

A Study on User Interface Based on Hand Gesture Recognition for Live Video Mixing System

Joohun Lee

Dept. of New Media Arts, Dong-Ah Institute of Media and Arts

Haehyun Jung

The Graduate School of Advanced Imaging Science, Chung-ang University

Jaejoong Lee

Mogence Lab Co. Ltd.

Abstract

Development in digital technology and video technology significantly contributes to increases of the utilization of video sources in a variety of performances. At the same time, the necessity to promptly control and change the video gets higher, especially in live performance. In these circumstances, if the performer could control the video depending on actions without need of direct contact to the device, live video performance will be much faster and more convenient. In other cases, DJ who carries out mixing music in club and DJ party would like to take a role of visual performer as well, or control background video in realtime corresponding to the music being played. In these cases, gesture recognition based non-contact interface could be advantageous over a traditional way of touching or contacting the device directly. To this end, the starting point of our study is to devise the user interface that enables a live video production making use of hand gesture recognition device. However, the hand gesture recognition based user interface might cause unintended wrong recognition due to similarity of device-controlling hand gesture and normal hand gesture. To prevent this, we introduce the gesture-recognition mode switcher using pre-gesture.

1. Introduction

Due to development of computers, natural way of manipulating computerized environment has been highlighted in terms of an interface method through which people is allowed to control a variety devices. As a consequence, there have been considerable efforts made to study on natural user interface(NUI). Unlike the traditional way of interface such as command line interface(CLI) or graphic user interface (GUI) which are based on the use of physical devices including mouse, keyboard, button and slide-bar, the NUI is the method that enables users to control their device not based on direct contact to the device but based on using non-contact recognition device like sensor. The gesture recognition user interface, which recognizes and identifies users gesture, has advantages that it is intuitive and simple to use. There are other studies that propose the method to analyze and use users physical body structure data captured by camera [1]. Another study proposes the method to capture the articulation of human hand using infrared (IR) camera [2]. In particular, one of widely used hand tracking device recently is the 'Leap Motion'. Working with IR cameras, the Leap Motion is capable of tracking and recognizing hand movement at a speed of 200 fps or higher. Since the Leap Motion achieves fast capturing an articulation features data, it is known to be a suitable interface for game environment in which immediate feedback is essential [3]. In reality, The Leap Motion has shown high level of user satisfaction when applied to the game environment [4].

In case that music and video are supposed to be provided simultaneously, the communication between DJ who controls music and VJ who controls video is critically

important in order to maximize visual and auditory effects. If DJ controls video as well as music, DJ would produce more effective performance [5].

On the other hand, development of digital technology and video technology contributes to performance production equipped with a variety of sensory effects satisfying visual and auditory sense, with help of multimedia technique. Recently, leaping beyond the limitations of traditional stage art depending on only mechanical devices, real-time performances are being actively performed effectively making use of video and audio which are representative components of multimedia. Due to the widespread popularity of performance providing music and video simultaneously such as DJing and VJing, the communication between DJ and VJ becomes much more important and the harmony between two has significant influence on satisfaction of audience and quality of performance. In this context, in order to allow DJ to control video in addition to music, hand gesture recognition based realtime video control interface is proposed, which is a type of non-contact control method [6]. The proposed method enables DJ to control music and video simultaneously. With hand-tracking recognition device, DJ who controls music can carry out the performance and present video at the same time, which means one person can take roles of DJ and VJ. This style of performance environment is likely to make performance more harmonized owing to better unification of music, video and performance.

