A Layered Authoring Model for Constructing Interactive Virtual Environment

Michitoshi Ishiwaka and Seiki Inoue

ATR Media Integration & Communications Research Laboratories

{ishiwaka, sinoue}@mic.atr.co.jp

Abstract: It is difficult and important for virtual studio systems and media-art products to be designed effective interactions between a performance person, who appears in the studio, and the virtual studio equipments, and, an appreciation person and artistic products. The difficulties originate with no even-grained component and no framework of interaction design work. In this article, a layered authoring model is proposed in order to facilitate system designers and artists to construct the system more easily and flexibly using fine-grained media components. And some experimental interactive virtual environments based on the model are shown.

1. Introduction

Video productions using virtual studios and production of media art and other works of art using virtual reality (VR) technology are becoming increasingly popular. A virtual studio makes it possible to present sophisticated images by varying computer graphics (CG), synthesized images, and background music in response to the actions of a performer situated in the studio. In media art as well, an author can present images in a more effective way through interaction between the art appreciator and the work.

When examining such production and presentation environments from the viewpoint of system construction, we see that they often take the shape of a real-time distributed processing system. System construction in such a case is not simple. For example, the sensor subsystem that analyzes the movements of the performer must be linked in real time with the generation subsystem that varies synthesized images according to the results of analysis. In addition, system specifications are decided in part according to the elements making up the processes of image presentation, art creation, etc., and this increases the difficulty of system construction. An example of this situation is determining how the work responds to various approaches made to it by the appreciator.

To therefore simplify system construction, we consider the possibility of dividing the sensor system, presentation system, and exchange system that interconnects these systems into respective components at an appropriate level of granularity, and of integrating and re-integrating these components in a flexible manner. The issue to be addressed here involves the granularity of the components and associated viewpoints. For example, the granularity of the sensor system from the viewpoint of video presentation and image presentation would be a person's actions and behavior and even high-order information units such as whether that person is happy or sad. On the other hand, the granularity of the sensor system from the viewpoint of system construction would be determined by applying an appropriate technique to CCD-camera input, i.e., data input/output units in the form of coordinates specifying the positions of hands and feet every frame period. In short, there is a gap between the viewpoints of presentation and system construction. Nevertheless, by setting component granularity for each viewpoint in a step-like manner, we can expect the construction of an image-presentation environment to be simplified without suffering a loss in component reusability and freedom of expression.

We have previously proposed an integrated media manipulation environment called COMI&CS (Computer Organized Media Integration & Communications System)[1, 2]. Our aim here was to construct an environment that enables images to be presented in a relatively simple manner using multimedia components with appropriate granularity, and we have been researching and developing this system to this end. In this paper, we report on a prototype framework for multimedia components and interaction design that will make it possible to create at will an interactive virtual world as a means of allowing anyone to easily express themselves through images. Based on our experiences with this prototype, we examine the granularity of multimedia components and interaction design for image presentation.

The following section overviews the prototype interactive virtual-space construction environment and describes the multimedia components that make up the constituent elements of this environment. Section 3 organizes and examines component granularity and the viewpoints associated with system components, and presents a concise interaction design model. The paper concludes with a summary and description of future topics for study.

2. A case study: Constructing an interactive virtual world

An interactive virtual world provides a framework in which a character in the virtual world can interact with the constituent elements of that world. To say it another way, the virtual world acts in concert with the independent actions of the character. The hardware configuration of the prototype environment is shown in Fig. 1. Here, the character in the blue-back studio watches his own figure within the virtual world displayed on the projector in front.

