

Large Display Interaction using Video Avatar and Hand Gesture Recognition

Sang Chul Ahn, Tae-Seong Lee, Ig-Jae Kim, Yong-Moo Kwon,
and Hyoung-Gon Kim

Imaging Media Research Center, KIST
39-1 Hawolgok-dong, Sungbuk-gu, Seoul KOREA 136-791
{asc, kij, lts, ymk, hgk}@imrc.kist.re.kr

Abstract. This paper presents a new system for interacting with a large display using live video avatar of a user and hand gesture recognition. The system enables a user to appear on a screen as a video avatar and to interact with items therein. The user can interact with the large display remotely by walking and touching icons through his video avatar. In order to build the system, we developed live video composition, active IR vision-based hand gesture recognition, and 3D human body tracking system, and incorporated a voice recognition system, too. Using this system, a user can interact efficiently with an embedded computer in a space that equipped with a large display.

1 Introduction

HCI(Human Computer Interaction) has been one of the most important research issues since the computer became one of the necessities of our daily lives. Many researchers are working on developing new interaction paradigms for efficient computer usage over mice and keyboards. As related works, there have been some researches to control computers by vision-based interaction. For instance, Kjeldsen tried to use gestures for computer interface[1]. The Perceptual Window offered a head motion based interaction technique[2]. Further, HCI techniques have been applied to virtual environment control[3], and building a smart interactive space[4,5].

Nowadays, as computers are getting smaller and smaller, and are embedded inside a lot of things around us. This embedded computing environment doesn't allow us to use mouse and keyboard type interface any more. We need a new type of interaction mechanism with things that have computing capability. In this environment, we don't care where the computers are located, but only confirm the result of our commands through some output devices like a large screen display. In this case, interaction in a space is required. The smart interactive spaces[4,5] are the examples of this kind of interaction.

This paper proposes a new mechanism of interaction with a large display using live video avatar of user and computer vision based hand gesture recognition. The proposed system enables a user to appear on the screen as a video avatar, and to walk and navigate across the screen. The system also allows the user to interact with items

or icons in the screen by touching them with a hand. The proposed system replaces the mouse cursor with live video avatar of a user. The user can control the computer by walking, navigating, and touching icons while seeing himself as a video avatar on the screen. In order to build the system, we have developed live video composition, active IR vision-based hand gesture recognition systems, and incorporated voice recognition system, too. Later we added 3D human body tracking to the system for 3D interaction. The proposed system can be applied to any tasks that can be done in a computer, but it is more efficient in some cases such as games and Powerpoint presentations.

