ARENA-STYLE IMMERSIVE LIVE EXPERIENCE (ILE) SERVICES AND SYSTEMS: HIGHLY REALISTIC SENSATIONS FOR EVERYONE IN THE WORLD

Jiro Nagao[1], Kiyoshi Tanaka[1,2], Hideo Imanaka[3]
[1]NTT Service Evolution Labs, Japan, [2]Otsuma Women's University, Japan, [3]NTT Advanced Technologies, Japan,

Abstract *– Immersive Live Experiences (ILEs) enable audiences at remote sites to feel real-time highly realistic sensations, as if they are at the event site. This article provides a key feature of an implementation of ILE services, called Kirari! for Arena, as a use case of arena-style ILE, and its technical elements developed by NTT Labs. The major functionalities are object extraction from an arbitrary background, object tracking with depth sensors, low-latency synchronized data transport and four-sided pseudo-3D image presentation with depth expression. It also provides evaluations on the experience of Kirari! for Arena audiences, as well as its conformance to International Telecommunication Union, Telecommunication Standardization Sector (ITU-T) standards for ILE.*

Keywords – AI, immersive live experience, MPEG media transport, pseudo-3D images, real-time image extraction, standardization.

1. INTRODUCTION

The worldwide spread of TV enabled people to watch events happening anywhere in the world almost at the same time, which was unimaginable to people in the past. However, the experiences that TV can provide are still far from real. Information regarding the real size of things, natural sounds at the site, and the other sides of objects cannot be depicted by a 2D display.

Viewing styles of large events have been evolving recently, that is, a number of large music concerts and sports events have been relayed to remote sites for public screening. Even if the remote sites are far from the real event site, the audience can share their emotions while watching the event. Public screenings can be held in movie theaters or in front of large public screens.

If, in addition to video and audio, environmental information of an event such as vibrations and feelings of excitement is reproduced at the remote site, the experience of remote public screening can be further enhanced, resulting in the audience feeling as if they are at the real event site.

This feeling of being in a highly realistic atmosphere is called immersiveness, and enjoying an event in a remote site in real-time with immersiveness is called an Immersive Live Experience (ILE). Research and development to implement ILE has been conducted, and a number of ILE-related technologies have been standardized to ensure the interoperability between ILE service providers and to facilitate the market.

This article provides a key feature of an implementation of ILE called "Kirari! for Arena" [1], that provides an audience at a remote site with the unprecedented, highly immersive experience of watching a stage located in the middle of the audience from all directions. This article contributes by providing evaluations of audience experiences of Kirari! for Arena, as well as its conformance to International Telecommunication Union, Telecommunication Standardization Sector (ITU-T) standards for ILE.

2. STANDARDIZATION ACTIVITY OF ILE AND RELATED WORKS

2.1 ITU-T H.430 standard series

To provide interoperable ILE services worldwide, it is necessary to establish international standards on ILE systems, especially their requirements and framework. The ITU-T H.430 series specifies the requirements, framework, service scenarios, and MPEG Media Transport (MMT) profile for ILE [2][3][4][5]. The series of Recommendations is listed as follows.

- H.430.1: Requirements of ILE services.
- H.430.2: Architectural framework of ILE services.
- H.430.3: Service scenarios of ILE including use cases.
- H.430.4: Service configuration and MMT profile for ILE.

The first two (H.430.1–H.430.2) Recommendations provide the basic concept of ILE, and ITU-T H.430.3 provides service scenarios. ITU-T H.430.4 is the first to specify the technical aspect of ILE including how to transport additional information such as 3D positions of objects in synchronization with the audio and visual stream. More Recommendations with technical specifications may follow.

ITU-T H.430.1 defines the term "ILE," and specifies the high-level requirements of ILE. The characteristic experiences ILE provides are as follows.

- Watch the event from a remote place in real time.
- Feel the event as if it is happening right there, with a considerably higher sense of realism.
- Share excitement with people around them.

