Gesture + Play

Konrad Tollmar, David Demirdjian & Trevor Darrell

Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

http://www.ai.mit.edu

The Problem: Physical and perceptual interfaces to games and virtual environments are an exciting new interface paradigm. Recent advances in computer vision make real-time sensing of users' position and pose possible using relatively low-cost sensors. However, little thought has been given to the interface abstractions that are most appropriate in this medium, or to the relationship of the physical space with the perceptual controls.

Motivation: As a part of this project we have developed a perceptual interface toolkit based on stereo vision sensing. Stereo vision allows accurate estimation of 3-D position and orientation cues, and also allows robust segmentation of the image of a user from other objects or people in a scene. Hence could our toolkit track users' movement and gesture in real-time and that is robust to crowds, lighting changes, motion, and clothing variation (without any special background or lighting effects, like key-chroma or IR [2]).

Our toolkit returns the 3D position and articulated body posture of a user as they move in an arbitrary space (see figure 1). In our case does these gestures, actions, and motions cause motion and action in a virtual workspace, e.g. in a game or avatar-world [6]. However these estimates must be transformed into abstractions of gesture, motion, and action and mapped to controls that are useful for an application.

Hence our major objectives in this study is to compare different abstract interface regimes for body- and gesture-based control of games and virtual environments using a passive unthethered perceptual interfaces. What gestures should control which virtual actions? For example should navigation control be relative or absolute? Direct or indirect? How should the physical shape of the room (floor or other surfaces) be designed to support the interaction?

Previous Work: Full-body, vision-based perceptual interfaces to virtual worlds actually have a long history of installations in the computer graphics field as well as on the art scene. The pioneering work of Krueger in the 1970's first combined an image of the user in a virtual world, and allowed 2-D gestural interaction [3]. The Vivid group's Mandala system has shown similar interactions in a rich variety of game worlds since the early 1990's [7]. The MIT Media Lab's ALIVE system in 1993 [5] was the first to allow explicit 3-D interaction with a virtual environment and artificial creatures. Krueger's more recent "Small Planet" [4] installation allowed a user to fly above a virtual planet with arm gestures. Sid Fel's IAMASCOPE [1] presented a simple yet captivating visual interaction between a users appearance and a video kaleidoscope. As been observed one major drawback with most of these systems is that they require strictly controlled backgrounds and/or lighting. They also generally do not allow multiple users to interact with a virtual screen simultaneously - preventing many collaborative applications. In this work we integrate robust 3-D perceptual interface with virtual gaming environments.

Approach: Based on our stereo vision system we have build a prototype that explores three different interaction models for perceptual interaction in virtual worlds. This study includes both navigations in virtual space as well as interactions with virtual objects. Our current design envisions the following interactions: (i) Direct interaction. In this models full normal body movements are used to navigate in the virtual world. The real world movement has a direct mapping to movement in the virtual world that is intuitive and transparent to the users (ii) Gesture driven interaction. This model uses gestures to control the virtual environment. Abstract gestures are combined simple body movement, such as pointing or the flying metaphor in Krueger's "Small Planet". This interaction has to be learned but we think that some users might find it more efficient. (iii) Location driven interaction. Motion semantics are given by a position-defined state, which are made apparent by physical objects (icons) in the real world. For example moving to different positions on the floor trigger different events and movements in virtual world.

Future Work: It is widely believed that computers will be easier to use if we can communicate with them in ways that are more similar to human-to-human interactions in the real world. But how to accomplish this natural

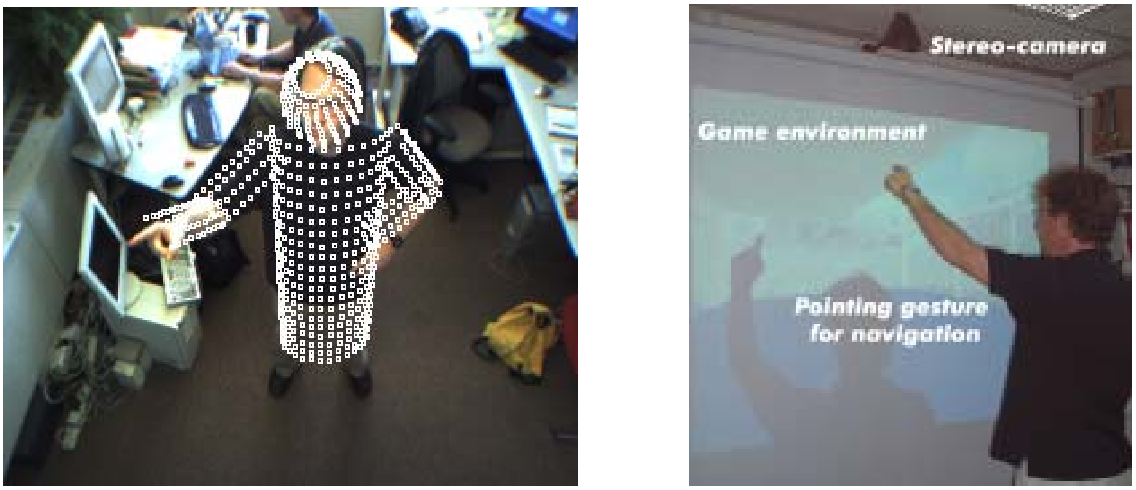

Figure 1: (left) Detection of body position and arm pose using stereo segmentation techniques. (right) A user interact with a virtual world using the gesture interaction model.

interaction is still largely undiscovered. Our initial idea was to use relative body motions to navigate around in the virtual world, e.g. by taking a half-step forward you will move forward, by leaning sidewise you will move sidewise. This was not as easy to use as we had hoped. One common problem was that users became disoriented and drifted from the central point since the use of relative movement was in conflict with the absolute movement that took place in the virtual world. From this we concluded that better feedback was needed. The other interaction models tended to provide a more robust control. However the lack of affordance and feedback, since the controls are invisible and the semantics are supposed to be natural, causes problems for the inexperienced users but is more efficient for the power user. This could be read as contradicting results from the original assumption that ”natural interfaces” should enable control for all kinds of users and the especially form novel users. We have also observed that ”natural interfaces” clearly increase the demands of technical performance of the perception system since users are less accepting to failures and expect performance more equal to human interpretation. We are now performing further studies where we test the relation between body movement, gestures and feedback mechanisms with different kinds of virtual worlds. Our long term goal is to integrate our multi person tracking system to explore multi-person interaction in virtual environments.

References:

[1] S. Fels, D. Reiners, and K. Mase. Iamascope: An interactive kaleidoscope. *Visual Proceedings of SIGGRAPH*, pages 76–77, 1997.

[2] G. Gordon, T. Darrell, M. Harville, and J. Woodfill. Background estimation and removal based on range and color. *Proceedings of CVPR99*, pages 459–464, 1999.

[3] M. Kreuger. *Artificial Realities II*. Addison-Wesley Publishing Co., 1991.

[4] M. Kreuger. Small planet. *Visual Proceedings of SIGGRAPH*, page 141, 1993.

[5] P. Maes, T. Darrell, B. Blumberg, and A. Pentland. The alive system: Wireless, full-body interaction with autonomous agents. *ACM Multimedia Systems*, 1996.

[6] Valve Software. The half-life game. *http://www.valvesoftware.com/*, 1996.

[7] J. Vincent and F. MacDougall. The mandala system. *CHI '90 Interactive experience*, 1990.