However, the hand gesture recognition using the Leap Motion has problem regarding the recognition range of sensor. The recognition range is as illustrated in the Figure. All the hand movements taking place in the range are recognized as the controlling gesture, which often results to mistakes that hand gestures not intended to be interpreted as command are recognized as command as well. If the Leap Motion is applied to the realtime performance, it is highly likely to make a mistake when recognizing DJ's hand gesture because DJ tends to move hand a lot for performing purpose. If the Leap Motion recognizes the hand gesture incorrectly, it might result to wrong controlling and unintended change in video. To prevent this kind of mistake, pre-gesture is required, which allows to tell that the subsequent gesture will be a gesture for controlling video or not. Specifically speaking, pre-gesture can be a signal that activates or deactivates the recognition mode, so that the movements occurring under the normal mode (gesture recognition mode deactivated) can be effectively excluded from the recognition.

In this paper, we propose video controlling user interface using the Leap Motion for a person who controls music in a live performance. The proposed user interface enables the user to control video in non-contact way using six distinct hand gestures. In addition, we suggest the way to prevent the mistaken recognition such that the user's movement for performance is interpreted as controlling gesture and vice versa. To prevent wrong control due to mistaken recognition by confusing between the user's movement for performance and the user's movement for controlling, we designed the mechanism to spot the user's gesture based on pre-gesture. The proposed mechanism sets up the pre-gesture which notifies on/off prior to starting the user's controlling gesture. Even if the user makes action which is similar to the controlling gesture but which is not controlling one, the proposed system can effectively spot the user's controlling gesture and reduce the mistaken recognition.

2. Hand-gesture Recognition Interface using the Leap Motion

In order for DJ to control video in addition to music in live performance, prompt and efficient user interface is essential. Non-contact control is a type of the NUIs, which provides considerably effective and simple interface for a user [7]. In many cases, videos are popularly used in live performance, owing to the development of video technology and computer controlling technology. In particular, the use of high resolution projector

and large LED display device is attributed to the encouraging the mood of the performance. One of public entertainment cultures, VJing is found in club, party and event or is performed by a visual artist. VJing is an representative example of maximizing the visual effect by way of realtime video control [8, 9]. If a performer can promptly control the video in live performance, VJing technique can be more effectively employed. Furthermore, if DJ who controls music could control video simultaneously, music and video are performed harmoniously, which could lead to good unification of visual and auditory effects. To achieve this, the interface that can promptly control video in realtime performance is required.

In order for DJ to control video in addition to music in performance, prompt and efficient user interface is essential. One of the NUIs, Non-contact control provides considerably effective and simple interface for a user. Among those devices, the Leap Motion which is commercially developed and widely used, is suitable for this environment. The device recognizes human hand gesture through recognizing hand articulation features at speed of 200 fps. Therefore this device can be effective in realtime controlling. Our UI device is intended to be used to control video while doing DJing, which means our device needs to be non-contact type. Since our device's UI is required to be intuitive and user-friendly, the Leap Motion is suitable as recognition technique. Figure 1 illustrates the architecture of realtime video controlling system based on hand gesture recognition using the Leap Motion.

