

GAYAIT: an Interactive Video and Sound Art System handling a large number of Video Clips and its Applications

Shinji Mizuno ¹⁾ (Member) Ryosuke Hirano ²⁾ Yukihiro Tsutsumi ³⁾

1) Faculty of Information Science, Aichi Institute of Technology

2) Toko Engineering Inc.

3) Office Crescendo Inc.

s_mizuno@aitech.ac.jp

Abstract

In this paper, we propose "GAYAIT": an interactive video and sound art system that handles a large number of video clips. The system has about 25~80 virtual displays in a 3DCG space, and each of them plays video clips simultaneously. In each video clip, a person is talking a message or singing a song. The user can pick up one or two of video clips with gestures. When the user chooses video clips, the size and the sound of the chosen video clips are magnified. As a result, the selected video clips come into focus, and the message or the song of the video clips can be understood clearly. The user can change video clips selection interactively, and could understand the personality of the people in the video. The user can also feel an instant change between a mass and an individual in our system.

1. Introduction

Media arts and digital contents using computer technologies such as image processing, sound processing, and robotics are going popular recently. Interactive media arts that can respond a user's motion have been used for advertising and communication, and they have especially attracted much attention [1][2][3]. A digital content that uses holographic to realize more realistic interactive experience in a 3DCG space has been introduced [4].

In this paper, we propose "GAYAIT": an interactive media system that handles a large number of video clips [5][6][7]. The system has about 25~80 virtual displays in a 3DCG space, and each of them plays video clips simultaneously. In each video clip, a person is talking some messages, singing a song, and so on. The user can pick up one or two of video clips with gestures. When the user chooses one of video clips, the size and the sound of the designated video clip is magnified. As a result, the selected video clip comes into focus, and the message or the song of the video clip can be understood clearly. The user can change video clips selection interactively, and could understand the personality of the people in the video. The user can also feel an instant change between a mass and an individual in our

system. The user can select two video clips simultaneously.

Each video clip plays its sound at each position, thus the user can feel the 3D space from mixed stereophonic sounds, and can recognize the change and the movement of sound sources when the user changes video clips selection in succession.

The contents synthesized by the GAYAIT system would change according to the video clips on each virtual display. When using video clips in each of which a person is talking a message and these clips are played simultaneously, the system represents a mass and creates a sound of a cloud, thus it is

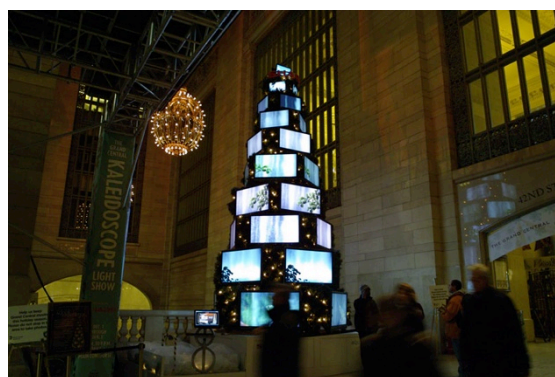


Figure 1. AQUOS Experience in the Grand Central 2008.

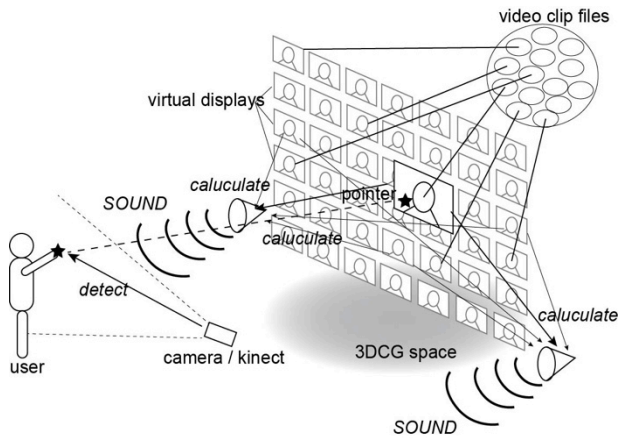


Figure 2. The over view of the "GAYAIT" system.

difficult to catch each message. When using video clips in each of which a person is singing a song, the user can control and enjoy a vocal solo and a chorus interactively. The system can also synthesize a harmony and a round of a song.

