Integration of Avatars and Autonomous Virtual Humans in Networked Virtual Environments

Tolga K. Capin[1], Igor Sunday Pandzic[2],
Nadia Magnenat Thalmann[2], Daniel Thalmann[1]

[1]*Computer Graphics Laboratory*
Swiss Federal Institute of Technology
CH1015 Lausanne, Switzerland
{capin,thalmann}@lig.di.epfl.ch
http://ligwww.epfl.ch

[2]*MIRALAB-CUI*
University of Geneva
24 rue de Général-Dufour
CH1211 Geneva 4, Switzerland
{Igor.Pandzic,Nadia.Thalmann}@cui.unige.ch
http://miralabwww.unige.ch

Abstract

In this paper, we survey problems and solutions for inserting virtual humans in Networked Virtual Environments. Using virtual humans as participant embodiment increases the collaboration in Networked Virtual Environments, as it provides a direct relationship between how we interact with the real world and the virtual world representation. We show the differences between avatars and autononous Virtual Humans in terms of motion control.

1 Introduction

Trends towards networked applications and Computer Supported Collaborative Work (CSCW), together with a wide interest for graphical systems and Virtual Environments, have in the recent years raised interest for research in the field of Networked Virtual Environments (NVEs) [1]. NVEs are systems that allow multiple geographically distant users to interact in a common virtual environment. The users themselves are represented within the environment using a graphical embodiment.

The user can evolve within the environment and interact with it. All events that have an impact on the environment are transmitted to other sites so that all environments can be updated and kept consistent, giving the impression for the users of being in the same, unique environment. The users become a part of the environment, embodied by a graphical representation that should ideally be human-like.

Networked Virtual Environment (NVE) systems are suitable for numerous collaborative applications ranging from games to medicine [2, 3], for example:

- virtual teleconferencing with multimedia object exchange,
- all sorts of collaborative work involving 3D design,
- multi-user game environments,
- teleshopping involving 3D models, images, sound (e.g. real estate, furniture, cars),
- medical applications (distance diagnostics, virtual surgery for training),
- distance learning/training,
- virtual studio/set with networked media integration,
- virtual travel agency.

Networked Virtual Environments (NVEs) have been an active area of research for several years now, and a number of working systems exist [4, 5, 6, 7, 8, 9]. They differ largely in networking solutions, number of users supported, interaction capabilities and application [10], but share the same basic principle. Although NVEs have been around as a topic of research for quite some time, in most of the existing systems the embodiments are fairly simple, ranging from primitive cube-like appearances [11], non-articulated human-like or cartoon-like avatars [12] to articulated body representations using rigid body segments [4, 5]. Ohya et al. [7] report the use of human representations with animated bodies and faces in a virtual teleconferencing application.

The realism in participant representation involves two elements: believable appearance and realistic movements. Realism becomes even more important in multi-user networked virtual environments (NVE), as participants' representation is used for communication. The local program of the participants typically store the whole or a subset of the scene description, and they use their own avatars to move around the scene and render from their own viewpoint. This avatar representation in NVEs has crucial functions in addition to those of single-user virtual environments:

- *perception (to see if anyone is around)*: the participants need to be able to tell at a glance who else is present in the same VE, and this should be done in a continuous manner. The realistic embodiment makes it easy to distinguish embodiments from other virtual objects.
- *localization (to see where the other person is):* the position and orientation of other participants can convey different meanings. Particularly, orientation of the embodiments may convey a special intention related to nonverbal communication.
- *identification (to recognize the person):* the embodiments make it easy to differentiate different participants in the NVE. Using this embodiment regularly, the participant has a bounded, authentic, and coherent representation in the virtual world. In addition, by changing decoration of the body through clothes and accessories, the representation has an emergent identity.
- *visualization of others' interest focus (to see where their attention is directed):* to understand where the other participants' attention concentrate, may be critical to supporting interaction. For example, for CSCW applications, it may make it easy to focus the discussion. Or, for nonverbal communication purposes, the gaze direction helps to control turn-taking in conversation, as well as modifying, strengthening or weakening of what is said verbally,
- *visualization of others' actions (to see what the other person is doing and what she means through gestures):* Action point corresponds to where in the virtual world a person is manipulating. This is crucial in applications where synchronous collaboration among participants is important (e.g. modifying different parts of an object).
- *social representation of self through decoration of the avatar (to know what the other participants' task or status is):* the decoration of the avatar can convey meanings which shape the interaction. This decoration can be constant, such as a uniform; or it can change from day to day or even within one day, such as accessories that the avatar wears.