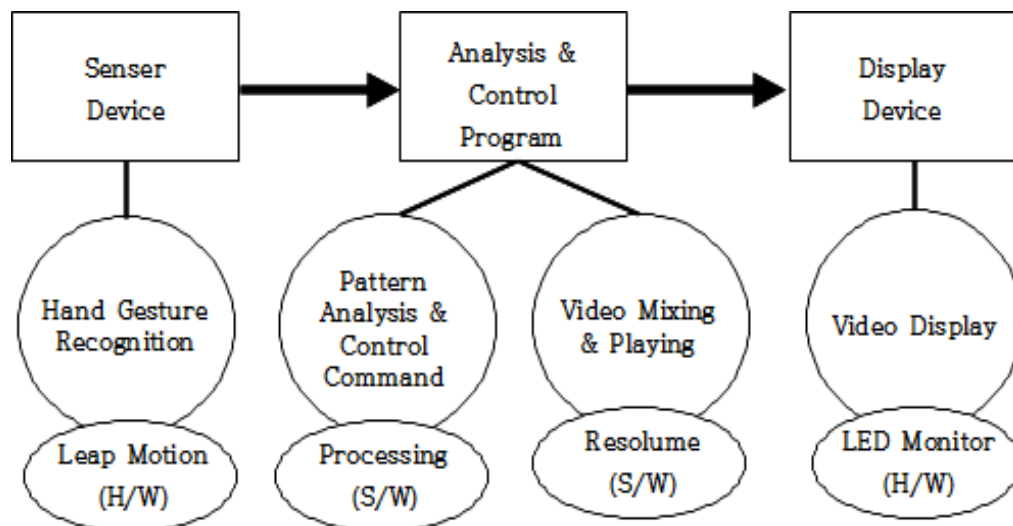


Figure 1. Block Diagram of the Proposed System

The system shown in Figure 1 has been implemented using SDK provided by the Leap Motion developing company. The system extracts x, y, and z-axis data for each finger and calculates the recognition pattern and acceleration of finger's movement. Based on the computed data, we assign the number (0-9) to ten fingers. Number 0 is assigned to the thumb of left hand, while number 5 is assigned to the thumb of right hand.

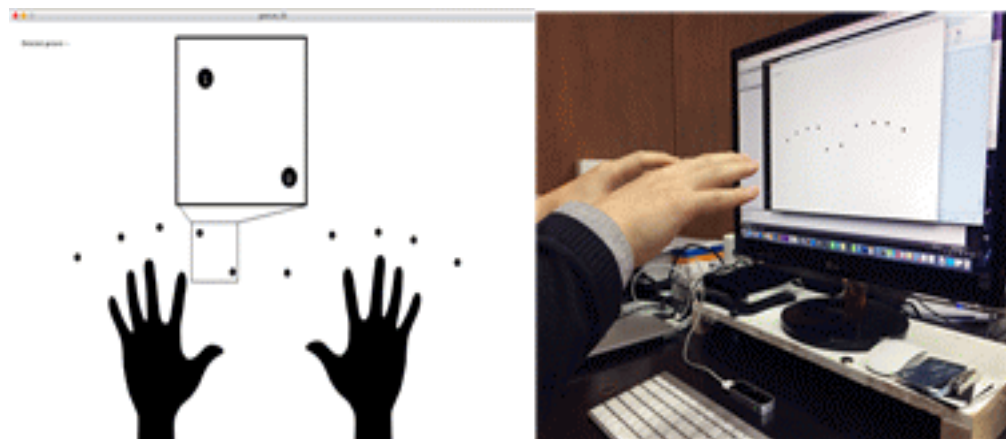


Figure 2. Extraction of Coordinates of Ten Fingers and Assigning Id

The user can generate hand gesture in a state that three fingers are open as shown in Figure 3. The hand gesture is extracted and recognized by the Leap Motion. The fingers movement is categorized into one of six different patterns which have spatial difference. The fingers movement extracted by the Leap Motion is compared to six patterns and categorized. Each pattern indicates distinct command. Six distinct finger movement patterns are illustrated in Figure 4. Once we integrate the six patterns into a video switcher software, we can successfully control video in conjunction with the Leap Motion as presented in Figure 5. If the recognition result is connected to video controlling software, the user, DJ is able to control video channel as he/she desires as well as doing DJing.