2.2 Related works

Various studies have been conducted to realize immersiveness, and a number of technologies have been applied to commercial services. 3D movie theaters as commercial services have become popular [6][7][8]. Moreover, vibration, light, air flow, as well as 360-degree 3D displays have started to be used in commercial theaters [9][10]. These 3D theaters require the audience to wear Head Mounted Displays (HMDs) or 3D glasses.

There have been a number of studies on watching sports events with HMDs [11][12]. In these studies, the audience can see from the viewpoint of athletes or from a free point of view. However, those approaches have a problem in which the HMDs prevent the audience from communicating with those around them while watching the event, resulting in a difficulty of sharing their emotions between themselves.

Other studies have focused on improving the immersiveness and comprehension of scenes by reproducing vibro-tactile sensations [11][13][14]. These technologies can realize ILE if the sensations of real events are transported and reproduced in real-time along with video and audio. Effectively combining these technologies with Kirari! for Arena can also further enhance its immersiveness.

Highly realistic experiences without HMDs have been an active research target as well [15][16][17], including a number of forms of multi-angle display devices [16][17][18].

These commercial services and studies, however, do not realize ILE because they either could not use a real event scene as input, could not process and show images in real-time, or did not have a real-time transport system to send the necessary content to remote sites.

3. TECHNOLOGY SUITE FOR ILE

This section introduces the Kirari! technology suite. Kirari! has been developed to provide technology components to realize immersive experiences. This section also introduces a number of past events where Kirari! demonstrated its ability to realize immersive experiences.

3.1 Kirari!

The Kirari! technology suite, which can produce a number of implementations of ILE, is a suite of technologies developed by NTT that are necessary to reproduce events at remote sites with high realism [19]. The suite consists of technologies to capture, measure, process, transport and display an event.

Figure 1 shows the technology map of Kirari!

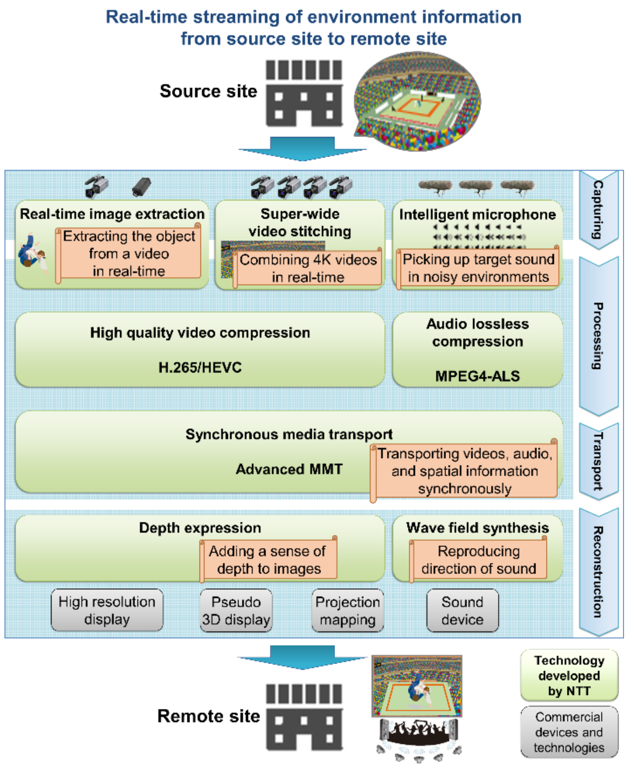

Fig. 1 – Kirari! technology suite

Real-time image extraction can identify the region of an object (e.g. performer on the stage) in an image in real-time so that the object image can be used to reconstruct the scene at the remote site.

Super-wide video stitching can produce super-wide (e.g. > 12K) video by combining multiple camera images in real-time for such events as super high definition live streaming.

Intelligent microphones can pick up sounds of interest in a noisy environment, which can be used, in turn, by wave field synthesis to reconstruct the direction of the sound so that the sound comes from the object.

