

## Interaction Techniques with Virtual Humans in Mixed Environments

Selim Balcisoy, Marcelo Kallmann, Rémy Torre, Pascal Fua, Daniel Thalmann  
Computer Graphics Laboratory, Swiss Federal Institute of Technology Lausanne  
ssbalcis, kallmann, rtorre, fua, thalmann@lig.di.epfl.ch

### Abstract

*Homogenous Virtual Environments evolved into Mixed Environments thanks to recent Augmented Reality techniques. We have done research on novel interaction techniques in virtual environments exploiting the usage of 3D virtual space. We will briefly present our Augmented Reality set-up for mixed reality experiments and we will analyze two distinct mixed environment case studies, and present two new interaction techniques where we use virtual humans as mediators between the real and virtual world.*

Keywords: Interaction Techniques, Mixed Environments, Augmented Reality, Virtual Humans.

### 1 Introduction

Most of the obstacles making the long awaited switch from a 2D computing environment to a 3D are diminishing. Fast workstations with high-end graphics cards are affordable, animation techniques are getting more complex and realistic, new input devices are emerging. However users are still limited to their mouse, keyboard and screen for computerized work and entertainment. In this paper we present two new interaction techniques for entertainment and work using mixed environments populated with virtual humans.

Recent developments in VR and human animation have led to the integration of Virtual Humans [9] into interactive 3D environments. With emerge of fast workstations allow us animate them in real-time for VR and Interactive Television applications. Virtual Humans can be guided or autonomous. Autonomous virtual humans can respond to perceived information and act on their own using their behavioral mechanisms. They can be used for crowd simulation, behavioral experiments or playing interactive games. Guided virtual humans are used as avatars,

representing a user in a virtual environment. Guidance of virtual humans is performed using 3D input devices like magnetic trackers for motion capture, spaceball or 2D graphical user interfaces.

Interaction with computers is dominated by the WIMP (Windows, Icons, Menus, Pointer) and desktop metaphor, which are originally developed for 2D environments to match the computing capabilities available in late 1970 and early 1980 [1]. These interaction techniques are well understood and established in application development. These successful interaction metaphors for 2D, like WIMP and desktop, are still used widely for commercial virtual environment applications with some modifications. These commercial applications are mostly used in entertainment industry, and usage of virtual environments or VR did not make expected rapid high usability in other areas, such as: manufacturing or medical. One of the reasons for this is that we do not know enough about the nature of user interaction in VEs to create systems, which allow people to do real work [2].

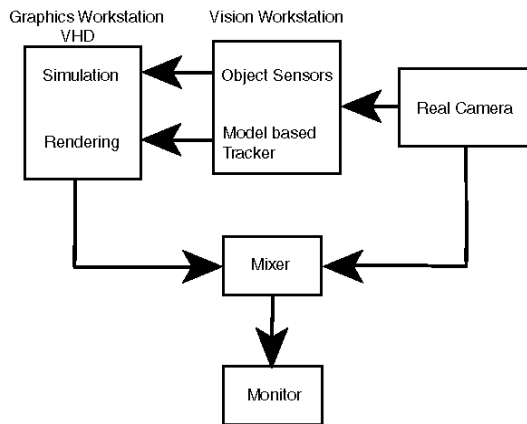
Research on user interfaces for 3D interactive virtual environments (VE) addresses developing interactions with a new medium. In this paper we present two interaction techniques using mixed environments with virtual humans:

- Direct manipulation of objects to interact in a mixed environment.
- Employing a virtual human as avatar to interact in a mixed environment.

We developed an Augmented Reality [10], [11] set-up to demonstrate our ideas with two case studies.

### 2 Experimental Augmented Reality Set-up

Our experimental Augmented Reality Set-up contains rendering, environment and virtual human simulation, mixing and computer vision sensor modules. The Figure 1 presents an overview of our system.