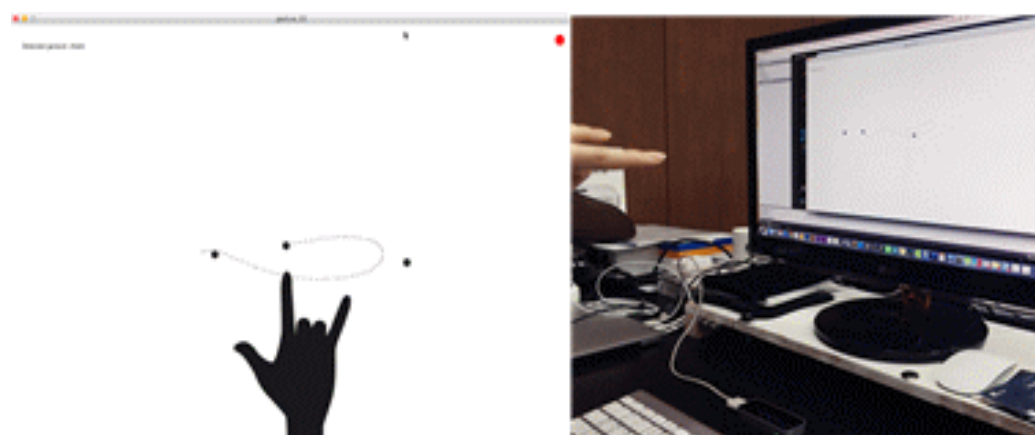


Figure 3. Gesture Recognition Process in the Proposed System

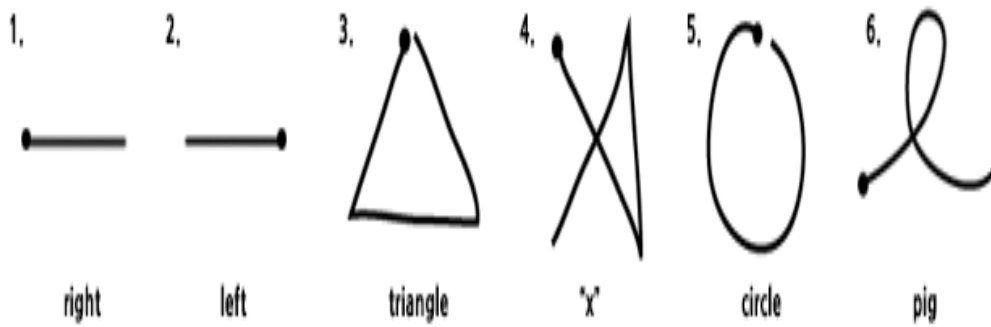


Figure 4. Six Command Recognition Patterns



Figure 5. Flow in the Controlling Process of Video Mixing Software through Hand Gesture Recognition by the Leap Motion

3. Improvement of Hand Gesture Interface using Pre-gesture

The Leap Motion has the recognition range as illustrated in Figure 6. As found in Figure 6, the Leap Motion has limitation that the user is required to move within the range because the recognition is carried out based on movement taken by IR camera. The Leap Motion also has drawback that it is likely to confuse between the movement for performance and the movement for control because DJ tends to move a lot and make big action while doing performance. Therefore it is necessary to design the device such that it can differentiate the two actions effectively. If this problem is ignored, wrong-action that the device mistakenly regards the user's performance action as controlling action and/or action-failure that the device mistakenly regards the user's controlling action as performance action might take place. To solve these problems, we introduce the way to explicitly express the beginning of controlling action using pre-gesture before starting the controlling gesture.