3.2 Immersive events by Kirari!

Kirari! has been demonstrated in many events to create highly realistic sensations. For example, in "Chokabuki" events [20][21][22], various effects were provided, such as "double characters" in which real-time image extraction was used to produce a duplicate of a Kabuki performer next to the real performer to simulate two versions of the same character on the stage at the same time, and sound effects by wave field synthesis technology to reproduce a sound field including its direction to simulate the sound coming from within the audience floor.

The "KABUKI LION" event demonstrated the world's first pseudo-3D presentation through remote international transport using Kirari! technology. Real-time image extraction was used on the video stream of a Kabuki performer, and the pseudo-3D image of him accommodated the press remotely [20].

Another example is "Futsu-no-sukima" [23], an exhibition aiming to create novel value, where infrared-powered real-time object extraction enabled about 1000 people wearing various items of clothing to experience joining a teleconference with the room background of their choice [24].

4. KIRARI! FOR ARENA

4.1 Overview

One of the presentation styles ITU-T H.430.2 describes is the arena style where there is an arena or stage with a surrounding audience in a remote site for events such as a Judo match or juggling street performance.

"Kirari! for Arena," which is an implementation of the arena-style ILE, is realized by using the Kirari! technology suite. Kirari! for Arena can provide an immersive experience where an audience can watch an event from all (front, back, left, right) directions at a remote site with high realism as if they were watching it at the real event site. Implementation details of Kirari! for Arena can be found in [1].

Figure 2 shows the displayed image on the display device (right), and the real object (left). The real object in this case is the real performer and the displayed image is reproduced by Kirari! for Arena. With Kirari! for Arena, the audience can observe remote events from all directions on a special four-sided display device. The audience can watch each side of the performer from the corresponding side of the display. In addition to the mandatory requirements of an arena-style ILE, Kirari! for Arena can reproduce the texture of the "floor" such as the mats in a Judo event.

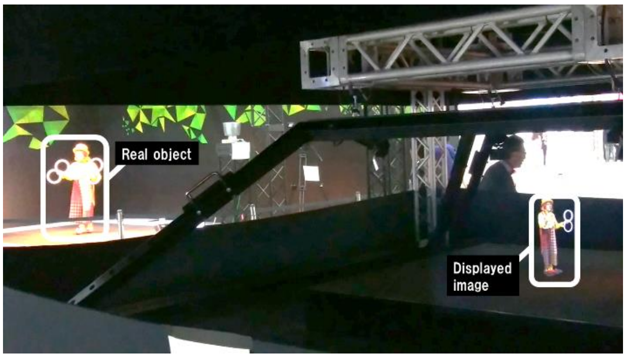

Fig. 2 – Real-time live display by Kirari! for Arena

Figure 3 shows the capturing and viewing configurations. A position sensor is installed along with the cameras to measure the coordinate (3D) position of the object. By transmitting and processing these images and position information in real time, the audience at the remote sites can feel a sense of depth in the images displayed on the four-sided display device.

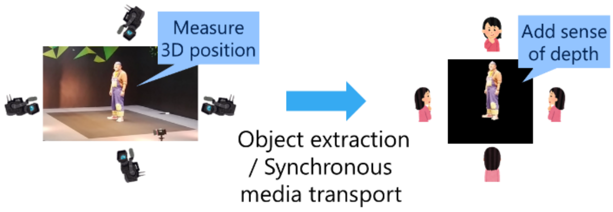

Fig. 3 – Capturing and viewing configurations of Kirari! for Arena

4.2 System configuration

Kirari! for Arena consists of four functions; Capture & Extraction, Measurement & Tracking, Information Integration & Transport, and, Depth Expression & Presentation. Figure 4 shows the functions of Kirari! for Arena.