There are some video arts that use a lot of video clips. Nam June Paik is famous for his video art that uses a lot of displays and plays video clips on them [8]. A Christmas tree with a lot of liquid crystal displays was exhibited at Grand Central Terminal, New York in 2008 as shown in Figure 1. It is possible to mix and play several music video clips on Google Chrome Music Mixer [9]. If using a PC, it is possible to play many video clips simultaneously by running media players simultaneously. However the number of video clips handling simultaneously is not so many, and it is difficult to control the way of playing them interactively.

GAYAIT is a novel media art, which uses a lot of video clips simultaneously and can control the way of playing them interactively. The user can experience and enjoy an instant change between a mass and an individual. In this paper, we introduce the method and the applications of GAYAIT.

2. System Overview

Figure 2 shows the overview of the proposed system.

The system is composed of a large screen, a PC, and a video camera. The system has a virtual stage with about 25~80 virtual displays in a 3DCG space, and shows them on a screen. Each of the virtual displays is connected to its own video clip file and plays the video and the sound of each video clip independently. The sound of each video clip is monophonic and played at each position of the virtual display, and the synthesized sound is stereophonic which consider the position of each video clip. When the content of each video clips is a person talking a message and they are played simultaneously, the system represents a mass and creates a sound of a cloud as shown in Figure 3(a). It is difficult to catch each message in this case.

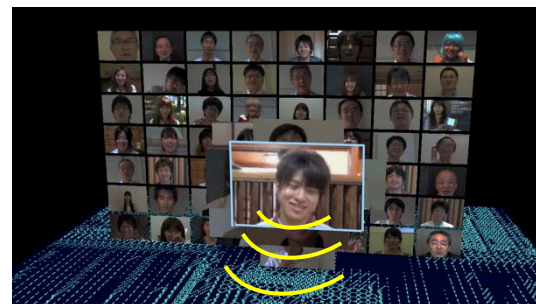
The user stands in front of the screen while holding a color



(a) About 25~80 virtual displays are put in a 3DCG space, and each of them plays video clips simultaneously. The system would synthesize sounds of a crowd



(b) A user stands in front of a screen and points to one of video clips.



(c) The size and the sound of the pointed video clip are magnified. The user can control the way of playing video clips interactively.



(d) The user can also choose two video clips and magnify them.

Figure 3. Control the way of playing video clips by a gesture.

ball, moves a pointer in a 3DCG space with their hands, and points to one of virtual displays as shown in Figure 3(b). The system monitors the user with a video camera, and detects the color ball from video images, and calculates the position of a pointer in the 3DCG space. When one virtual display is chosen by the pointer, the virtual display comes closer and is enlarged in a 3DCG space, and the video image on it increases in size as



Figure 4. Further emphasizing the chosen video clip by moving a user's hand forward.

shown in Figure 3(c). In addition, the sound of the chosen video clip increases in volume, and sounds of other video clips decrease in volume.

As a result, the video clip on the chosen virtual display would be emphasized, and the message of the video clip would be clearly understood by the user. This allows the user to understand the personality of the people in the video clip. When the user moves the pointer, the system emphasizes video clips one after another, and the user can control the way of playing video clips interactively. The user could experience an instant change between individual and mass in our system.

In this paper, we introduce some new gestures to control the GAYAIT system. After choosing video clips, the user can further emphasize the video clip by moving their hands forward (Figure 4). Using two balls, the user can move two pointers and can select two video clips simultaneously (Figure 3(d)), The user can also magnify and rotate the chosen video clip (Figure 5). The user can change two operations by changing an operation mode or a color of a ball. When the user uses balls of the same color, the user can operate two pointers. When the user uses balls of different colors, one ball is used for choosing a video clip and the other one is used for further magnifying and rotating the chosen video clip.

The user can also recognize the change and the movement of sound sources when the user moves a pointer and changes video clips selection one after another.

3. Methods for GAYAIT

The GAYAIT system uses computer graphics technologies, image processing technologies, and sound processing technologies, and it is implemented on a PC in C++.

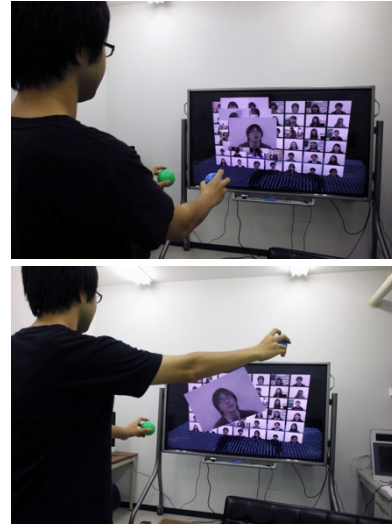


Figure 5: Further magnifying and rotating the chosen video clip using two balls.