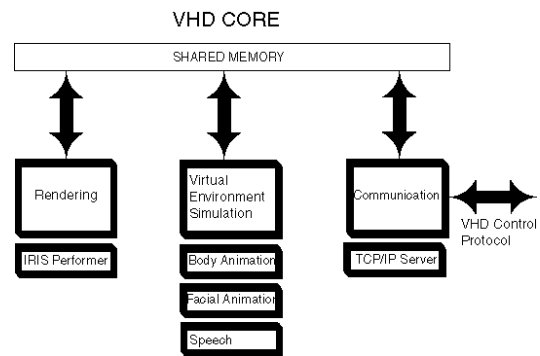
**Figure 1: Overview of experimental Augmented Reality set-up**

Our system gets images from real environment, and processes it at a dedicated workstation to perform vision based tracking and detection algorithms. Results from these algorithms are sent to a graphics workstation, where a virtual environment simulation is running. This simulation controls virtual humans and objects in the VE and renders a synthetic image. This image is sent to a video mixer and blended into filmed image. The result, an illusion of a mixed environment is displayed on a large monitor.

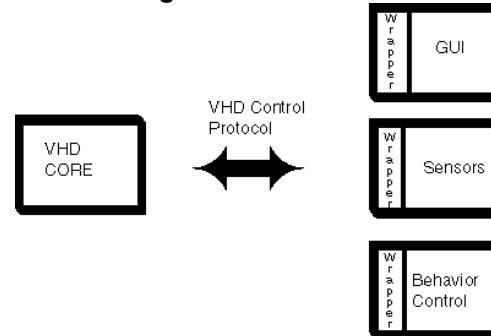
On the graphics workstation, a SGI Onyx2 with Infinite Reality Graphics board, we run an integrated real-time virtual environment simulation application, Virtual Human Director (VHD), which provides:

- Virtual human simulation: Speech, facial and body animation with motion generators and keyframed sequences, based on inverse kinematics and motion blending algorithms.
- Rendering : IRIX Performer, OpenGL
- Connections to external applications: External applications like AI controllers, GUIs can connect to VHD kernel over a TCP/IP port. They can send and retrieve simulation data, and control virtual humans, objects and camera parameters, using a communication protocol called virtual human control protocol.

VHD core is a multiprocessor application, and processes are communication with each other over a shared memory data structure. The Figure 2 depicts the VHD kernel and Figure 3 presents usage of virtual human control protocol.



**Figure 2:VHD Core.**



**Figure 3: Connections with external applications.**

In our experiments we used three external applications connecting to VHD. The first one is the Graphical User Interface (GUI), which is a common GUI to control virtual humans and camera with menus in a 2D fashion. A GUI is necessary for overall simulation control.

Our second extension is vision-based sensors. We developed a monocular real-time tracker based on the concept of Model-Based Optimization [5] [12]. A parametric model of a feature, such as a box, is extracted from one or more images by automatically adjusting the model's state variables until a minimum value of an objective function is obtained. The optimization procedure yields a description that simultaneously satisfies or nearly satisfies all constraints, and as result, is likely to be a good model of the feature. We also developed a detector, which visually scans a 2D array of squares and detects changes of luminance values on those squares.

Autonomous virtual humans are controlled by external Artificial Intelligence applications. In our experiments we used two types of such applications. One is for game control, and one for behavior modeling. The first application generates the next game move given the current situation. It is freeware software. The latter requires a Python language interface. This interface lets Python scripts generated by behavior modeling applications to perform in VHD. Both applications are presented more in detail in the case studies.

To create the illusion of a mixed environment, we blend the real video stream into the rendered one using chroma keying technique utilizing an over the shelf video mixer. To combine real and rendered scenes, we need to project synthetic models (virtual humans, objects) at the right location in real images. This requires a calibration of the camera used to film the scene, that is to compute the internal and external camera parameters that define the projection matrix [6]. In our setup we calibrate the camera, and we use our model-based tracker to register an object of interest correctly. We also use this tracker to track the camera motion in real-time.