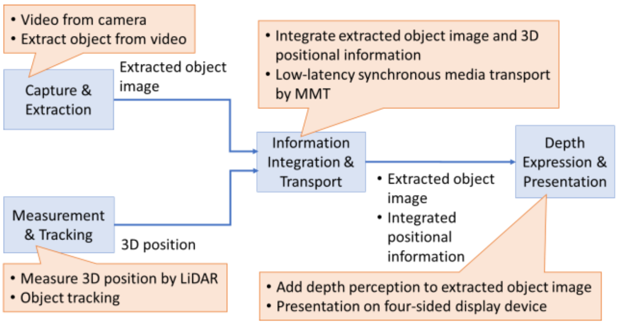

Fig. 4 – Functions of Kirari! for Arena

The Depth Expression & Presentation function uses a special four-sided display device, which plays a major role in giving a sense of reality to the reconstructed images.

The display device is composed of four half-mirrors as shown in Fig. 5 combined in the shape of a quadrangular pyramid, and each half-mirror reflects the image of a flat display installed above it, facing down. The reflection presents an aerial virtual image of objects to the audience (Fig. 6). Inside the half-mirrors, a physical plate (stage) is installed as the "floor" of the virtual images.

The device uses the principle of Pepper's ghost [16][17] to show pseudo-3D images on an arena stage without HMDs. The principle uses the reflection of a half-mirror to create a form of illusion by showing both the real objects behind the half-mirror and the reflected image at the same position. There are other approaches to show 3D images such as those that use an array of lenticular lenses [25]. However, the Pepper's ghost principle was selected for the following reasons.

Kirari! for Arena requires the views from all four directions to be shown at the center of the stage, that is, the four images of the performer should be positioned at the same physical spot. Transparent LED panels cannot be used because four of them would have to be placed at the center, intersecting with each other, which is physically impossible. 3D displays with lenticular lenses can show aerial virtual images outside of their own physical devices but still need the devices behind the images. Therefore, even if the four floating images can be placed at the center, the devices behind the images would obstruct the views. The Pepper's Ghost principle does not cause such problems and enables aerial virtual images to be placed at the center of the stage, intersecting with each other without obstruction. This principle was also selected because of its cost efficiency and scalability.

Fig. 5 – Four-sided display of Kirari! for Arena

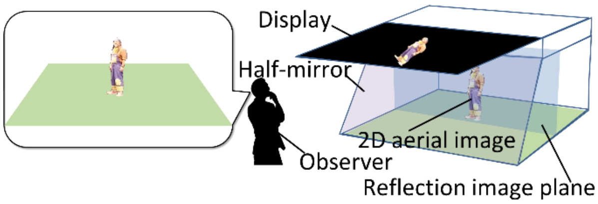

Fig. 6 – Device configuration to show 2D aerial image

Here, the audience perceives the aerial image as a real object because the image reflected by the half-mirrors appears to be floating in the air without any support. The Depth Expression & Presentation function also processes the object images to increase the sense of depth.

To enable the audience to perceive as if a real object is there, the object needs to be finely extracted. In addition, real-time processing is essential to realize a live experience. Real-time object extraction is conducted by the Capture & Extraction function.

The Measurement & Tracking function measures and tracks the 3D position of the object to provide the depth information (3D position information) used by the Depth Expression & Presentation function.

A large-capacity, low-latency media transport system is indispensable for displaying high-resolution video streaming at remote sites in real-time. Kirari! for Arena requires transport technology that can send four video streams synchronously. The Information Integration & Transport function integrates the extracted object images from the Capture & Extraction function and the position information from the Measurement & Tracking function, and transmits the large amount of data to the Depth Expression & Presentation function in a low-latency and synchronized manner.

4.3 Capture & Extraction function

Kirari! for Arena processes the captured video to isolate (extract) only the object in real time from an arbitrary background to create a video stream that is the source of the virtual image. For this purpose, the Kirari! Integrated Real-time Image Extraction System (KIRIE) was developed [1]. KIRIE can finely extract objects from 4K videos with arbitrary background images in real time. The performance of the extraction process is illustrated in Fig. 7.