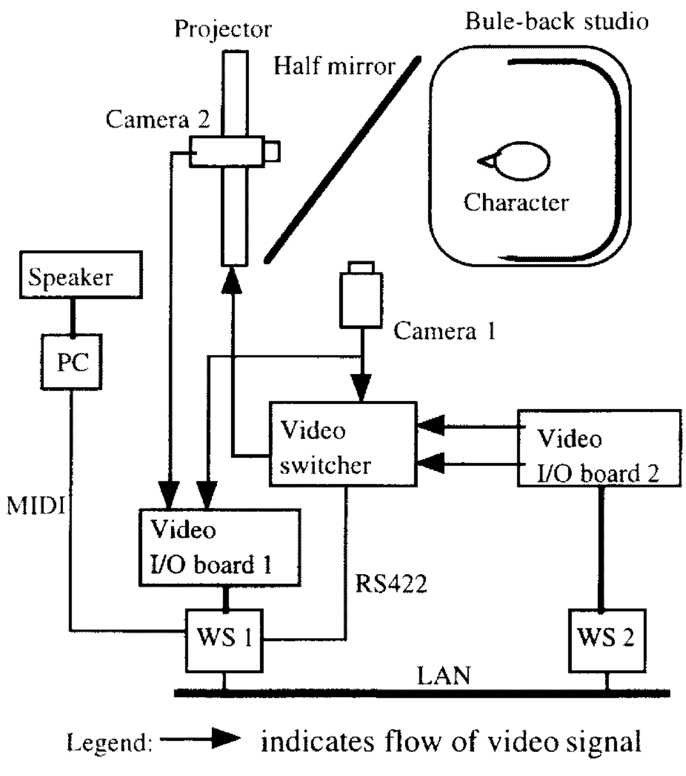

Figure 1: Hardware configuration diagram

The character can also hear sounds and music in response to his actions.

The following describes the sensor system that detects actions made by the character in the virtual world, the presentation system that consists of elements making up the virtual world, and the exchange system that specifies the interactive relationship between the character's actions and the virtual world.

Sensor system

The sensor system captures the actions of the character in the virtual world and outputs information related to those actions. It consists of input devices like cameras and magnetic sensors, and programs that analyze the data received from these devices and output results. In the prototype system, actions of the character in the virtual world consist mainly of the position he stands at in the studio and signals that he generates with gestures like raising his right hand. Two cameras are used in this sensor system: one camera faces the performer and points down towards the floor to pick up the position of his foot steps (Fig. 1, camera 2), and the other camera picks up the entire image of the performer from the front (Fig. 1, camera 1) to detect his signals. The following three steps summarize this procedure.

1) The images taken by each camera are separated into character region and background region using hue and then digitized.

2) The character's standing position is detected by first determining the rectangle circumscribing the character region and then determining the coordinates of the center point of the rectangle's base.

3) Signals are detected by comparing the following feature quantities with pre-registered feature quantities: the area of the character region, the height-to-width ratio of the circumscribing rectangle, and the occupation ratio and center of gravity of the character region in each of five rectangles formed by dividing the circumscribing rectangle into five equal parts [3].

Presentation system

The presentation system features internal data made up of pictures and music that form the constituent elements of the virtual world, and outputs these elements according to control information from the outside. In the current implementation, we developed a component (Fig. 1, WS2) that outputs three-dimensional computer graphics and actual pictures with depth information[4] as visual space in the virtual world, and a component (Fig. 1, PC) that outputs short musical phrases as sound space in the virtual world.

The visual information is combined with the image of the character in the studio resulting in the presentation of a three-dimensional virtual world. Specifically, the system outputs picture

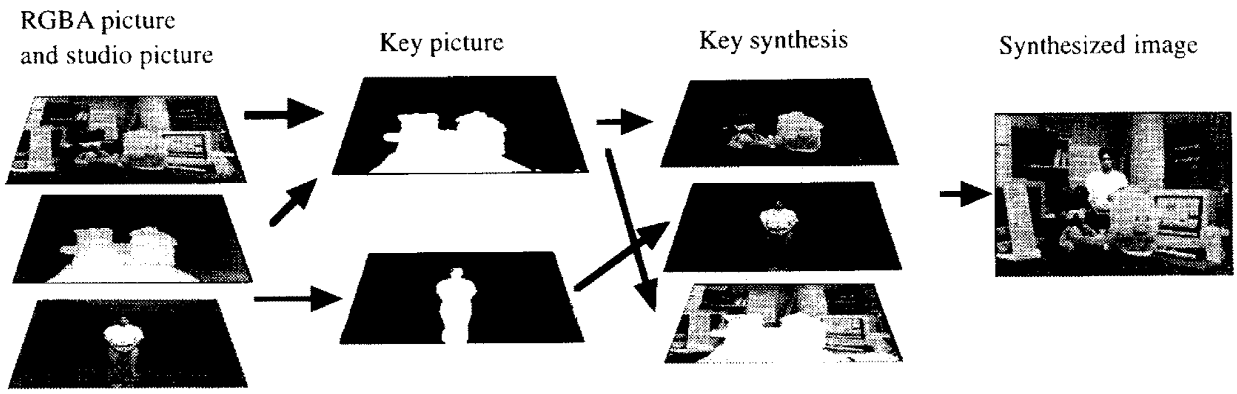

Figure 2: Example of picture synthesis

data storing image data for display over the RGB channel and depth data for the alpha channel via a video board (Fig. 1, video input/output board 2), and combines this data with studio images using image synthesizing equipment (Fig. 1, switcher). Here, the use of depth pictures as the key source makes it possible to perform three-dimensional synthesis. An example of this process is shown in Fig. 2.

