyeball and the relationship between the field of view and required resolution, we can calculate the number of pixels required in the entire field of view to be approximately 70 million per eye. As this number is quite large, we will continue to carry out evaluation experiments using the curved screen mentioned later and derive a practical number of pixels for an ideal HMD.

We next consider "space sharing for program viewing", which will enable multiple viewers to share this sense of presence and reality when viewing a program. Here, we define space sharing for program viewing as a state in which viewers who are actually in the same room or in the same virtual space through the use of AR or VR technology, share the same background information (context) for the program.

Here, viewers sharing the same space are considered to be able to convey their level of interest or emotion to each other through speech, facial expression, or other means. As such, even if viewers are not necessarily in the same real space, by sharing their individual virtual spaces through AR or VR technology, which harbors great potential for spatial expression, they should be able to convey interest and emotion through speech and facial expression in the same ways as if they were actually close to each other.

By also sharing contextual information related to the program, the emotional response to the content will also increase through the shared viewing experience. For example, viewers can share their excitement by speaking with each other through broadband communication networks while watching live coverage of Olympic competitions at the same time. Broadcasting disseminates the same information at the same time, and by combining it with state-of-the-art broadband communication networks such as 5G, we can build an environment that creates shared experiences.

Fig. 4 Simulated large half-cylinder screen experience

By combining more advanced AR and VR devices with broadcasting and broadband in the future, we will be able to implement space sharing for program viewing, enabling us to enjoy content with family and friends at remote locations.

4 RESULTS OF SPACE-SHARING RESEARCH

To enable experiences in which multiple people are transported into the space depicted in images, we prototyped a system that projects high-resolution images onto a large curved screen, as shown in Fig. 4. We projected images captured with three 8K cameras onto the large screen with a diameter of approximately 11 m and height of 4 m [5]. Viewers were able to see the images and experience greater immersivity than ever before and rated the experience very highly. We are also developing a more compact system using a curved OLED panel, and intend to continue research in the future, studying factors that induce a sense of immersion and space sharing from the perspectives of visual psychology and cognitive science.

As an example of research on space-sharing services, we introduce a system that utilizes a see-through-video HMD and enables users to share viewing of omnidirectional video with people who are nearby in real space (Fig. 5) [6,7]. The see-through-video HMD is a VR HMD with stereo cameras mounted on the front and displays images from the viewpoint of the eyes directly on the HMD. This produces an effect similar to that of AR glasses. The stereo cameras are able to measure the distance to the subject in real time, and in areas closer than a set distance, such as 1 or 2 m, the images captured by the stereo cameras are displayed directly. In areas farther than the set distance, omnidirectional images are displayed. In this way, viewers can see other nearby viewers as well as what they are looking at and their reaction. This gives the impression that they are watching the omnidirectional video together even when each person is wearing an HMD.

To evoke a sense of sharing a space, it is also important to reproduce the spatial sound faithfully [8,9]. With VR and AR, one can not only hear sound sources from in front, but also turn around and listen to them from behind. A viewer can hear quite a difference in a person's voice when in front of or behind the person. To enable a viewer to hear the sound of a source from any location with the correct spatial impression, the sound must be reproduced with volume and frequency characteristics that correspond to the viewer's listening position relative to the sound source. Therefore, we are considering building a database of the emission characteristics from a sound source to any viewer's position (Fig. 6). We have measured such emission characteristics for a Japanese speaker as the sound source from 128 points with the viewer's location on the upper hemisphere from the horizontal plane at ear height.

In the future, we will continue research on full-object audio, which will enable the detailed separation and control of sound objects, and technology to reproduce sound sources processed with their emission characteristics, regardless of whether the sound sources are near or far from the viewer.