In this paper, we survey problems and solutions for inserting virtual humans in NVEs. We have built an architecture, VLNET (Virtual Life NETwork) [9 13], and we tried to integrate artificial life techniques with virtual reality techniques in order to create truly virtual environments shared by real people, and with autonomous living virtual humans with their own behavior, who can perceive the environment and interact with participants. Figure 1 and Figure 2 show example applications of the system.

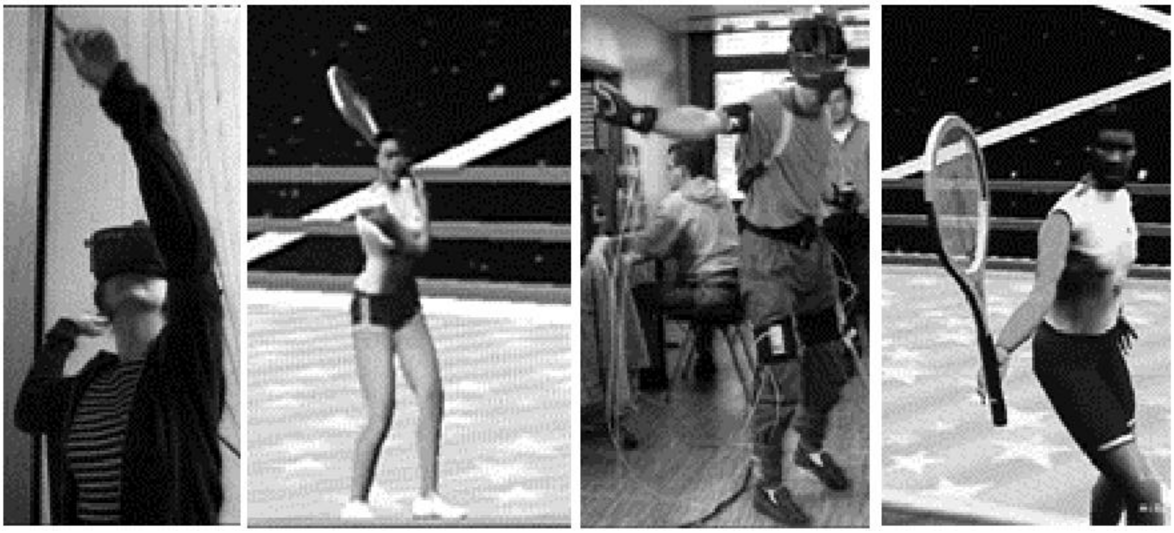

Figure 1. Networked Virtual Tennis

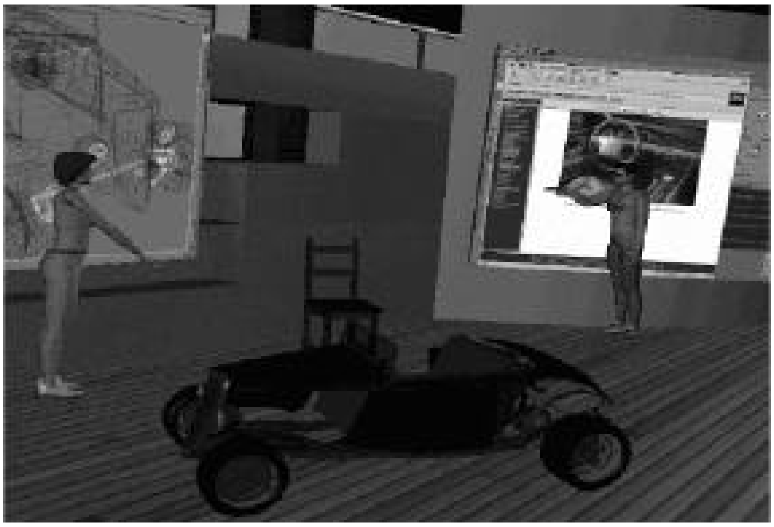

Figure 2. Virtual meeting

2 Basic of Virtual Humans

2.1 Articulated Structure for Virtual Human Modeling

For our Virtual Humans [14], we use the HUMANOID [15] articulated human body model with *75* degrees of freedom without the hands, with additional 30 degrees of freedom for each hand. The skeleton is represented by a 3D articulated hierarchy of joints, each with realistic maximum and minimum limits. The skeleton is encapsulated with geometrical, topological, and inertial characteristics of different body limbs. The body structure has a fixed topology template of joints, and different body instances are created by specifying 5 scaling parameters: global scaling, frontal scaling, high and low lateral scaling, and the spine origin ratio between the lower and upper body.

Attached to the skeleton, is a second layer that consists of blobs (metaballs) to represent muscle and skin [16]. The method's main advantage lies in permitting us to cover the entire human body with only a small number of blobs. The method is based on deformations of the body based on cross sections, and is summarized as follows: the body is divided into 17 parts : head, neck, upper torso, lower torso, hip, left and right upper arm, lower arm, hand, upper leg, lower leg, and foot. Because of their complexity, head, hands and feet are not represented by blobs, but instead with triangle meshes. For the other parts a cross-sectional table is used for deformation. This cross-sectional table is created only once for each body by dividing each body part into a number of cross-sections and computing the outermost intersection points with the blobs. These points represent the skin contour and are stored in the body description file. During runtime the skin contour is attached to the skeleton, and at each step is interpolated around the link depending on the joint angles. From this interpolated skin contour the deformation component creates the new body triangle mesh.

2.2 The Type of Virtual Humans

We can divide the virtual human control methods into three classes:

- *Directly controlled virtual humans*: the state vector of the virtual human is modified directly (e.g. using sensors attached to the body) by providing the new DOF values directly (e.g., by sensors attached to the body).
- *User-guided virtual humans*: the external driver *guides* the virtual human by defining tasks to perform, and the virtual human uses its motor skills to perform this action by coordinated joint movements (e.g. walk, sit).