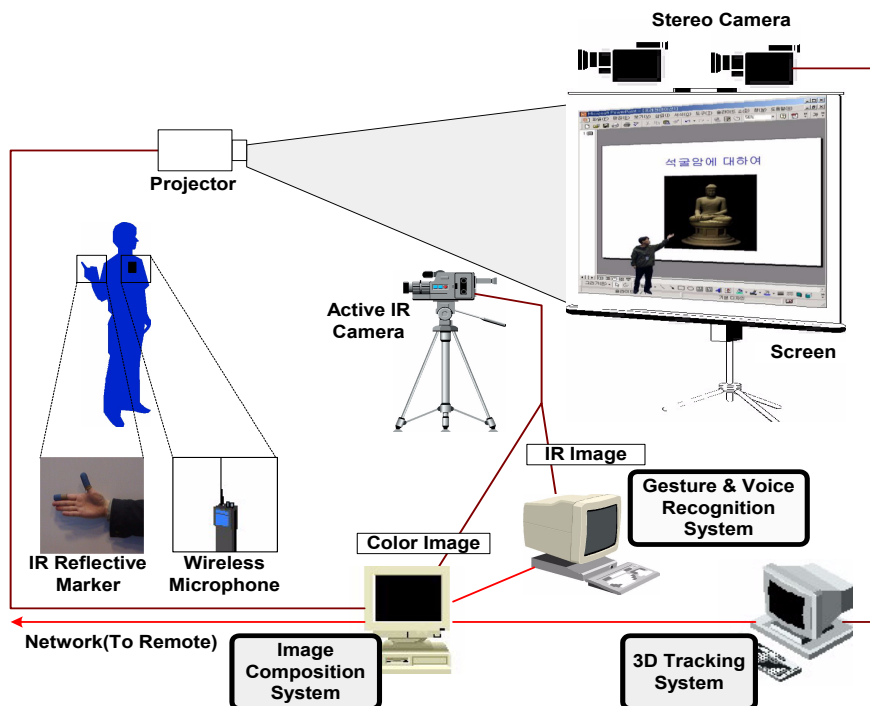


Fig. 1. Configuration of the proposed large screen interaction system

2 System Overview

The proposed system consists of three subsystems: *Image composition subsystem*, *Gesture & voice recognition subsystem*, and *3D Tracking subsystem*. Figure 1 shows the configuration of the system, and three computers represent the subsystems. The user interacts with the large screen in front of a camera while looking at the screen. The image composition subsystem extracts the user image and overlays it on the screen. So, the user sees himself as a video avatar on the screen during interaction. The user can interact with icons and items within the screen. He can use hand gesture and voice to control them. For robust hand gesture recognition, we used an IR (Infra

Red) camera and IR reflective markers. The 3D tracking subsystem uses a stereo camera and tracks the user's position. The user can use this subsystem for 3D interaction such as walking around a 3D object. As a whole, this proposed system gives us a more intuitive interaction method with a large screen.

3 Video Avatar

The proposed system uses a live video avatar of a user to give more intuitive sense of interaction. Since the image of a user is overlaid on a screen and moves across the screen, the user can feel as if he is in the computer screen. The live video avatar also has the role of mouse cursor in controlling and interacting with the screen.

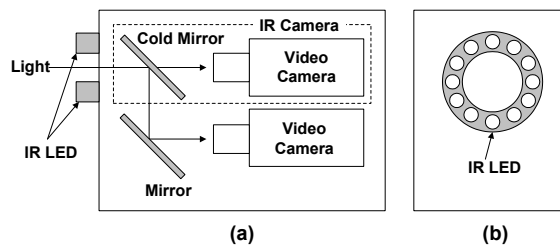


Fig. 2. Active IR camera, (a) side cut view, (b) front view

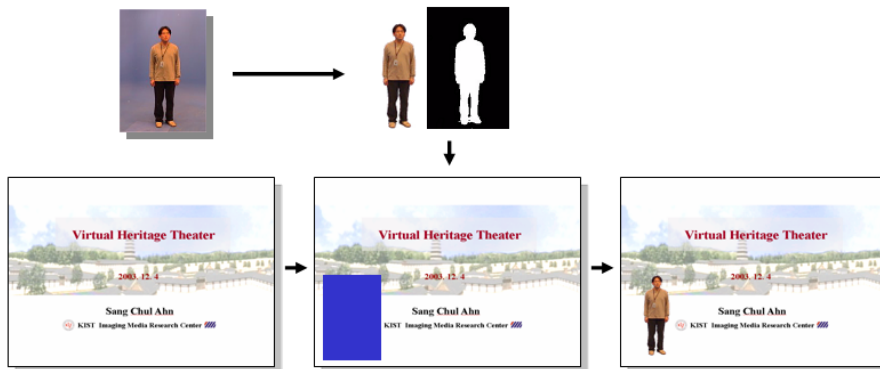


Fig. 3. Video avatar composition

In order to make a live video avatar, the image composition subsystem captures the images of a user with video camera. In Figure 1, we can see that an active IR camera is capturing the images of the user. This active IR camera is a combination of an IR camera, an IR light source, and color video camera. The structure of the active IR camera is shown in Figure 2. The IR camera and IR light source are used in the gesture & voice recognition subsystem, which will be explained in the next section. The image composition subsystem uses the output of the color video camera. The image

composition subsystem extracts the image of a user from the background, and uses it as a live video avatar. Background subtraction or chroma-keying method can be used in this process. The video avatar is overlaid on the screen. The composition of video avatar can be done with a video processing hardware or a chroma-keying hardware. However, we found that we could implement it in realtime using a chroma-keying function of the Windows XP. The user can control the position of the video avatar by walking around the stage as long as he remains within the field of view of the video camera. Figure 3 shows the image composition mechanism of video avatar.