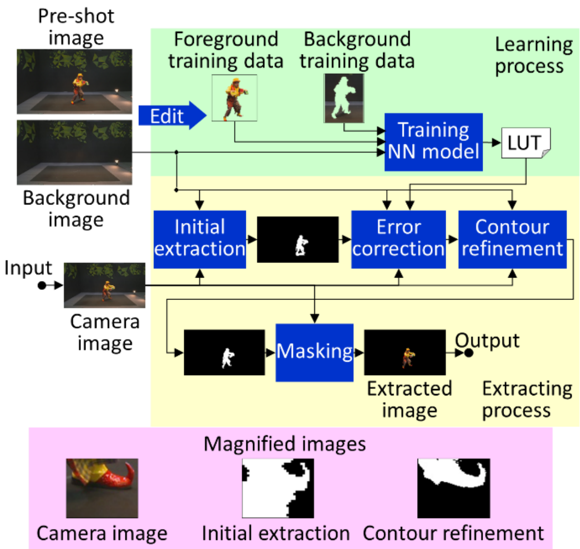

Fig. 7 – Extraction process

First, KIRIE classifies the input images from the camera into binary values by background difference as the initial extraction.

Artificial Intelligence (AI) is used to correct the extraction errors from the initial extraction [26]. More specifically, it classifies the pixels by a Neural Network (NN). The NN is a 10-layer convolutional network with 6-dimensional input (RGB of foreground and background pixels) and 2-dimensional output (the posterior probability of being foreground and background). The NN is trained in advance by the pixel pairs of the foreground and background images at each coordinate position. To prevent the repeated evaluation of the NN during the error correction process, a Look-Up Table (LUT) was generated with the quantification of 6-bits for each of the RGB values. The LUT considerably reduced the computation load and helped to achieve real-time processing.

Then, to refine the contours of the objects, KIRIE reclassifies them as the correct foreground or background pixels by referring to the colors and labels in the vicinity [27][28].

Finally, the object regions are extracted from the original color image by masking the other regions in accordance with the classification result.

The extraction error of the object extraction compared with that of the conventional method has improved by 29–48% [26].

4.4 Measurement & Tracking function

This function measures the 3D positions of the objects at the event site and tracks their positions as they move. It uses a Laser imaging Detection And Ranging (LiDAR) device to measure the positions [29][30], track the objects, and output the object labels along with the 3D position information.

4.5 Information Integration & Transport function

This function integrates information such as the object extraction result from the Capture & Extraction function and the 3D position information of the object from the Measurement & Tracking function. The integrated data is transported to the Depth Expression & Presentation function with time synchronization and low latency. This realizes a real-time synchronized display of video from four directions.

Position information of objects is transported by a special profile of the MMT protocol tailored for ILE. This profile is designed to transport metadata such as position information of objects and lighting control signals [31][32]. The video and audio encoded/decoded by a low-latency HEVC/MMT codec system [33] are also synchronously transported along with the metadata by MMT. A video size of 3840 x 2160 pixels at 59.94 fps was encoded by HEVC with a bitrate of 10 Mbps. The syntax of the metadata is defined by ITU-T H.430.4 [5].

4.6 Depth Expression & Presentation function

This function adds the depth expression to the extracted object images in accordance with the 3D position information, and the processed images are displayed on the four-sided display device.

In general, when shooting large events such as sports competitions and music concerts for live public screening, the cameras are placed so as not to interfere with the audience, usually from a different height (Fig. 8 left). Therefore, if the camera images are displayed without any processing to match the viewpoint of the audience, the objects are not shown where they should be. They might look either higher (like the objects are floating) or lower (sunk into the floor) than the ideal positions. Therefore, the positions of the objects (object images produced by the real-time object extraction) are transformed by viewpoint transform for Kirari! for Arena [34].

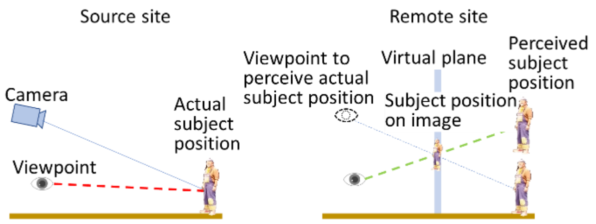

Fig. 8 – Object position perceived by a viewpoint different from a camera

The parameters of the transform are acquired in advance from the physical positions of the cameras and their calibration with the performing stage.