- *Autonomous virtual humans*: the virtual human is assumed to have an internal state which is built by its goals and sensor information from the environment, and the participant modifies this state by defining high level motivations, and state changes (e.g. turning on vision behavior).

The control methods are not independent, higher-level controls require lower level capabilities. Autonomous behavior assumes motor skills to accomplish control, and motor skills modify individual DOFs.

In NVEs, the two first types of Virtual Humans are usually called Avatars.

2.3 Direct Control of Avatars

For this type of control, a complete representation of the participant's virtual body should have the same movements as the real participant body for more immersive interaction. This can be best achieved by using a large number of sensors to track every degree of freedom in the real body. Molet et al. [17] discuss that a minimum of 14 sensors are required to manage a biomechanically correct posture and Semwal et al. [18] present a closed-form algorithm to approximate the body using up to 10 sensors. However, many of the current VE systems use only head and hand tracking. Therefore, the limited tracking information should be connected with human model information and different motion generators in order to “extrapolate” the joints of the body which are not tracked. This is more than a simple inverse kinematics problem, because there are generally multiple solutions for the joint angles to reach to the same position, and the most realistic posture should be selected. In addition, the joint constraints should be considered for setting the joint angles. The main lowest-level approaches to this extrapolation problem are: inverse kinematics using constraints [19], closed form solutions [18], and table lookup solutions [20].

3 Guided Avatars

Guided virtual humans are avatars which are driven by the user but which do not correspond directly to the user motion. In VLNET, an example of virtual human guidance is guided navigation. We may consider three guided control methods: walking, picking , and bending.

3.1 Guided Control for Walking

The participant uses the input devices, typically a mouse or a SpaceBall, to update the transformation of the eye position of the virtual actor. This local control is used by computing the incremental change in the body position, and estimating the rotation and velocity of the center of body. The walking motor [21] uses the instantaneous velocity of motion, to compute the walking cycle length and time, by which it computes the necessary joint angles. This model can also include kinematical personification depending on the individuality of the user, and it is based on the mathematical parameterization coming from biomechanical experimental data. Figure 3 shows an example of walking motion in real-time. The model that we are using, supports frontal walking, and lateral stepping, running, turnaround and backwards stepping are not supported.