4 Gesture & Voice Interaction

The proposed system uses hand gesture and voice as the main interaction tools. The gesture & voice recognition subsystem enables the user interaction by vision-based hand gesture recognition and voice recognition. For instance, a user can issue a command to play music by saying “Music Start” to a microphone. By hand gesture “Up / Down / Left / Right”, he can also move his video avatar across the screen to select or play an item. This is useful when it is too far from him to reach by walking within the field of view of camera. The implemented user commands are shown in Table 1. As can be seen, there are many voice commands, and hand gesture commands are included in the voice commands. However, hand gesture commands are more useful for 2D/3D motion than voice commands. Voice commands are more useful for discrete actions such as start or end of some action. Some of the commands are implemented for interactive Powerpoint presentation.

Table 1. The implemented commands for interaction

Type	Commands	Comments
Hand gesture	Left / Right Up / Down	2D avatar move or 3D navigation
	Forward / Backward	3D navigation
	Select	Left mouse button Double click
Voice	Left / Right Up / Down	2D avatar move or 3D navigation
	Slideshow Next / Previous	Powerpoint full screen Proceed or go back
	Select / Close	Selection/Window close
	Pen Mode Pen Mode Exit	Powerpoint pen mode
	Screen Keyboard	
	Navigation Forward / Backward	3D navigation start 3D navigation
	Navigation Stop	3D navigation stop
	Music Start Equalizer	Media Player start Media Player equalizer
Help	Call a helping MS Agent	

Although computer vision-based hand gesture recognition has been widely studied, it also inherits the shortcoming of most computer vision algorithms: sensitiveness to lighting condition. Thus, we adopted active IR based recognition method for robust recognition. As mentioned before, an active IR(Infra Red) camera is used for capturing a user's hand gesture. The IR camera can be made with a normal video camera and an IR filter. In Figure 2, the upper part composes the IR camera. The cold mirror is an IR filter that absorbs IR rays while reflecting visible rays. We used a cold mirror that absorbs the rays of above 800nm in wave length. Additionally we made two IR reflective thimbles for user's hand. The IR reflective thimbles were made with retro-reflective material so that they can be viewed best from the camera with IR light source. Figure 4 shows the IR reflective thimbles and a user's hand wearing them. Since thumb and index finger are mostly used for selecting and pointing, we decided to use the thimbles for thumb and index finger.

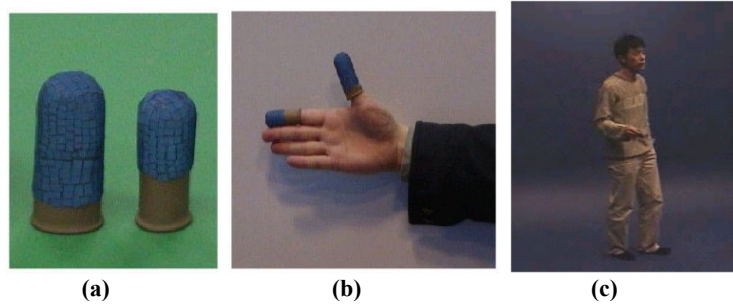


Fig. 4. IR reflective thimbles, (a) thimbles, (b) a hand wearing thimbles, (c) user interacting by hand gesture.

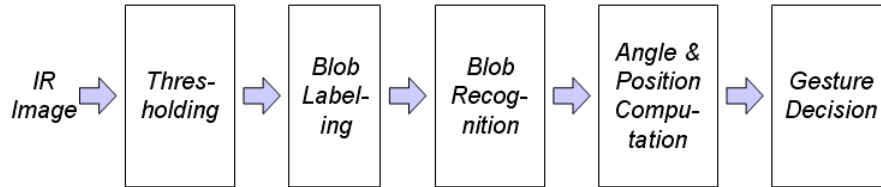


Fig. 5. The flow chart of hand gesture recognition.