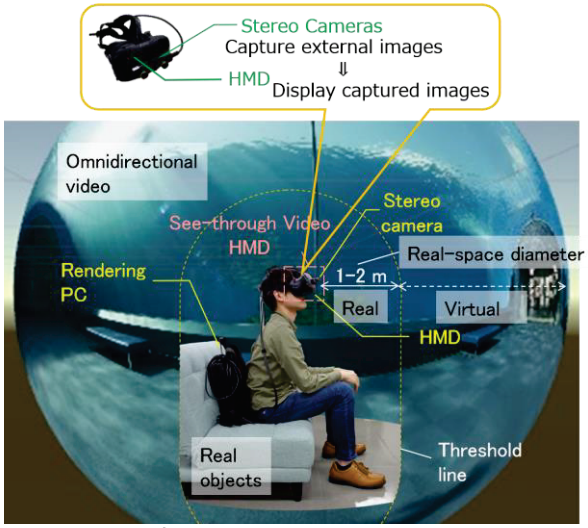

Fig. 5 Sharing omnidirectional images

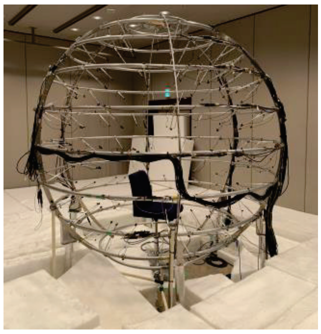

Fig. 6 Equipment for measuring emission characteristics from sound source

5 CLOSING THOUGHTS

In this article, we have introduced the concepts of and some of our research on Diverse Vision, which is part of our research on future media services. We expect that the quality and diversity of media services handling spatial information will continue to increase in the future with advances in a wide range of ICT technologies, but there are still many issues that must be resolved in implementing them.

One difficulty is that the creation of content for this new medium will not only involve capturing images and sound, but also require advances in sensing technologies that can obtain new 3D information, such as the surface characteristics of objects and lighting conditions, in real time, so that objects can be processed in accordance with the spatial conditions. Spatial signal processing technology will also be needed to reproduce the space on the basis of this information by processing and composing the information. To support these technologies, there is also a need for new computing technologies that will realize signal processing at ultra-high speed, and transmission technologies for exchanging this information.

Meanwhile, on the user side, there is a need for immersive display devices that will enable viewers to enjoy virtual spaces that they can hardly distinguish from real spaces. In this article, we described specifications for an ideal HMD on the basis of the relationship between the field of view and resolution, but other factors beyond resolution will also need to be studied, such as time resolution, color expression, brightness, bit depth, and the speed of response to head movement. It is also necessary to develop viewing devices that can be used for long periods of time with less eye strain and wearing fatigue, as well as technologies to reduce their weight, size, and power consumption.

Research for Diverse Vision has only just begun, so there are still many issues to be resolved before the practical implementation can be realized. Nonetheless, by solving them through R&D, we will attain seamless connections between real and virtual spaces and establish new media technologies that will enable us to move freely between them.

REFERENCES

[1] NHK STRL Open House 2019, Exhibition E1, 2019, https://www.nhk.or.jp/strl/open2019/tenji/e1_e.html

[2] H. M. Traquair, “The normal field of vision,” in *An Introduction to Clinical Perimetry*, M.D., F.R.C.S. ed., London, UK, Henry Kimpton, pp. 4, 1938.

[3] W. J. Lee, J. H. Kim, Y. U. Shin and H. W. Lim, “Differences in eye movement range based on age and gaze direction,” Eye, Vol. 33, pp. 1145-1151, 2019.

[4] V. Virsu and J. Rovamo, “Visual resolution, contrast sensitivity, and the cortical magnification factor,” Experimental Brain Research, Vol. 37, pp. 475-494, 1979.

[5] D. Koide, H. Kawakita, K. Yoshino, K. Ono and K. Hisatomi, “Development of high-resolution virtual reality system by projecting to large cylindrical screen,” IEEE International Conference on Consumer Electronics (ICCE), DOI: 10.1109/ICCE46568.2020.9043096, 2020.

[6] H. Kawakita, K. Yoshino, D. Koide, K. Hisatomi, Y. Kawamura and K. Imamura, “HIAR/VR for various viewing styles in the future of broadcasting,” International Broadcasting Convention (IBC2019) Conference, Tech Talk: VR meets AR, Real and virtual spaces become one, 2019.

[7] K. Yoshino, “AR/VR-based space-sharing services: Watching programs together with performers and distant friends and family members,” NHK STRL Bulletin, Broadcast Technology, No. 81, Summer 2020.

[8] S. Sakamoto, S. Hongo and Y. Suzuki, “3D sound-space sensing method based on numerous symmetrically arranged microphones,” IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, Vol. E97-A, No.9, pp.1893-1901, September 2014

[9] F. J. Fahy, “Other application of sound intensity measurement,” in *Sound Intensity*, 2nd ed., London and New York, UK and USA, Spon Press, pp.250-253, 1995.