### 3 Interaction Techniques in Mixed Environments

Interaction is defined as mutual or reciprocal action or influence. In real world everything and everybody interacts with everybody else. The most basic form of interaction is the physical one, where objects follow Newtonian mechanics for collision and support. The physical interaction changes the physical properties of an object, like shape, size, position and material properties.

Several methods have been developed for interactions inside a homogenous Virtual Environment, and between an application and its user. Several research groups [3], [8], [13], [14], [15] worked on novel interaction techniques and tools for virtual environments:

- Adaptation of 2D techniques in 3D,
- Usage of new input devices and trackers,
- Direct manipulation, Physical mnemonics, Gestural actions

These techniques have some limitations:

- No reference to actual working habits in a real environment,
- Lack of haptic input and feedback,
- Cumbersome input devices.

There is clear need for a natural way of interacting with and designing of a virtual environment. A mixed environment will allow users to be in a real environment and perceive the world as it is. Studies have shown that using some type of real environment (objects, room) helps them to accomplish real work tasks. On the other hand for entertainment purposes, there is an interest for mixed virtual sets from production companies. A mixed virtual set will allow designers to create a mixed environment, where some object can be real or virtual depending the visual impact and more important rendering power. Every rendering system has limited polygon budget, and it is more interesting to spend it on some virtual characters

than a complete synthetic background for interactive productions. Using an image based technique for creating an environment solves this problem.

There will be audio-visual feedback from the simulation in form of a see-through HMD output, or a large monitor in a mixed environment. However both feedback technologies have their own limitations such as wires, weight, head tracking in case of a HMD, or lack of visual realism and precision in case of a large monitor. We need a reference point or in other words a mediator between real and virtual worlds.

Our approach is that using mixed environments and virtual humans as a mediator between a real environment and a virtual environment, we can achieve a natural way of interaction between human and machine. A virtual human can act very human like in terms of animation, ergonomics, and perform very precise handling of a virtual object at the same time.

We will define interaction with a virtual human as: triggering some meaningful reaction in the environment in response actions such as body, social gestures or verbal output from a participant.

#### 3.1 Direct manipulation of objects to interact in a mixed environment

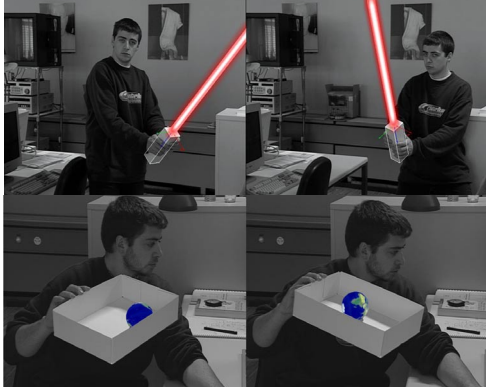
One of our goals is that participants in a mixed environment are interacting with the simulation not over a GUI or any kind of input device, but by becoming part of the environment. Their natural interaction with objects will change the state of simulation and trigger some meaningful response.

A good example for becoming part of the environment is a fight simulator application, where a user's actions are captured by a magnetic tracking system, and a posture recognition is performed [4]. In this system participant uses his own body to interact with the virtual environment. An autonomous virtual human responses according the result of a gesture/posture recognition algorithm.

To have a similar interactive application in a mixed environment without any cumbersome tracking device, requires a real-time non-invasive tracker for complete human body. Only a computer vision based tracking system would fit this specification, however current vision trackers are not up to full body tracking and giving similar results like a magnetic tracker. To overcome this limitation we decide to track simpler objects than a human, but still having a large range of interaction possibilities. We

propose an approach, where tracked objects have semantic values known to real humans and to the mixed environment simulation.

We made several experiments to test the robustness of our algorithm before using it in a large-scale application. The Figure 4 presents several snapshots of tracking experiments.