Figure 5 shows the flowchart of hand gesture recognition algorithm. Since the IR reflective thimbles look white in an IR image, we can extract the regions by simple thresholding. Then, we apply the labeling operation to the regions, and find thimble regions of thumb and index finger by size. The hand gesture is recognized by relative position and direction of thimbles. The center position (x_c, y_c) of a thimble is determined using the 0th and 1st moments as follows,

$$(x_c, y_c) = \left(\frac{M_{10}}{M_{00}}, \frac{M_{01}}{M_{00}} \right), \quad M_{ij} = \sum_x \sum_y x^i y^j I(x, y),$$

where $I(x, y)$ is the intensity value of IR image at (x, y) position. M_{ij} is x and y directional moment. We can also compute directional angle using 2nd moment as follows,

$$\theta = \frac{1}{2} \tan^{-1} \left(\frac{b}{a-c} \right), \quad a = \frac{M_{20}}{M_{00}} - x_c^2, \quad b = 2 \left(\frac{M_{11}}{M_{00}} - x_c y_c \right), \quad c = \frac{M_{02}}{M_{00}} - y_c^2.$$

Then, we check the following 5 conditions to determine relative position.

Condition 1: $y_Y - y_G \geq \frac{H_Y + H_G}{2} + \alpha_1$; Yellow blob is below green blob

Condition 2: $y_G - y_Y \geq \frac{H_Y + H_G}{2} + \alpha_1$; Yellow blob is above green blob

Condition 3: $x_Y - x_G \geq \frac{W_Y + W_G}{2} + \alpha_2$; Yellow blob is right of green blob

Condition 4: $x_G - x_Y \geq \frac{W_Y + W_G}{2} + \alpha_2$; Yellow blob is left of green blob

Condition 5: $|x_G - x_Y| \geq \frac{W_Y + W_G}{2}$; Two blobs are close each other vertically

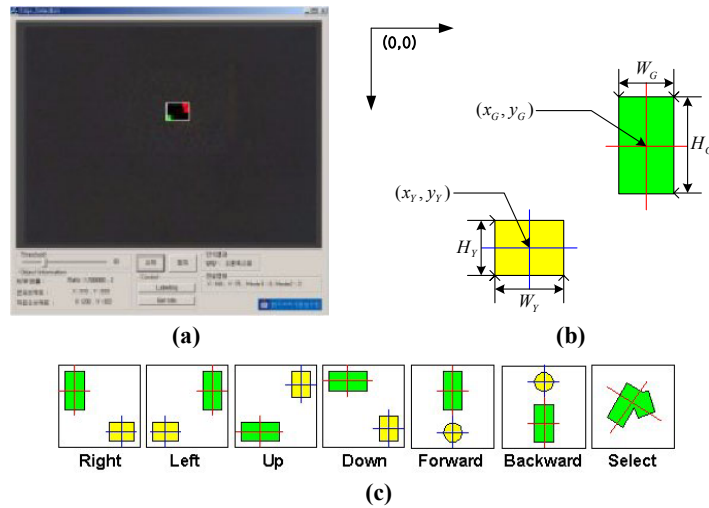


Fig. 6. (a) Screen shot of the hand gesture recognition system, (b) Metrics of thimbles, (c) typical configuration of hand gesture image.

Finally we interpret hand gestures as shown in Figure 6. Note that the green(longer) and yellow blobs represent thimbles for thumb and index finger, respectively. α_1, α_2 are predefined small values to make up the distance between thumb and index finger. For instance, if all the directional angles of yellow and green blobs are 90 degree, and the condition 1 and 5 are satisfied, the hand gesture is recognized as the “forward” command.

Once the hand gesture is recognized, it is converted to a command and sent to image composition subsystem by network. Then, image composition subsystem applies

an appropriate action to the screen. Here we have to make the IR image have the same coordinates to that of normal video image. That's why we use the active IR camera structure as in Figure 2.

5 3D Interaction

In the proposed system, a user can use 3D interaction. Figure 7 shows one of the examples, where a user shows a 3D model of Buddha and his video avatar is walking around it. It can be seen that part of the avatar is occluded by the statue. In this way a video avatar can walk and navigate the 3D space. We implemented it by invoking our 3D model viewer program. The 3D tracking subsystem extracts a user out of the images from stereo camera, and gets his 3D position. The information is sent to image composition subsystem so that it can overlay a video avatar at appropriate 3D position. Figure 7(b) shows a screen shot of 3D tracking system.