The position of the object reconstructed with the 3D position information was more stable than previous studies [35][36] using background subtraction and cropped input image.

5. EVALUATION AND DISCUSSION

Kirari! for Arena was demonstrated at an exhibition in December 2018. There were more than 200 exhibits and the total guest count was more than 14,000 people.

Commercially available PCs (specifications varied up to two 22-core 2.2-GHz CPUs, 32 GB RAM, GPU) were used to construct the Kirari! for Arena system. The system could process 4K (3840 x 2160 pixels) images with a frame rate of 59.94 fps.

Evaluating experience is quite a problematic task, but audiences of such experiences provide responses that can be used to estimate an objective evaluation of the system. A total of 98 relevant responses interviewed by the exhibitors were collected through the exhibition, and 20 of them were useful for evaluating the system. First, the responses were sorted into positive and negative stances. Then, they were further categorized into smaller categories for each stance. Table 1 shows the categorization results.

As these are free comment responses, the 5 categorized as "Realistic" does not necessarily mean the remaining 15 did not evaluate the system as "Realistic." On the contrary, the fact that at least 5 people out of 20 actively expressed their impression that the display was "Realistic" means the system gave realistic sensations to a fair number of people. Similarly, the positive responses (16) were four times as many as the negative responses (4). This shows the efficacy of the system. There are 2 responses categorized as "Not stereoscopic enough." It is natural to have a certain number of such responses because the display is 2D in principle. Providing a stronger sense of depth would be future work.

Table 1 – Audience responses

Stance	Category	Quantity	Subtotal	Total
Positive	Useful	10	16	20
	Realistic	5		
	Low latency	1		
Negative	Not stereoscopic enough	2	4	
	Has latency	2		

Table 2 shows some examples of the responses categorized as "Useful" and "Realistic." They are shown because these categories have a wide variety of responses. The "Realistic" responses show that the performer looked as if he was really in the device. Such realistic sensations were stimulated because the image from the source site was effectively fused into the surroundings in the display device at the remote site by combining real-time image extraction and depth expression. These responses could confirm that the atmosphere was reconstructed by the system.

Table 2 – Examples of responses

Category	Responses
Useful	It would be great if we have this in our home.
Useful	It would be good for streaming of live concert.
Useful	I want to watch Sumo on this.
Realistic	Looks as if he's really there.
Realistic	It's too real it's scary.

5.1 Conformance to ITU-T H.430 series (ILE) of standards

Kirari! for Arena conforms to the ITU-T standards by fulfilling all of the mandatory requirements.

Table 3 summarizes the requirements specified in ITU-T H.430.1 and the conformance status of Kirari! for Arena. The numbers in the column "Kirari! for Arena" indicates the subclauses in this article describing the functions and evaluation results that fulfill the requirements. Although wave field synthesis technology with a speaker array was not used for the exhibition, Kirari! for Arena can also be constructed with it to reconstruct the direction of sound.

Table 3 – List of ILE requirements and conformance status of Kirari! for Arena

No.	**Requirements**	**Manda-tory**	**Kirari! for Arena**
1	Life-size display		
2	Direction of sound	✔	✔
3	Atmosphere reconstruction		✔ (5.1)
4	Spatial reconstruction	✔	✔ (4.6)
5	Synchronous representation	✔	✔ (4.5)
6	Augmented information		
7	Real-time object extraction	✔	✔ (4.3)
8	Spatial object information		✔ (4.4)
9	Synchronous transmission	✔	✔ (4.5)
10	Data storage		
11	Reconstruction processing	✔	✔ (4.6)
12	Auditory lateralization		
13	Video stitching		

5.2 Discussion on further study

Although Kirari! for Arena conforms to ITU-T H.430, it has several aspects to be improved because it was an experimental development. First, it was not life-size. Second, auditory lateralization should be integrated for a stronger sense of reality. Past Kirari! events proved their feasibility and effectiveness [20]. Augmented information may also enhance the audience experience. These issues would require further research and development to determine the appropriate methods for real-time transport and processing. Data storage and video stitching may also be integrated wherever such functions are necessary.