**Figure 4: Snapshots from tracking experiments.**

By tracking an object where the user has a priori knowledge of how to operate with it, provides a natural interaction with the environment. In the second experiment, having a dynamic ball in a box, users already know/expect what will happen when they move the box in 3D. When the ball starts to roll upwards or downwards in a natural way, users will accept this immediately, as a real ball would have done similar movements. On the other hand the first experiment it is more entertainment than physical simulation, requires knowledge of popular culture. We tested our experiments with our laboratory members and visitors, mainly computer scientists. Users were exited to play around with the box and ball demo more than the light saber demo, probably due to realistic dynamic ball simulation.

These experiments showed us that a meaningful interaction with a mixed environment without cumbersome trackers is possible. As a large scale experiment to developed a checkers game simulator, where a real person plays checkers against a virtual opponent.

### 3.2 Case Study I: Checkers Game

In checkers game, like any other board game, all the interactions are happening on or very close to the game board. Additionally several objects on the board are moved by participants. A minimalist approach would be to let user play with his pieces and virtual pieces would change their places automatically similar to chess simulators from early 1980s. However we are accustomed to play against a real opponent in

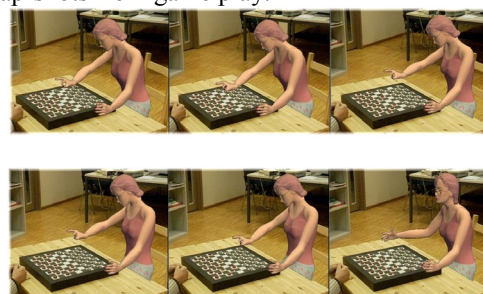
real life. A virtual human acting as an opponent will enrich the experience of playing the game, making it more realistic in several ways:

- Perception of time: Even if the checkers simulator would produce instant results, a virtual human will raise her hand, take a piece and make her move. The time needed for a move is analog for a real human, without considering time for thinking.
- Perception of 3D space: In a similar setup to our mixed reality setup, virtual pieces would change their places by themselves, in case of a virtual human, we can track her movements and have a longer time to follow which piece is played from where to where.
- Perception of opponent: A virtual human can be programmed or guided to speak and perform facial and body animation according to a situation.

The game play is similar to a real game, with only major difference where virtual player can not move real pieces when they are eliminated from game, this has to be done by the real player.

To develop our checkers game simulator we need to track real pieces and to place correctly a virtual human and virtual pieces we need to track the board. Tracking the board is performed by our model-based tracker, the result is used to track the camera movements.

The AR system uses a calibrated camera for correct registration of the board, the board coordinates are used to let the virtual camera follow the real one over the game play. Virtual human is controlled by a text based checkers game simulator, which calculates the next move for the virtual human. The Figure 5 presents snap-shots from game play.



**Figure 5 Checkers play between real and virtual human.**

This experiment shows that using available computer vision algorithms and virtual humans in a mixed environment, we can develop a natural interaction technique. In this case users manipulate real objects to trigger some meaningful actions in a mixed environment.

Next example will present another interaction technique, where semi-autonomous virtual humans manipulate virtual objects and demonstrate some meaningful results in a mixed environment.

### **3.3 Employing a virtual human as avatar to interact with a mixed environment**

Performing precise operations in a virtual environment is a difficult task. The input devices are not precise enough, there is lack of haptic feedback and direct modification requires complex interaction techniques like auto scaling. To analyze the problem, and develop a new interaction technique we can think of an imaginary product development scenario, where a CD-Driver has to be added to a workstation tower.

### **3.4 Case Study II: rapid prototyping in mixed environments**

We assume that we have the workstation tower already produced and we have a rough 3D model of it. Our designers have made 3D model of a CD-Driver, which they think suitable for this tower design. Now we have to place this CD-Driver at an appropriate position in the tower and figure out if this design is acceptable in a real work environment.

To find the position there are several solutions, one is to manufacture one CD-Driver prototype and test it on the tower, place it in a real office, let a test group play around with it, but manufacturing costs and time budget will be serious limitations. And if we don't like the design of the CD-Driver, it would cost more to make a new prototype.