6 Application

We applied the proposed system to build an intelligent room called "Smart Studio". In the Smart Studio we could interact with a large screen and use an embedded computer in the environment. We could enjoy music and navigate internet as we want. We could enjoy some games with whole body action, too. From our experiments, we noticed that the proposed system provides efficient interaction mechanism in space and that it is quite effective in games and Powerpoint presentation applications. Figure 8 shows some of the screen shots of various applications.

7 Conclusion

In this paper, we have presented a new system for interacting with a large screen, which enables a user to appear on the screen and to interact with icons and items therein. The proposed system incorporates live video composition, voice recognition, active IR vision-based hand gesture recognition, and 3D human body tracking. The screen output of the system looks similar to TV news, but is different in that a user can control every item on the screen in realtime. The proposed system has the following advantages. First, a user can feel somewhat immersive sense since he appears as a video avatar on the screen and can touch icons by hands. Secondly, a user can feel naturalness with controlling by walking or touching and picking icons and items. We applied the system to building an intelligent space and found that it worked efficiently. As embedded computing environment is coming to our lives, this type of interaction is expected to be used. Currently, we are extending this system to a teleconferencing system that shows video avatars of every participant and shares the same workspace.

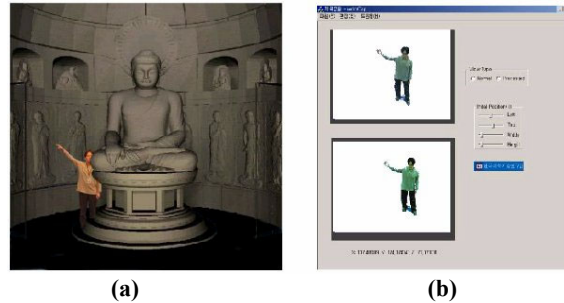


Fig. 7. 3D tracking result, (a) a user walking around a statue of Buddha, (b) screenshot of the 3D tracking system.

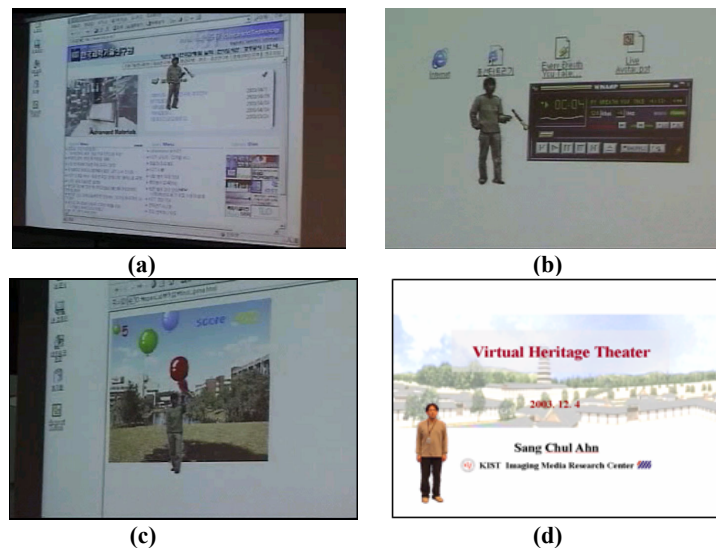


Fig. 8. Screen shots of various applications, (a) internet surfing (b) music control (c) game (d) Powerpoint presentation.

References

1. R. Kjeldsen, J. Kender: Toward the use of gesture in traditional user interfaces. Proceeding of Automatic FGR'96. (1996) 151-156
2. J.L. Crowley, J. Coutaz, and F. Berard: Things that see. Communications of the ACM. vol. 43. no. 3. Mar. (2000) 60-61
3. I.J. Kim, S. Lee, S.C. Ahn, Y.M. Kwon, H.G. Kim: 3D tracking of multi-objects using color and stereo for HCI. Proceeding of ICIP2002. (2002)
4. A.F. Bobick, S.S. Intille, J.W. Davis, F. Baird, C.S. Pinhanez, L.W. Campbell, Y.A. Ivanov, A. Schutte, and A. Wilson: The Kids Room. Communications of the ACM. vol. 43. no. 3. Mar. (2000) 60-61
5. Brumitt, B., Meyers, B., Krumm, J., Kern, A., and Shafer, S.: EasyLiving: Technologies for Intelligent Environments. Handheld and Ubiquitous Computing, September (2000)