Figure 3. An example of real-time walking sequence

3.2 Guided Control for Picking

The arm motion for picking an object is a similar problem to walking: given 6 degrees of freedom (position and orientation) of the sensed right hand with respect to body coordinate system, the arm motor should compute the joint angles within the right arm. This is more difficult than simple inverse kinematics problem, because there are generally multiple solutions for the joint angles to reach to the same position, and the most realistic posture should be selected. In addition, the joint constraints should be considered for setting the joint angles. For real-time purposes, we exploit a three-link chain with 7 degrees of freedom within the right arm, with the right shoulder as the root.

3.3 Guided Control for Bending

Using right arm for manipulating objects has limitations: only the objects within reach can be picked; and it is impossible to reach objects lower, on the floor for example. For this purpose, for moving the hand lower, we have implemented a bending behavior driver. Bending also uses inverse kinematics as in arm motion, however this is a two-step process. First, the driver checks if the pick matrix is within the reach of the right arm (i.e. the distance to shoulder is less than the length of the arm); if this test is false, the spine of the body is adjusted using inverse kinematics so that the shoulder is within the distance less than the arm length. Then, in the second step, the inverse kinematics algorithm is applied in the right arm chain as described above to reach the end effector.

4 Autonomous Virtual Humans

4.1 What is autonomy ?

An autonomous system is a system that is able to give to itself its proper laws, its conduct, as opposed to a heteronomous system which is driven from the outside. Our autonomous virtual humans are able to have a behavior, which means they must have a manner of conducting themselves. Autonomous Virtual Humans (e.g. Figure 4) should be able to have a behavior, which means they must have a manner of conducting themselves. Typically, the Virtual Human should perceive the objects and the other Virtual Humans in the environment through virtual sensors: visual, tactile and auditory sensors. Based on the perceived information, the actor's behavioral mechanism will determine the actions he will perform. An actor may simply evolve in his environment or he may interact with this environment or even communicate with other actors. In this latter case, we will consider the actor as a interactive perceptive actor.

Figure 4. Autonomous Virtual Humans

Virtual vision [22] is a main information channel between the environment and the virtual actor. The Virtual Human perceives his environment from a small window in which the environment is rendered from his point of view. As he can access depth values of the pixels, the color of the pixels and his own position, he can locate visible objects in his 3D environment. To recreate the a virtual audition, in a first step, we have to model a sound environment where the Virtual Human can directly access to positional and semantic sound source information of a audible sound event. For virtual tactile sensors, it may be based on spherical multi-sensors attached to the articulated figure. A sensor is activated for any collision with other objects. These sensors could be integrated in a general methodology for automatic grasping.

Behaviors may be also dependent on the emotional state of the actor. A nonverbal communication is concerned with postures and their indications on what people are feeling. Postures are the means to communicate and are defined by a specific position of the arms and legs and angles of the body.

4.2 Interfacing Autonomous Virtual Humans with the NVE

The behavior of the autonomous virtual human is typically affected by the VE. The autonomous actor program needs to receive the state of the VE from the NVE system. The input can be through synthetic sensors, such as the transformation of objects in the VE, on the other hand it can also be natural language text interface.

The NVE system should be open, allowing developer of the autonomous actor program to choose whichever animation technique he wants: he should have minimum concern about the other details of the NVE, and the autonomous virtual human program should not depend on a specific approach. The behaviors should be easily extensible, and not be limited to predefined behaviors. As NVE systems are already very complex themselves, our approach to Autonomous Behaviors (AB) in NVEs is interfacing the two systems externally rather than trying to integrate them completely in a single, large system. Such an approach also facilitates the development of various AB systems to be used with a single NVE system, making it a testbed for various algorithms.

This interfacing leads to a kind of symbiosis between an NVE system and an AB system, where the AB system provides the brains and the NVE system the body to move around, be seen and act upon objects, but also to see, hear and get the external information to the brain. In order to implement this strategy, the NVE system must provide an external interface to which AB systems can be hooked. This interface should be as simple as possible to allow easy connection of various autonomous behaviors. At the same time it should satisfy the basic requirements for a successful symbiosis of a NVE system and an AB system: allow the AB system to control its embodiment and act upon the environment, as well as gather information from the environment upon which to act.