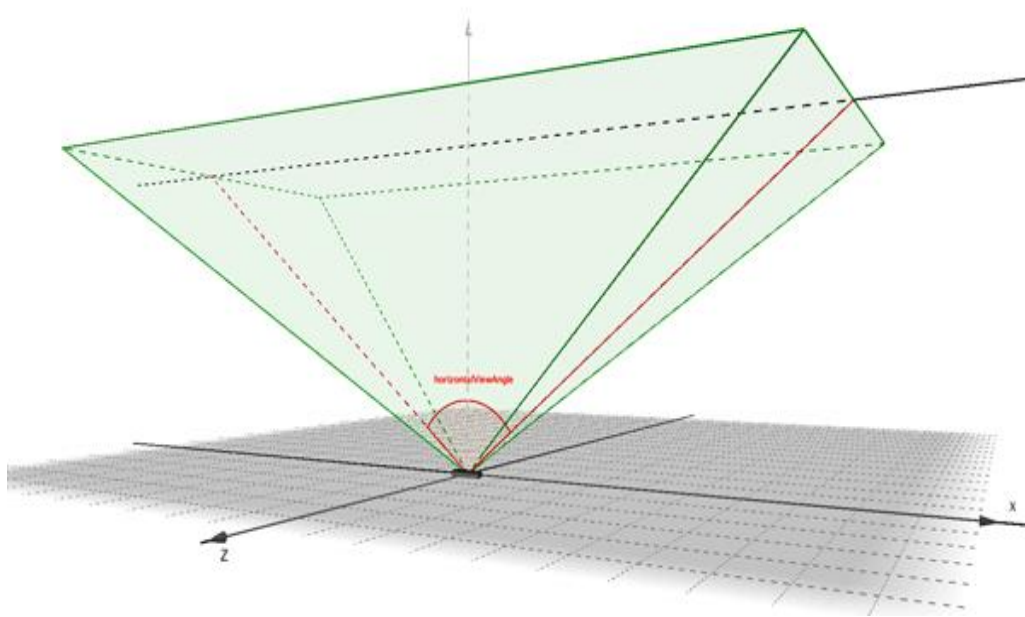


Figure 6. Gesture Recognition Range by the Leap Motion

In case of the traditional hand-tracking interface, errors involving movement recognition tend to occur in reality, because the interface could reckon the user's movement to manipulate other devices as VJing action. The reason why this happens is the performer necessarily makes action not involving with VJing in front of sensor sometimes. From the viewpoint of the interface device, there is no means to tell controlling action from performing action. Pre-gesture is a good way to express that it is beginning point of controlling action. To do this, we need to add a recognition mode in order to differentiate normal action and commanding action.

Figure 7 is a flow of algorithm used to switch to a recognition mode based on pre-gesture.

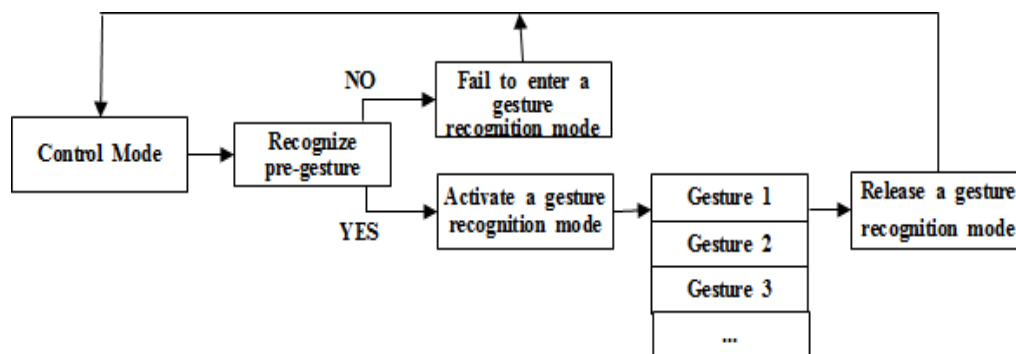


Figure 7. Flow of Switching to a Gesture Recognition Mode using Pre-gesture

As seen in Figure 7, the user, VJ carries out VJing in a normal control mode. If VJ wants to do controlling action, VJ could switch to a gesture recognition mode by showing pre-gesture explicitly. Once the pre-gesture is recognized by the system, the user enters a gesture mode. Since then, the actions taken during the gesture recognition mode are recognized as hand-gesture for controlling. In the gesture recognition mode, the system will do recognize hand-gesture action as predefined. If the recognition system does not

have any pre-gesture capability, all actions done by VJ are considered to be the controlling actions even when some of actions are not. In this system with pre-gesture capability, depending on whether the pre-gesture is recognized or not, gesture recognition mode will be activated or deactivated. Once the system enters the gesture recognition mode, the actions done by VJ are considered as the actions to control the video, so that the system executes the corresponding command as the result of recognition.