Another solution is to work with a CAD application, where we place the CD-Driver in the tower using a GUI. We can work very accurate and try many possible positions. Making other tests, like letting a user group test the solution in a realistic environment, is not easily possible inside a CAD application. And if we decide to use a virtual reality setup, to let users place the CD-Driver in the tower. In this case users are expected to wear a HMD for visual feedback and a glove with sensors, to track the hand movements. And still placement will be difficult and not precise enough.

There are two important tests to be performed. The first one is the position and orientation of the CD-Driver relative to the tower. The second one to determine the impact of different placements and different work environments on users. We propose to use virtual humans in a

mixed environment to make these tests to overcome some if not all of the limitations of other solutions. We propose that a virtual human places the CD-Driver at a predefined position first, pushes the open button puts a CD in it, closes it. Performs the same for another position, until a suitable position is found.

Employing a virtual human for rapid prototyping has the following advantages:

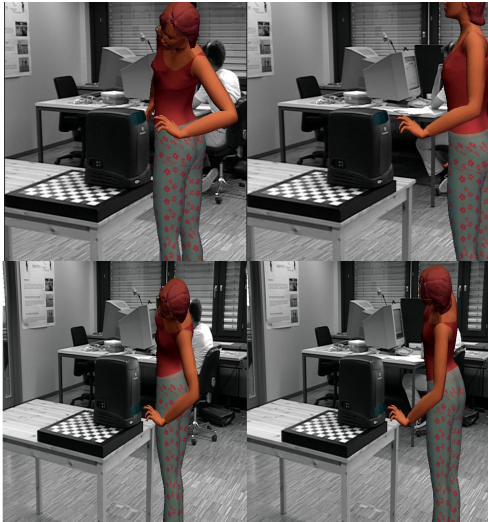
- **Precise Positioning:** Like in a CAD application, multiple positions can be defined before the experiment, and virtual human can test a set of possible positions.
- **High visual realism:** We use a real background and position a virtual object on a real one. We can put the real object in different locations and positions. The virtual human will perform the tests like a real human. For example if we position the O2 on the ground, she will bend.
- **Fast testing:** Such a mixed environment will produce multiple results and will reproduce them at any given time.

We predefine several positions of the CD-Driver on the tower using a specific feature modeler, where the behaviors of existing parts in the model, and the behaviors of the new parts to evaluate are defined. In our case we define several positions of the CD-Driver and its movement for opening and closing. This modeler is called SOMOD, from Smart Object Modeler. This modeler is described in [7] in detail. The behaviors are recorded as Python scripts.

We developed a Python interpreter for our experimental AR setup, which allows pre-recorded behavioral scripts to operate within the mixed environment. The data transfer between VHD and Python interpreter uses the virtual human transfer protocol.

We place the workstation tower, a SGI O2 on a table and register its position using the model-based tracker. The CD-Driver is placed at a given place in the beginning. In our experiment we designed three different positions for the CD-Driver. A virtual human tests the first one like, reaching to the CD-Driver, pushing its button, it opens, taking the CD out of it, or putting a CD in it, pushing the button again, closing the driver. Then the virtual human takes the driver and puts it into its next position, and performs tests like the first one. .

The Figure 6 presents two snapshots from tests. Note that in the upper image couple the driver is in a high position, and in the lower ones in a low one, the position of the virtual humans is also different in the left images than the right ones.



**Figure 6: A virtual human performing position tests for a virtual CD-Driver, on a real workstation.**

## 4 Conclusions

In this paper we presented two possible interaction techniques with 3D environments. We used a mixed environment as a simulation environment and virtual humans as mediators between the real and virtual world.

The first interaction technique detects changes in a real environment, like positioning of an object, and triggers actions in the simulation. Note that these actions should be natural responses, otherwise users may get confused. For example, this would be the case if a virtual object started to move without any visible force being implied on it. To avoid such confusions we use virtual humans to operate on virtual object. Users see a realistic animated human figure performing operation on the virtual object.