3.1 Constructing a virtual stage

The virtual stage of GAYAIT is constructed using 3DCG technology. Each virtual display in the stage is created by a plane rectangle and is put in a 3DCG space. Each virtual display is connected to its own video clip.

Each video clip is separated into a video file and a sound file beforehand. The video file format is Quick Time and the sound file format is WAV. The resolution and the frame rate of the video file would be decided depend on the spec of a PC.

The system processes the video files and the sound files separately to play video clips. A frame image of each video clip is taken out from its video file one after another, the size of the image is modified to fit the virtual display, and the image is mapped to the connected virtual display using texture mapping technology. This process is done for all virtual displays automatically at each time the system refreshes a frame.

3.2 Sound processing

Sound of each video clips is monophonic. Virtual sound sources are put at a position of each virtual display, and each virtual sound source plays the sound of its video clip. The GAYAIT system calculates the left volume and the right volume of each sound at a viewpoint based on the position of each virtual sound source, and mixes them. As a result, the system can synthesize stereophonic sounds by taking the position of each virtual sound source into consideration.

Each sound of video clips has a parameter of volume, and they would be used to control a volume of the sound when the user selects a video clip.

3.3 Detecting gestures of a user

A user stands in front of the screen with holding a color ball and makes some gestures to move a pointer in a 3DCG space.

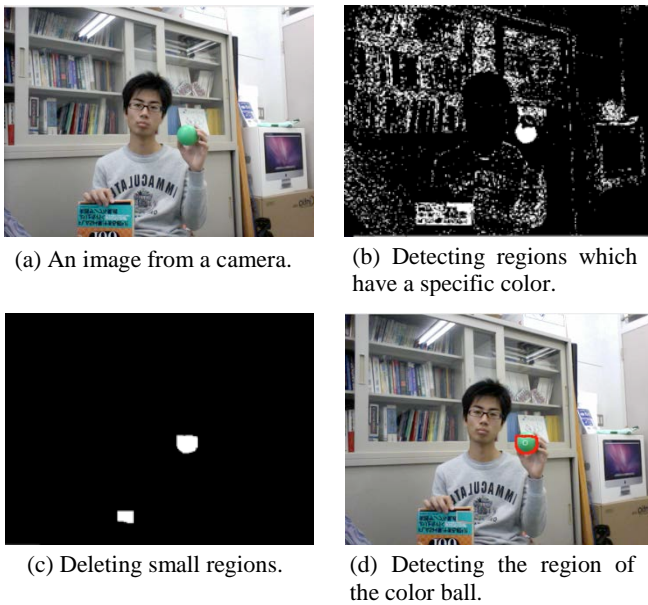


Figure 6. The process of detecting the position of the color ball.

The system monitors the user with a video camera and detects the position of the color ball to decide the position of the pointer.

Figure 6 shows the process of detecting the position of a color ball. First, the system converts an image taken with the video camera to the HSV color space, and detects regions which have a specific hue and a specific brightness of the color ball as shown in Figure 6(b). Then, small regions are deleted by erosion and dilation processes as shown in Figure 6(c). Finally, the biggest region is decided as the region of the color ball as shown in Figure 6(d). The center position of the detected region is projected to a 3DCG space and the position of the pointer is decided.

It is also possible to use two balls, a mouse, or a Kinect to decide the position of the pointer. When using two balls, the relative position of two balls is also used. The suitable way of detecting gestures of a user would change according to the situation of using the GAYAIT system.

3.4 Control of video images and sounds

The user moves a pointer with a color ball, and the position of the pointer would change the way of playing the video clips.

If the system does not detect a color ball and the position of the pointer is out of range, the GAYAIT system plays all video clips with same sizes and same volumes. When the system detects a color ball and decides the position of the pointer in a 3DCG space, the nearest virtual display from the pointer is considered as being chosen.