Listening information is output according to control information from the outside. In this process, the system selects a musical phrase and then varies music tempo, volume, and acoustical effects like reverberation in real time to present a sound space. In actuality, a MIDI sequencer is controlled here.

Exchange system

The exchange system treats information output from the sensor system as input and outputs control information to the presentation system. It consists of components having functions like selection of data to be passed from sensor-system output to control information, type conversion, and computer network communications.

The next section describes how the exchange system can be used to decide what kind of virtual world to construct.

3. Multimedia components and interaction design

By integrating the various components from the sensor system, presentation system, and exchange system described in section 2, a model for constructing one interactive virtual world can be considered. We look at interaction design from the viewpoints of image presentation and system construction, and present the interactive virtual world construction model shown in Table 1.

At the low-data I/O level, which includes the system's hardware configuration, data is used as material for image presentation.

At the media handling level, interaction is designed based on sensor-system and presentation-system information in symbolic form. Here, components of the sensor system and presentation system treat the various types of low data as their internal data and are configured to accept methods at the symbolic level. In this way, even if certain components of the sensor system and presentation system are changed in a new installation, the reusability of the exchange system can be maintained. For example, our current implementation makes use of a camera as one component of the sensor system. Changing this component to a magnetic sensor would not require that the symbol "right hand raised" be changed.

At the image handling level, interaction is

Table 1: Interactive Virtual World Construction Model

	Sensor System Components	Exchange System components	Presentation System Components
Image handling	Emotion sensor Ex: "Spectator looks happy", "Performer looks sad"	Image connection Ex: "**if** happy, **then** present a happy world"	Virtual world presentation Ex: "Happy image", "Happy music"
Media handling	Person's movement sensor Ex: "Right-hand raised", "Move to left"	Media connection Ex: "**if** right-hand raised, **then** generate background image"	Image and music generation Ex: "Generate and regenerate RGBA file"
Raw data I/O	Camera sensor Ex: "Silhouette picture as feature quantity"	Raw data connection Ex: "Hardware configuration"	Image and music materials Ex: "RGBA picture file", "MIDI file"

designed based on semantic units of information for image presentation. The components of the sensor system and presentation system here treat information from the media handling level as internal data and are configured to accept methods at the semantic level. Here as well, the reusability of the exchange system can be maintained even if interpretation changes with respect to sensor-system and presentation-system components.

An example of constructing an interactive virtual world at the media handling level is shown in Fig. 3.

Figures 3(a) and 3(b) present the expanse of three-dimensional space in the virtual world. Figure 3(c) presents both the expanse of space and lively motion in the virtual world as a ball rotates, stops, and reverses rotation in response to the character's movements. Moreover, although impossible to represent on paper, musical phrases change in response to movement to create an even more impressive performance.

Construction of an interactive virtual world at the image handling level is presently under study. Three major issues in this regard are described below.

(a)

(b)

(c)