On the standards side, it may be worth considering standardizing issues such as reference models of presentation environments. Further collaborations with immersive service-related standardization bodies, such as ISO/IEC JTC1, especially MPEG, 3GPP, and DVB, also need to be continued.

6. CONCLUSION

ILE services can bring people unprecedented experiences. This article introduced Kirari! for Arena as one implementation of ILE, which reconstructs the atmosphere of sports or entertainment events using a four-sided display and depth expression to provide audiences anywhere in the world with a highly realistic sensation. The display device has a relatively simple configuration that consists of four displays and four half-mirrors, compared with existing presentation devices containing many projectors and special optical devices. Kirari! for Arena is realized by combining functions for real-time image extraction, object tracking, synchronous media transport, and depth expression. Item-wise evaluation shows that it conforms to ITU-T H.430. Further standardization work on ILE and collaboration with related standardization bodies are required.

REFERENCES

[1] J. Nagao, H. Miyashita, T. Sano, K. Hasegawa, T. Isaka (2019), "Kirari! for Arena : Real-Time Remote Reproduction of 4-Directional Views with Depth Perception," Journal of the Imaging Society of Japan, Vol. 58, Issue 3.

[2] Recommendation ITU-T H.430.1 (2018), Requirements for immersive live experience (ILE) services.

[3] Recommendation ITU-T H.430.2 (2018), Architectural framework for immersive live experience (ILE) services.

[4] Recommendation ITU-T H.430.3 (2018), Service scenario of immersive live experience (ILE).

[5] Recommendation ITU-T H.430.4 (2019), Service configuration, media transport protocols and signalling information of MMT for Immersive Live Experience (ILE) systems.

[6] Dolby 3D. 2018. Retrieved from http://www.dolby.com/us/en/technologies/dolby-3d.html.

[7] RealD. 2019. Retrieved from http://reald.com/.

[8] XPANDVISION™ 3D Cinema. 2019. Retrieved from http://xpandvision.com/cinema/.

[9] 4DX. 2019. Retrieved from http://www.cj4dx.com/.

[10] Media Mation MX4D. 2019. Retrieved from http://mx-4d.com/.

[11] Y. Mizushina, W. Fujiwara, T. Sudou, C. L. Fernando, K. Minamizawa, S. Tachi (2015), "Interactive instant replay: sharing sports experience using 360-degrees spherical images and haptic sensation based on the coupled body motion," AH '15.

[12] N. Inamoto, H. Saito (2015), "Free viewpoint video synthesis and presentation from multiple sporting videos," ICME, 2005.

[13] B-C. Lee, J. Lee, J. Cha, C. Seo, J. Ryu (2005), "Immersive Live Sports Experience with Vibrotactile Sensation," INTERACT 2005, LNCS 3585, pp. 1042-1045.

[14] M-S. Iekura, H. Hayakawa, K. Onoda, Y. Kamiyama, K. Minamizawa, M. Inami (2015), "SMASH: synchronization media of athletes and spectator through haptic," SA'15 Symposium on MGIA.

[15] H. Kim, H. Yamamoto, N. Koizumi, S. Maekawa, and T. Naemura (2015), "HoVerTable: Combining Dual-sided Vertical Mid-air Images with a Horizontal Tabletop Display," The Transactions of Human Interface Society, Vol. 17, No .3.

[16] H. Kim, S. Nagao, S. Maekawa, and T. Naemura (2014), "MRsionCase: A Glasses-free Mixed Reality Showcase for Surrounding Multiple Viewers," IEEE Transactions on Media Technology and Applications, Vol. 2, No. 3, pp. 200–208.

[17] X. Luo, J. Lawrence, and S. M. Seitz (2017), "Pepper's Cone: An Inexpensive Do-It-Yourself 3D DisplayA," UIST 2017, pp. 623-633.

