Gesture Interface for a Virtual Walk-through

Kenji Mase and Rieko Kadobayashi
ATR Media Integration & Communications Research Labs, Kyoto, Japan
{mase,rieko}@mic.atr.co.jp

1 Introduction

Museums are great archives of society and display wonderful natural phenomena, historical artifacts, artistic masterpieces, and human knowledge for visitors to come and see. We are investigating the presentation styles of future museums and have proposed the Meta-Museum concept[1]. Meta-Museum aims to provide a communication environment between experts, e.g., researchers and curators, and non-experts, e.g., visitors, by using seamless integrated presentations of virtual objects and exhibits.

One of the issues confronting realization of such a Meta-Museum is how to provide easy-to-use interfaces for exploring a virtual world and integrated presentations. We have developed VisTA (Visualization Tool of Archaeological data), which is a temporal transition simulation system for an ancient village. VisTA users have the freedom of walking-through a virtual village, changing the hypothetical lifetime of buildings and accessing reference records through WWW browsers.

In this paper, we discuss design of the gesture interface of the VisTA-Walk system, which substitutes mouse interactions with full body gesture interactions, e.g., walking-through a virtual space and accessing reference information about objects. We use the pfinder[2] program module as a perceptual (gestural) user interface library and integrate it with VisTA. Events from the pfinder program are translated as mouse events and fed to the VisTA system. In the following sections, we describe the system configuration of VisTA-Walk and how gestures are used for a walk-through system interface.

2 VisTA-Walk

VisTA simulates the temporal transition of ancient villages for archaeological researchers and provides users with a walk-through viewer. The reference records of excavations are accessible via a WWW browser located on-screen. Web pages such as vestige site excavation records are hyper-linked from a vestige database shown with a viewer. The database contains a typical instance model and the lifetime of each house.

Figure 1 illustrates the VisTA-Walk system and shows an interaction scene. *Pfinder* is the gesture recognition program and uses one video camera mounted on top of a screen. It was originally designed for full body interactions in interactive virtual environments[3]. We have set the camera on top of a large screen (170 inches), too, and use several of the pfinder outputs, e.g., standing position, crouching position, and stretching of arms (left and/or right), which are translated into mouse events. VisTA-Walk contains models of the land of a village, its vestiges and buildings, and the hypothetical lifetime of each building. The models are written in and controlled under Open Inventor. The database keeps reference records about the vestiges in the form of a HTML. A reference is displayed on the browser if a building is selected in the main viewer.

3 Gesture Interface of VisTA-Walk

The following are major interactions of VisTA: (i) walking-through a virtual space for exploration, and (ii) pointing at an object in a virtual space to get reference records. It is acceptable for a user to use different parts of the body for these two independent interactions. Based on this principle, a plausible example of interaction design would be that a standing position controls walking-through and hand gestures control pointing.

3.1 Gestures for Walk-through

How can we assign a standing position to the control events of walking-through? There are at least three choices of metaphor for control: mouse, joystick[3], or steering wheel and accelerator. The positioning action of a mouse is that of a locator-type device, while a joystick and a steering wheel are valuator-type devices. We employ a steering wheel-like control scheme for gesture interaction in the VisTA-Walk system, while providing mouse control for desktop interaction.

Similar to the movement of a steering wheel, stepping aside from the neutral position is used to steer the direction of walking. Stepping forward or backward is assigned to an accelerator with which a user moves the viewing position forward or backward in the virtual environment. In order to stop at a desired point in a scene, a user must return to the neutral position.

The speed of steering and walking is proportional to the stepping distance from the neutral position, and the