The chosen virtual display comes closer to the viewpoint and is enlarged in a 3DCG space, and the video image on it increases in size as shown in Figure 7. Virtual displays around

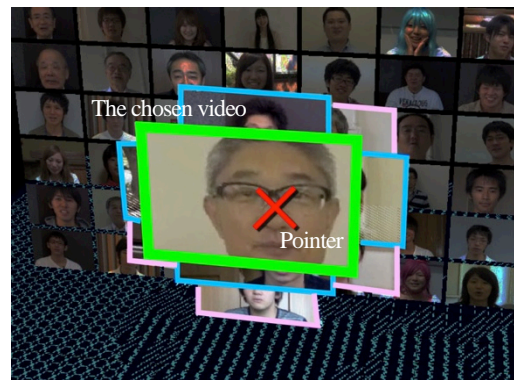


Figure 7. Choosing a video clip with a pointer.

the chosen virtual display also increase in size based on the distance between the pointer and each virtual display.

The sound source of the chosen video clip also comes closer to the viewpoint with the virtual display, and as a result the synthesized sound at the viewpoint increases in sound. It is not enough to emphasize the video clip, thus the parameters are changed to increase the sound of the chosen video clip and to decrease other video clips in volume.

As a result, the video clip on the chosen virtual display would be emphasized and the user can catch the sound of the chosen video clip. When the user moves the pointer, the system emphasizes video clips one after another and the user can control the playback of video the clips interactively. The user can also recognize the movement of sound sources.

The system calculates the change in the size of a ball region. They are used to change the size and the volume of the chosen video clip. When the user moves a ball forward after choosing a video clip, the size of a ball region will increase and the size and the volume of the chosen video clip is magnified further.

Using two balls, the user can move two pointers and can select two video clips simultaneously. It is possible to magnify or rotate the chosen video clip interactively. They are realized by using the relative position of one ball from another ball.

4. Experiments

The GAYAIT system is implemented on a PC in C++. OpenGL is used for creating the virtual stage, OpenCV is used for detecting gestures of the user, and OpenAL [10] is used for sound control processes. The spec of the PC is as follows.

- OS: Mac OSX 10.6.8
- CPU: Intel Core i7 2.8GHz
- GPU: NVIDIA GeForce 330M
- Memory: 8GB

The resolution of each video file is 160 x 120(pixels) or 320 x 240(pixels), and the frame rate is 10(fps). Each sound file is 24,000(kHz), 16(bit), monophonic.



Figure 8. Replacing video clips.



Figure 9. Movement of the selection of video clips.

4.1 Using message video clips

Video clips in each of which a person is talking a message are used in the GAYAIT system. The number of virtual displays is 56, and the system prepares about 90 video clips. The lengths of video clips are varied. The shortest video clip is 15(s) and the longest video clip is 420(s). The system plays those video clips on virtual displays repeatedly and independently. It is possible to add video clips at any time. 56 video clips are selected at random, replaced periodically, and connected to each virtual display as shown in Figure 8. They are played simultaneously

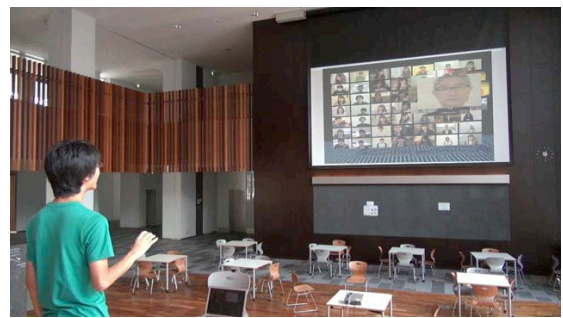


Figure 10. An exhibition of the GAYAIT system with a large screen at a campus festival.

on each virtual display.

If no video clip is chosen, the system plays all video clips with same sizes and same volumes. A stereophonic sound for each video clips is synthesized according to the position of it and mixes them, thus the system synthesized a stereophonic sound like a noise of a crowd in a hall, and it is difficult to understand each message.

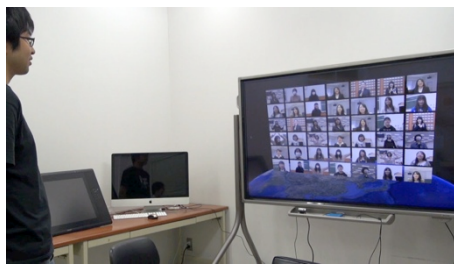
When the user chooses one of the video clips, the size and the volume of the chosen video clip is magnified, and the volumes of other video clips are reduced. As a result, the selected video clip comes into focus, and the message of the video clip can be understood clearly. The size and the volume of video clips around the chosen video clip are also magnified a little, and the user can catch the messages a little.