Also the pre-gesture needs to be explicitly different from normal actions. If the pre-gesture is similar to the general actions, normal action which is not pre-gesture is likely to be recognized as pre-gesture, which turns the system to gesture-mode and causes wrong control to VJing. The pre-gesture defined in this paper is as illustrated in Figure 8.

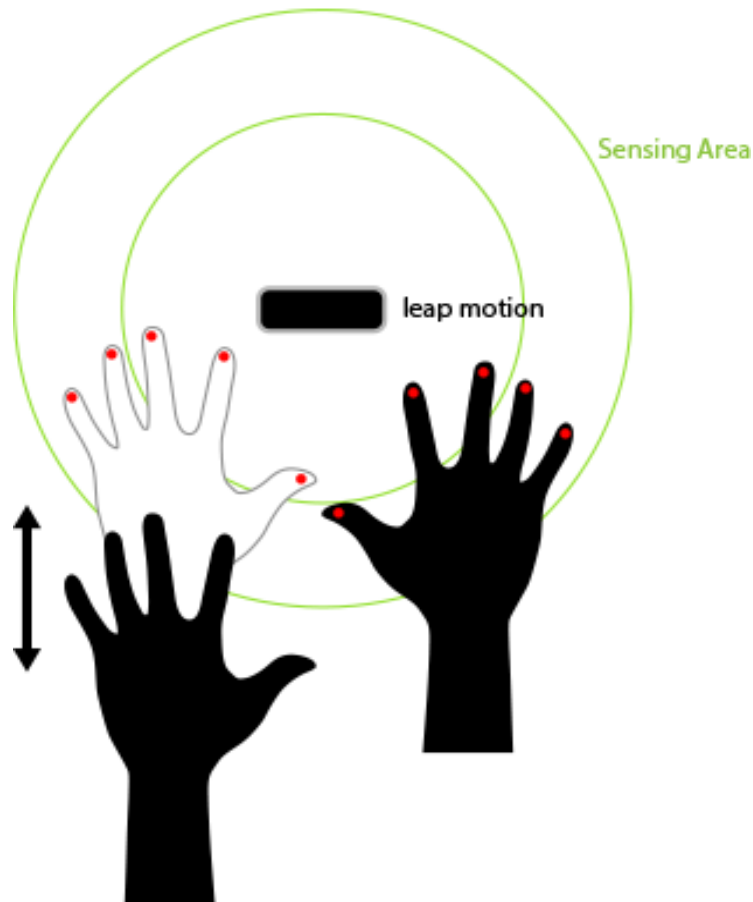


Figure 8. Pre-gesture used to Enter Gesture-mode

Figure 8 shows the pre-gesture necessary for entering gesture-mode. While both hands are being recognized within the recognition range, if the user pulls one hand out of the range and put it back within the range twice repeatedly, the system decides to enter gesture recognition mode. On the contrary, while the user stays in gesture recognition mode, if the user places the other hand inside of the range or pulls both hands out of the range, gesture recognition mode is released automatically. As described above, pre-gesture that can control to activate and deactivate gesture recognition mode could reduce the occurrences of wrong recognition of hand action which is not intended for controlling. Hence, the proposed system is effective to control realtime video in addition to music in more accurate manner by preventing the occurrence of wrong-action.

