Gesture + Play

Konrad Tollmar, David Demirdjian and Trevor Darrell
Artificial Intelligence Lab – MIT
200 Technology Square, Cambridge MA02139, USA
konrad@ai.mit.edu

ABSTRACT
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. In this demo we present an ongoing study of perceptual interface abstractions that compares side-by-side the performance of several different interface regimes for exploring virtual game worlds. Our demonstration platform serves as an experimental test bed to collect data for our ongoing research, and as an example of state-of-the-art perceptual interfaces.

Keywords
gaming, virtual environments, perceptual interface

INTRODUCTION
Perceptual interfaces use machine perception to allow users to interact with computers. This demonstration explores possible interaction models for perceptual interfaces in games and virtual environments. In our case estimates of body position and pose are the raw measurements of our perceptual interface. However these estimates must be transformed into abstractions of gesture, motion, and action and mapped to controls that are useful for an application. The mapping of recognized tokens to application controls depends on both the application and the structure of the physical space. This study invites comparison of 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?

As a part of our work we are developing 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. We have developed software that can 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 multiple users as they move in an arbitrary workspace (see figure 1). Gestures, actions, and motions in this space cause motion and action in a virtual workspace, e.g. in a game or avatar-world.

For this study we are focusing primarily on game environments, for the playful and evocative nature of their visual content and the variety of possible perceptual interaction styles. The virtual world is created using the HALF-LIFE game engine from Sierra/Valve [refx], which enabled us to create rich environments with texture mapping and animations that also contain active objects with behavior.

Our installation is physically realized as a simple video projection wall, a stereo-camera and an ample open space in front (see figure 2).

BACKGROUND
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 [6]. 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.

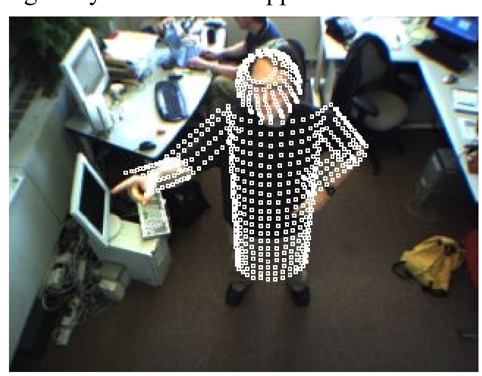

Figure 1: Detection of body position and arm pose using stereo segmentation techniques.

In this work we integrate robust 3-D perceptual interface with virtual gaming environments.

INTERACTION MODELS

Based on our stereo vision system we have build a prototype that explore 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:

- *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.

- *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.

- *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.

Figure 2: A user interact with a virtual world using the gesture interaction model.

DISCUSSION

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 interaction is still largely undiscovered.

Our system is still very much work in progress, and the main point in this demo is to present our approach in our study. However we have observed some early experiences.

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 a contradictive 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 (e.g. outdoor vers indoor). For example, we are experimenting whether physical modifications in the real world (e.g. non-uniform-height floor surface and physical objects icons) could enable a richer interaction. Our next step will then be to integrate our multi person tracking system to explore multi person interaction in virtual environments.

REFERENCES

1. Fels S. S., Reiners D., and Mase K.. Iamascope: An interactive kaleidoscope. Visual Proceedings of SIGGRAPH '97: The Electric Garden, pages 76--77, 1997.
2. Gordon, G., Darrell, T., Harville, M., Woodfill, J., Background Estimation and Removal Based on Range and Color, *Proceedings of CVPR99*, Ft. Collins CO, pp II:459-464, 1999.
3. Krueger, M., Artificial Realities II, Addison-Wesley Publishing Co., Reading, Massachusetts, 1991.
4. Krueger M., Small Planet, Visual Proceedings of SIGGRAPH'93, Machine Culture, ACM SIGGRAPH, pp. 141, 1993.
5. Maes P., Darrell T., Blumberg B. and Pentland A., The ALIVE System: Wireless, Full-Body Interaction with Autonomous Agents, ACM Multimedia Systems, Special Issue on Multimedia and Multisensory Virtual Worlds, ACM Press, Spring 1996.
6. Half-Life, http://www.valvesoftware.com/ , 1996.
7. Vincent, J. V., MacDougall, F. The Mandala System, *CHI* '90 *Interactive experience*, April 1-5 1990. Seattle, Washington