When the user move a pointer with a color ball, the system emphasizes video clips one after another, and the user can control the playback of the video clips as shown in Figure 9. A sound of each video clips is synthesized according to the position of it, and it is possible to recognize the movement of the position of the sound source.

We exhibited the GAYAIT system using a large screen at a campus festival as shown in Figure 10, and about 50 people experienced it. Some people gave their impressions that the feeling of operating the GAYAIT system was just like looking at a crowd of people through a magnifying glass. Many people enjoyed emphasizing personalities from a crowd.

4.2 Using music video clips

Video clips in each of which a person is singing a song or playing a musical instrument are used in the GAYAIT system as shown in Figure 11. All people in the video clips sing a same song independently with a same accompaniment. The number of virtual displays is 42, and they play their music video clips simultaneously. All video clips are previously adjusted to the same timing and the same length.



(a) Synthesizing a stereophonic chorus sound.



(b) Synthesizing a vocal solo sound.



(c) Synthesizing a twin vocal sound.

Figure 11. Using music video clips on GAYAIT.

The system synthesizes a stereophonic sound for each video clips according to the position. When no music video clip is chosen, the system mixes all sounds with same volumes, thus the system synthesizes a stereophonic sound like a chorus in a concert hall. On the other hand, when the user chooses one of the video clips, the sound of the chosen music video clip is magnified, and the system synthesizes a vocal solo sound. When choosing two video clips, the system synthesizes a twin vocal sound. It is possible to change a chorus, a vocal solo, and a twin vocal sound interactively at any time. When the user moves a pointer, it is possible to synthesize a sound that a song is sung by many singers in turn, and the user can control the way and the order of singers interactively. There is no other performance like this.

When using singing video clips that include different tones such as soprano and bass, the system can synthesize a harmony sound. If the timing of each song is delayed properly, the system can synthesize a round of a song. When the system starts a round of a song from the right side to left side, the user can hear the song is moving from the right side to the left side.

4.3 Applications

An instant change between each personality and a mass could attract attention of many people, thus our system would be useful for advertisement, entertainment, and education.



Figure 12. Using the GAYAIT system in Nagoya City Science Museum.

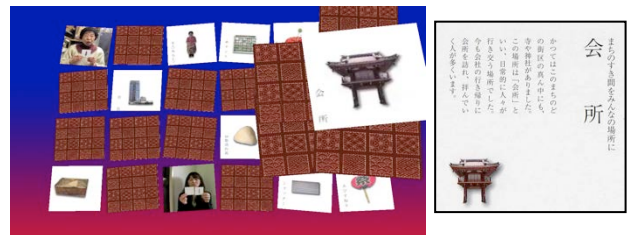


Figure 13. The GAYAIT system for the Choja-machi karuta cards.

The GAYAIT system was used at a wedding party to present celebration messages from friends. The bride and groom could get celebration from each person directly, and could feel an atmosphere of celebration from a mass. Many guests gave their impressions of the GAYAIT system that playing many celebration message videos simultaneously with GAYAIT is very unique and much more attractive than playing videos one after another. The GAYAIT system was very popular with the host and the guests of the wedding party and a certain company has a plan to use GAYAIT continuously as one of stage effects in a wedding party.

The GAYAIT system has been used at Nagoya City Science Museum. Nagoya City Museum was rebuilt in 2011, and has a lot of message video clips of curators and visitors of the museum in which people are talking about memories of the old museum and expectations of the new museum. The GAYAIT system is used in the exhibition room of the museum where the visitors can operate the system with a color ball or a touch panel, and watch messages. Figure 12 shows the system in the museum.

The GAYAIT system is also used to promote Choja-machi: a shopping area in Nagoya, Japan. Choja-machi is an old shopping area and the people are promoting Choja-machi with a "karuta" card game [11][12]. Karuta is a set of traditional Japanese playing cards, and the people of Choja-machi made original karuta cards. The Choja-machi karuta cards would be used not only for playing a karuta cards game, but also creating a digital content with the GAYAIT system. Images of the karuta cards and video clips of promoting messages about Choja-machi

are used in the GAYAIT system. Figure 13 shows the GAYAIT system for the Choja-machi karuta cards. In a Choja-machi festival, many participants sang an original song independently and organized a virtual chorus on GAYAIT.