We propose a second interaction technique, where users perform precise operations on real and virtual objects in a mixed environment using a semi-autonomous virtual human. In this way we can pre-define several operations like positions to reach relative to an object.

Human-machine interaction is an open research topic. With emerge of new technologies such as Virtual Reality and Augmented Reality, we are exploring new ways of interacting with computers in 3D environments. Like successful 2D interaction techniques, some 3D interaction techniques and metaphors will become successful, if only they become used for assisting some real work or creating new entertainment content.

## 5 References

- [1] Ken Hinckley, Rany Pausch, Dennis Proffitt and Neal F. Kassel, Two-Handed Virtual Manipulation ACM Transactions on Computer-Human Interaction, Vol. 5, No. 3, September 1998, Pages 260–302.
- [2] Robert W. Lindeman John L. Sibert James K. Hahn, Towards Usable VR: An Empirical Study of User Interfaces for Immersive Virtual Environments, CHI '99 Pittsburgh PA USA
- [3] Mark R. Mine, Frederick P., Brooks Jr., Carlo H. Sequin, Moving Objects in Space: Exploiting Proprioception In Virtual-Environment Interaction, Proceedings of SIGGRAPH 97, Computer Graphics Proceedings, Annual Conference Series, pp. 19-26 (August 1997, Los Angeles, California).
- [4] Emering L., et al.. "Interacting with Virtual Humans through Body Actions", IEEE Computer Graphics & Applications, Vol.18, No1, pp8-11, 1998
- [5] Fua. P. RADIUS: Image Understanding for Intelligence Imagery, chapter Model-Based Optimization: An Approach to Fast, Accurate, and Consistent Site Modelling from Imagery. Morgan Kaufmann, 1997. O. Firschein and T.M. Strat, Eds. Available as Tech Note 570, Artificial Intelligence Center, SRI International.
- [6] J.P. Tarel and J.M. Vezien. Camcal Manual: A Complete Software Solution for Camera Calibration. Technical Report 0196, INRIA, September 1996.
- [7] M. Kallmann and D. Thalmann, "Modeling Objects for Interaction Tasks", EGCAS'98 - 9th Eurographics Workshop on Animation and Simulation, Lisbon, Portugal, 73-86, 1998.
- [8] An Approach to Natural Gesture in Virtual Environments, Alan Wexelblat, ACM Transactions on Computer-Human Interaction, Vol. 2, No. 3, September 1995, Pages 179-200.
- [9] Badler, N. and Webber, B. L. 1993. Simulating Humans: Computer Graphics Animation and Control. Oxford University Press, New York.
- [10] R.T. Azuma. A Survey of Augmented Reality. Presence, Teleoperators and Virtual Environments, 6(4):355–385, August 1997.
- [11] G. Klinker, K. Ahlers, D. Breen, P.-Y. Chevalier, C. Crampton, D. Greer, D. Koller, A. Kramer, E. Rose, M. Tuceryan, and R. Whitaker. Confluence of Computer Vision and Interactive Graphics for Augmented Reality. Presence: Teleoperations and Virtual Environments, 6(4):433–451, 1997.
- [12] E. Marchand, P. Boutheymy, F. Chaumette, and V. Moreau. Robust real-time Visual Tracking Using a 2D-3D Model-Based Approach. In International Conference on Computer Vision, pages 262–268, Corfu, Greece, September 1999.
- [13] Z. Szalavari, E. Eckstein, and M. Gervautz. Collaborative Gaming in Augmented Reality. In ACM Symposium on Virtual Reality Software and Technology, pages 195–204, Taipei, Taiwan, November 1998.
- [14] M. Anabuki, H. Kakuta, H. Yamamoto, H. Tamura. Welbo: An embodied conversational agent living in mixed reality space. CHI 2000 pp. 10-11
- [15] J. Underkoffler, B. Ullmer, H. Ishii. Emancipated Pixels: real-world graphics in the luminous room. SIGGRAPH'99