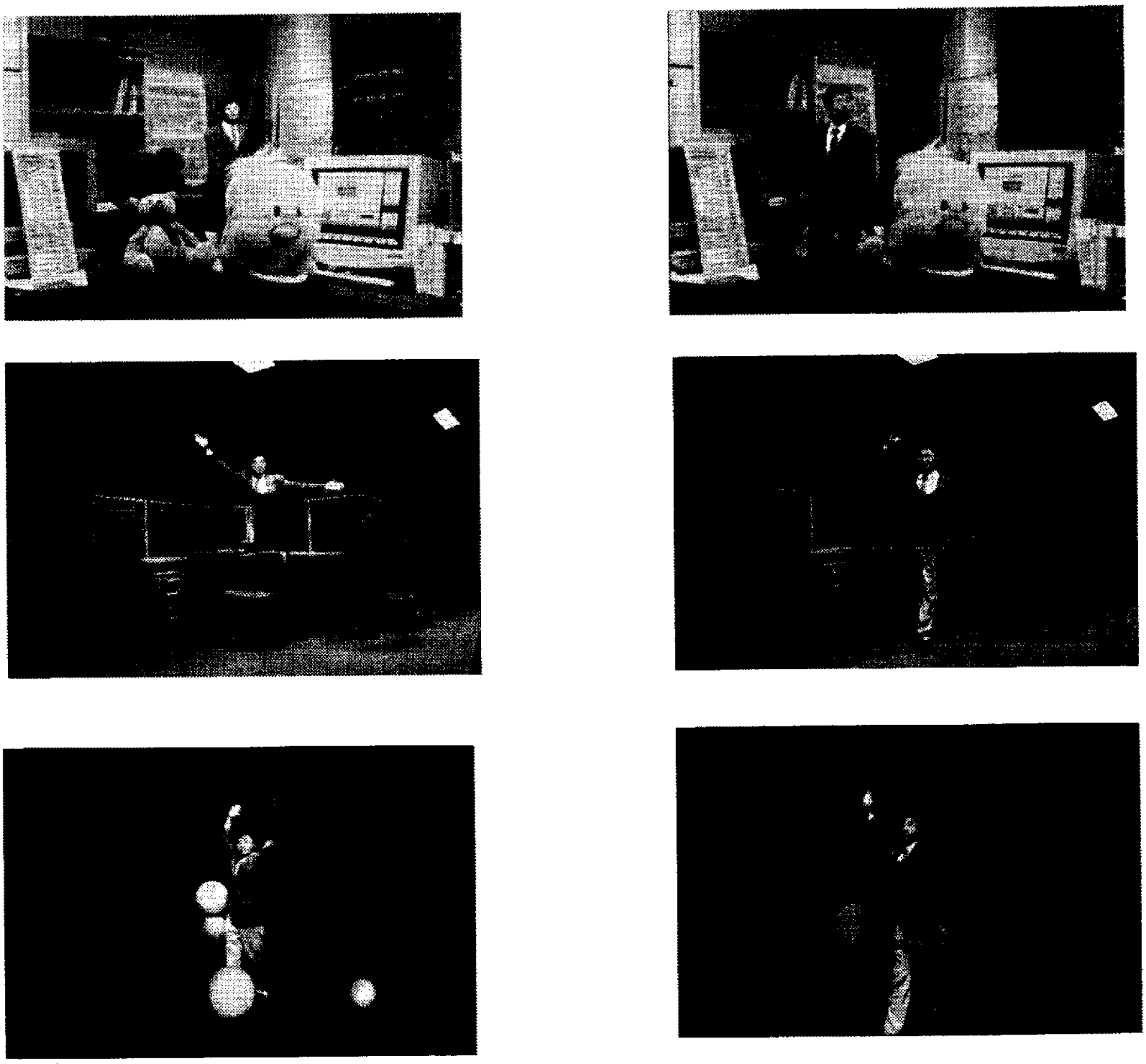

Figure 3: Example of constructing an interactive virtual world

1) Implementation of media handling components with modifying meaning

For example, sensor-system components with meanings like "quickly" and "smoothly" that modify actions would be installed, as would similar components in the presentation system.

2) Introduction of an image presentation state

For example, considering presentation in the virtual studio, it may be desirable at times to convey different meanings for the same action like "right hand raised" in the virtual world. In such a case, presentation state control would be achieved at the image handling level.

3) Implementation of image handling exchange-system components as a context-dependent interpretation system

In relation to issues 1) and 2) above, we are studying a configuration in which word meaning is assigned to the media handling level and a text-interpretation system is assigned to the image handling level as a way of dealing with multiple meanings of modifiers and context-dependent image presentation and thereby avoid inconsistencies.

4. Conclusion

This paper has described multimedia components and interaction design with the goal of enabling anyone to construct at will an interactive virtual space and to create images by interacting with that space. It was shown that dividing the sensor system, presentation system, and exchange system into components, whose step-like granularity corresponds to a symbolic level as seen from the viewpoint of system construction and a semantic level as seen from the viewpoint of image presentation, can be an effective means of presenting images.

As the next step in this research, we plan to further enhance the functions of each system's components and to continue our study of interactive virtual world construction at the image handling level as described in section 3. We also plan to evaluate the image presentation environment by performing more experiments in the construction of interactive virtual worlds.

References

[1]M. Ishiwaka and S. Inoue: "An Integrated Media Handling Environment COMI&CS and its Handling Parameter Exchange Method", IPSJ Technical Report CG80-7, 1996. (in Japanese)

[2]S. Inoue, M. Ishiwaka, S. Tanaka and J. Park: "A Study on Image Expression Environment IE Room", ITE Annual Convention, 1997. (in Japanese)

[3]M. Ishiwaka and S. Inoue: "A Study of Constructing Media Handling Environment according to Human Poses", IPSJ 55th Convention, 3-164, 1997. (in Japanese)

[4]J. Park and S. Inoue: "Depth Extraction Using Multiple Cameras for Video Composition", IEICE Technical Report, PRMU96-113, 1996. (in Japanese)