In this paper, the system increased the chosen video clip in volume, and the user could catch the message. If the system does not change the volume of the chosen video clip, the GAYAIT system might be useful to experiment the cocktail party effect [13].

In our experiment, we confirmed that it is possible to handle at least 80 video clips on virtual displays simultaneously. It depends on a PC spec used for the GAYAIT system.

5. Conclusion

In this paper, we proposed a novel interactive media art and digital content: GAYAIT. This system enables us to handle a large number of video clips simultaneously and to control the way of playback them interactively by using computer graphics, image processing, and sound processing technologies. The GAYAIT system can create very unique contents with videos and sounds, and they can attract many people's attentions. The GAYAIT system has ability to use for advertisement, entertainment, education, and so on.

We would improve and extend the system by detecting more complex gestures. Exploring other interactive media contents using the GAYAIT system, and using live video clips are also future works.

Acknowledgement

The authors would like to thank Dr. Hiroaki Sawano of Aichi Institute of Technology, Dr. Katsuhiko Mouri of Nagoya City Science Museum, the members of Shinji Mizuno laboratory and Sawano laboratory of Aichi Institute of Technology, and the members of Nagoya Nursing and Midwifery College for help of this work, valuable discussion, and making video clips. This work is supported in part by Ministry of Education, Science, Sports and Culture of Japan, Grant-in-Aid for Scientific Research (C) (23500139).

References

- [1] N. Ohshima, K. Okazawa, H. Honda, M. Okada, TableTalkPlus : An artifact for promoting mutuality and social bonding among dialogue participants, *J. of Human Interface Society*, Vol. 11, No. 1, pp. 105-114, 2009.
- [2] R. Nakamori, T. Aoki, I. Sii, Neighbors' Shadow: Moderate Communication with Familiar Strangers, 7th International Conference on Advances in Computer Entertainment Technology (ACE 2010), pp. 17-19, 2010.
- [3] M. R. Mine, D. Barnard, B. Yang, D. Baker, Tavola: Holographic User Experience, SIGGRAPH 2012 Emerging Technology, DVD

Proceedings, 2012.

- [4] Y. Fei, D. Kryze, A. Melle, Thermal Interactive Media, SIGGRAPH 2011 Emerging Technology, DVD Proceedings, 2011.
- [5] S. Mizuno, R. Hirano, Y. Tsutsumi, GAYAIT: an Interactive Media System with Individual and Whole characters, Proc. of Interaction 2012 Demonstration, IEXB-13, 2012.
- [6] S. Mizuno, R. Hirano, Y. Tsutsumi, GAYAIT: an Interactive Media System using a large number of Video Clips and its Applications, Proc. of NICOGRAPH International 2012, pp. 56-61, 2012.
- [7] S. Mizuno, R. Hirano, Y. Tsutsumi, An Interactive Media System "GAYAIT" and its Applications, Proc. of DICOMO 2012, pp. 2438-2443, 2012.
- [8] Num June Paik Studios, <http://www.paikstudios.com/>.
- [9] Google Chrome Music Mixer, <http://www.google.co.jp/landing/motto/>.
- [10] OpenAL: <http://connect.creativelabs.com/openal/>.
- [11] hanare: <http://kin2hanare.jp/>.
- [12] H. Sawano, S. Mizuno, A. Hara, Digital Contents with Carta for Promotion of the Town, Proc. of DICOMO 2012, pp. 2425-2428, 2012.
- [13] E. C. Cherry, Some Experiments on the Recognition of Speech, with One and with Two Ears, *J. Acoust. Soc. Am.* Vol 25, Issue 5, pp. 975-979, 1953.



Shinji Mizuno is currently an Associate Professor at Faculty of Information Science, Aichi Institute of Technology. He received the BS, MS, and Ph.D degrees from Nagoya University, in 1993, 1995 and 1999, respectively. He was a Research Associate at Toyohashi University of Technology from 1999 to 2009. His research interests include computer graphics, image processing, virtual reality, and interactive arts. He is a member of the Society for Art and Science, IPSJ, IIEEJ, and VRSJ.



Ryosuke Hirano received the BS degree from Aichi Institute of Technology in 2012. He joined Toko Engineering Inc. in 2012. His research interests include computer graphics and interactive arts.



Yukihiro Tsutsumi is a Japanese film director. He is a Visiting Professor at Aichi Institute of Technology from 2010.