4. Conclusion

As high resolution project and large LED display device are developed, it becomes easy and simple to connect and control a wide variety of video sources in realtime basis at the performing place, which contributes to encouragement of audiences [10]. In this paper, we propose the hand gesture recognition user interface that enables single performer to control multiple devices playing video and music simultaneously in a multimedia performance production. Since the proposed interface is a non-contact type of user interface, it can be effective for DJ to control video while he/she carries out music mixing. The mechanism of the proposed interface is based on simple six hand gestures each of which corresponds to the command necessary to change the video channel. It is advantageous over the tradition interface such that only one performer can easily play both roles of DJ and VJing. However, it has also downside that this system could cause unnecessary error in the process of hand gesture recognition in regards to the recognition range of the Leap Motion. The Leap Motion is supposed to capture all hand gestures within the sensing range and interpret the gesture to control commands even if some of hand gestures are only actions for performance. Consequently, confusion of video-controlling gesture and performance gesture is likely to happen. To prevent these confusions, we propose the improvement of our interface in which the user can switch the system's state to gesture recognition mode explicitly using pre-gesture. Prior to carrying out video-controlling gesture, the user needs to activate the gesture recognition mode by presenting predefined pre-gesture. Once the system recognizes the pre-gesture, it turns to the gesture recognition mode. Only hand gestures presented during the gesture recognition mode will be captured and recognized as controlling gesture that triggers command. Similarly, the user can deactivate the gesture recognition mode by presenting predefined pre-gesture. The hand gestures presented during this normal mode will not be recognized by the system, which could prevent unnecessary wrong action due to wrong recognition. The improved system can be said to be effective to control video and music in more accurate manner by preventing wrong actions due to wrong recognition.

References

- [1] K.-M. Lee, "Designing Physical Game Interfaces Using Gestures", Journal of Korean Society of Design Science, vol. 24, no. 1, (2011), pp. 136-144.
- [2] J.-H. Rye, "Architectural Design Evaluation with Game Engine and Hand Gesture Interface", Architectural Institute of Korea, vol. 34, no. 1, (2014), pp. 49-50.
- [3] M. J. Kim, J. Heo, J.-H. Kim, S.-Y. Park and J. Chang, "Development and Evaluation of Leapmotion-based Interface Considering Intuitive Hand Gestures", Korean Society For Computer Game, vol. 27, no. 4, (2014), pp. 69-75.
- [4] M. J. Kim, J. Heo, J.-H. Kim, S.-Y. Park and J. Chang, "Development and Evaluation of Leapmotion-based Interface Considering Intuitive Hand Gestures", Korean Society for Game, vol. 27, no. 4, (2014), pp. 69-75.
- [5] J. Lee, "An User Interface of Live Video Mixing System Using Hand Gesture Recognition", Advanced Science and Technology Letters Art, Culture, Game, Graphics, Broadcasting and Digital Contents, vol. 101, (2015), pp. 5-9.
- [6] J. Lee, "An User Interface of Live Video Mixing System Using Hand Gesture Recognition", Advanced Science and Technology Letters Art, Culture, Game, Graphics, Broadcasting and Digital Contents, vol. 101, (2014), pp. 5-9.
- [7] J. Lee, "An User Interface of Live Video Mixing System Using Hand Gesture Recognition", Advanced Science and Technology Letters Art, Culture, Game, Graphics, Broadcasting and Digital Contents, vol. 101, (2015), pp. 5-9.
- [8] M. Kim, D. Kim and H. Kim, "A study on space and object expression using projection mapping", Journal of Digital Design, vol. 29, (2011), pp. 564.
- [9] E. Koh, "A Study on Performing Content through Projection Mapping", Graduate School of Techno Design Kookmin University, Master's thesis, (2012).
- [10] S. Noh, J. Lee and J. W. Park, "Implementation of Real-time VJing System for Live Projection Mapping", Journal of Korea Contents Society, vol. 13, no. 6, (2013), pp. 55-66.

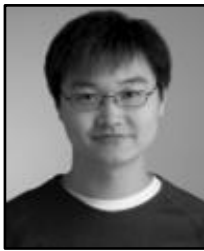
Authors



Joohun Lee, Dong-An Institute of Media and Arts, Dept. of New media Contents Professor.



Haehyun Jung, H2MEDIA CEO The Graduate school of Advanced Imaging Science, Multimedia & Film Chung-Ang University Mogencelab



Jaejoong Lee, The Graduate school of Advanced Imaging Science, Multimedia & Film Chung-Ang University Mogencelab