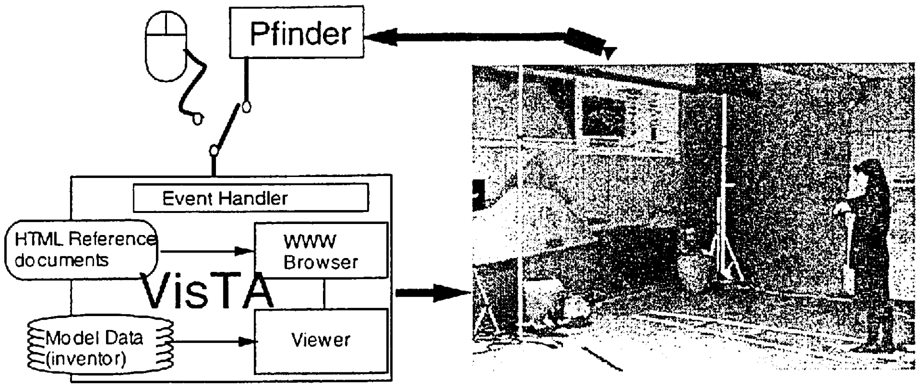

Figure 1: A scene in the VisTA-Walk system and the system block diagram: gesture-based walk-though and information access in a virtual ancient village. A TV camera is set on top of the screen.

ratio of speed to distance affects the assessment of usability. If the speed is too fast, a user can quickly arrive within the proximity of a destination point, but it is hard to stop exactly at the desired point. On the other hand, if the speed is too slow, a user may step far from the neutral point to quickly arrive at a point; stopping at the desired point necessitates moving back by several steps, and this causes inaccuracy. However, the usability test tells that users will enjoy traveling in a slow speed and arrive to the desired point withough difficulty.

Walking-through is realized by means of controlling the viewing position of a virtual 3-D space with a virtual camera. A 2-D position on the terrain surface is necessary and sufficient when we only simulate walking actions. However, changes in the height of the viewing position and the viewing elevation facilitate immersive sensations for a user.

3.2 Gestures for Pointing Objects

The current interpretation of pointing gestures are a "left one" or a "right one". The VisTA-Walk gesture interpreter uses the output recognition results of Pfinder, e.g., "stretch left/right hand" or "raise left/right hand". The interpreter chooses an object on the left (right) side when it detects a "left (right) one" request from Pfinder outputs. The pointing actions recognized by Pfinder are insufficient to locate a particular object among many, and this is due to the camera position. Pointing and stretching a hand toward the screen is hard to detect because the camera is on top of the screen. Stereoscopic multiple camera arrangement is necessary to get complete 3-D information about hand and body gestures[4], and such an arrangement is presently unrealistic, because many cameras would be necessary to cover the complete area of a user in motion.

Ambiguity is not the problem at all because a user is situated in the virtual space and is able to move around. Suppose a user wants to indicate a particular object from among many at a distance. She only needs to come closer to the object and then stretch her hand toward the object. Moreover, this is the case when no distinctive feature other than the location is available. We will integrate other media such as voice to complement a situation such as saying "I want that house on the left with the red roof."

4 Conclusion

A gesture-based walk-through interface in virtual space supplies a novel, immersive experience to people. We can easily get the feeling of walking by stepping back and forth, even in a limited physical space. The barrier or fear of using a computer with a mouse and keyboard has been greatly reduced by the VisTA-Walk gesture interface, and we think this is good for general visitors of museums.

Acknowledgments

The authors would like thank to Yasuyoshi Sakai and Ryohei Nakatsu, for their support and encouragements in this research. The Pfinder gesture recognition system used in the VisTA-Walk was provided by the Perceptual Computing Section, Media Laboratory, MIT.

References

[1] Kenji Mase, Rieko Kadobayashi, and Ryohei Nakatsu, "Meta-museum: A supportive augumented reality environment for knowledge sharing," *Int'l Conf on Virtual Systems and Multimedia '96*, pp. 107–110, Gifu, Japan Sept. 1996.

[2] Cris Wren, Azi Azarbayejani, Trevor Darrell, and Alex Pentland, "Pfinder: Real-Time Tracking of the Human Body," *2nd International Conf. on Automatic Face and Gesture Recognition*, Killington, Vermont, Oct. 1996.

[3] Christopher R. Wren, et. al. "Perceptive Spaces for Performance and Entertainment: Untethered Interaction using Computer Vision and Audition," *Applied Artificial Intelligence*, **11**, 4, pp. 267–284, June 1997.

[4] Masaaki Fukumoto, Kenji Mase, and Yasuhito Suenaga, "Finger-pointer: Pointing interface by image processing," *Comput. & Graphics.*, **18**, 5, pp. 633–642, May 1994.