[18] O. Bimber, B. Fröhlich, D. Schmalstieg, and L. Miguel Encarnação (2001), "The Virtual Showcase," IEEE Computer Graphics and Applications, pp. 48-55, Nov/Dec.

[19] H. Takada (2017), "Immersive Telepresence Technology "Kirari!"," Journal of the Imaging Society of Japan, pp. 366-373, Vol. 56, No. 4.

[20] Nippon Telegraph and Telephone Corporation News. 2016. Retrieved from http://www.ntt.co.jp/news2016/1604/160419a.html.

[21] Nippon Telegraph and Telephone Corporation News. 2017. Retrieved from http://www.ntt.co.jp/news2017/1704/170417c.html.

[22] Nippon Telegraph and Telephone Corporation News. 2018. Retrieved from http://www.ntt.co.jp/news2018/1803/180313b.html.

[23] Environmental Planning Laboratory. 2018. Retrieved from https://www.epl.co.jp/futsu-no-sukima/.

[24] J. Nagao, M. Yamaguchi, H. Nagata, and K. Hidaka (2018), "Robust Object Extraction from Similar Background Color using RGB-IR Camera," Proceedings of the 2018 ITE Annual Convention Vol. 2018, 13C-3.

[25] R. Pei, Z. Geng, and Z. X. Zhang (2016), "Subpixel Multiplexing Method for 3D Lenticular Display," Journal of Display Technology, Vol. 12, No. 10, October.

[26] H. Kakinuma, H. Miyashita, Y. Tonomura, H. Nagata, and K. Hidaka (2017), "Real-time Image Segmentation using Machine Learning Techniques for 4K Video," Proceedings of the ITE Winter Annual Convention, Vol. 2017, 15B-2.

[27] H. Miyashita, K. Takeuchi, M. Yamaguchi, H. Nagata, and A. Ono (2016), "Fast and Accurate Image Segmentation Technique using Multiple Sensors," IEICE Technical Report, Vol. 116, No. 73 (MVE2016 1-5), pp. 17–22.

[28] H. Miyashita, K. Takeuchi, H. Nagata, and A. Ono (2017), "Fast Image Segmentation for 4K Real-time Video Streaming," IEICE Technical Report, Vol. 117, No. 73 (MVE2017 1-13), pp. 189–190.

[29] K. Kidono, A. Watanabe, T. Naito, and J. Miura (2011), "Pedestrian Recognition Using High-definition LIDER," Journal of the Robotics Society of Japan, Vol. 29, No. 10, pp. 963–970.

[30] Y. Ishii, T. Tokunaga, Y.Tonomura, and K. Hidaka (2017), "Kirari! Tracker: Human Detection and Tracking System by Using LiDAR and Deep Learning Engine," ITE Winter Annual Convention, ITE, 15B-3.

[31] Y. Tonomura (2018), "Kirari!: An ultra-realistic viewing experience to the world," Proceeding of ITE Annual Convention.

[32] H. Imanaka, Y. Tonomura, and K. Tanaka (2018), "New style of sport watching by Immersive Live Experience (ILE) and its standardization status," Journal of the ITU Association of Japan, Vol. 48, No. 2, pp. 7–10.

[33] M. Ono, K. Namba, T. Yamaguchi, and A. Ono (2018), "Development of the law-latency HEVC/MMT codec system," Proceeding of ITE Annual Convention, 22C-3.

[34] T. Isaka, M. Makiguchi, and H. Takada (2018), "Viewpoint conversion of 2D floating image using subject space position and its application to real-time system," Proceedings of the 2018 the Institute of Image Information and Television Engineers Winter Annual Convention, 24D-3.

[35] W. K. Hyung and Y. S. Sung (2002), "Enhanced Lane: Interactive Image Segmentation by Incremental Path Map Construction," Graphical Models, Vol. 64, No. 5, pp. 282–303.

[36] W. K. Hyung and Y. S. Sung (2002), "Tour into the Video: Image based Navigation Scheme for Video Sequence of Dynamic Scenes," Proc. of VRST 2002, pp. 73–80.