We may now identify the functionalities that the NVE system must provide to the AB system through the open interface:

- *Embodiment,* or a graphical representation, is a fundamental requirement to allow presence of the virtual actor in the environment. Though this can be a very simple graphical representation (e.g. a textured cube), some of the more advanced functionalities (e.g. facial and gestural communication) require a more human-like structure as support.
- *Locomotion* is necessary for getting from one place to another, and might involve generation of walking motion or simple sliding around. Essentially, the AB system must be able to control the position of the embodiment in the environment.
- *Capacity to act upon objects* in the environment is important because it allows the AB system to interact with the environment. The interface should provide at least the possibility to grab and move objects.
- Without some *feedback from the environment* our virtual actors might be autonomous, but blind and deaf. It is essential for them to be able to gather information from the environment about objects and other users.
- If the AB system is to simulate virtual humans, it is necessary for real humans to be able to communicate with them through our most common communication channel - the *verbal communication.*.
- It is desirable for a virtual actor to have the capacity of *facial and gestural communication* and therefore the ability to have a more natural behavior by showing emotions on the face or passing messages through gestures.

5 Conclusion

In this paper, we have shown the impact of Virtual humans in NVEs. We have explained the role of sophisticated avatars to represent the users. We also emphasized the importance of Autonomous Virtual Humans.

Acknowledgments

We are grateful to the assistants at LIG and MiraLAB for their contributions in the human models. The development was partly sponsored by the SPP program and the Federal Office for Education and Science in the framework of the European ACTS Project VPARK.

References

[1] Durlach NI, Mavor AS, eds. (1995) Virtual Reality: Scientific and Technological Challenges, Committee on Virtual Reality Research and Development, National Research Council, National Academy of Sciences Press, ISBN 0-309-05135-5.

[2] Doenges PK, Capin TK, Lavagetto F, Ostermann J, Pandzic IS, Petajan ED (1997) MPEG-4: Audio/Video & Synthetic Graphics/Audio for Mixed Media, Image Communication Journal, Special issue on MPEG-4.

[3] Zyda M, Sheehan J, eds. (1997) Modeling and Simulation: Linking Entertainment and Defense, ISBN 0-309-05842-2, National Academy Press.

[4] Barrus JW, Waters RC, Anderson DB (1996) Locales and Beacons: Efficient and Precise Support For Large Multi-User Virtual Environments, Proc. IEEE VRAIS, pp. 204-213.

[5] Carlsson C, Hagsand O (1993) DIVE - a Multi-User Virtual Reality System, Proc. IEEE VRAIS '93, Seattle, Washington, pp. 394-400.

[6] Macedonia et al. 1994 Macedonia MR, Zyda MJ, Pratt DR, Barham PT, Zestwitz (1994) NPSNET: A Network Software Architecture for Large-Scale Virtual Environments, Presence: Teleoperators and Virtual Environments, Vol. 3, No. 4, pp. 265-287.

[7] Ohya J, Kitamura Y, Kishino F, Terashima N (1995) Virtual Space Teleconferencing: Real-Time Reproduction of 3D Human Images, Journal of Visual Communication and Image Representation, Vol. 6, No. 1, pp. 1-25.

[8] Singh G, Serra L, Png W, Wong A, Ng H (1995) BrickNet: Sharing Object Behaviors on the Net, Proc. IEEE VRAIS '95, pp. 19-27.

[9] Pandzic IS, Capin TK, Magnenat Thalmann N, Thalmann D (1997) VLNET: A Body-Centered Networked Virtual Environment, Presence: Teleoperators and Virtual Environments, Vol. 6, No. 6, pp. 676-686.

[10] Macedonia MR, Zyda MJ (1997) A Taxonomy for Networked Virtual Environments, IEEE Multimedia, Vol. 4, No. 1, pp. 48-56.

[11] Greenhalgh C, Benford S (1995) MASSIVE, A Distributed Virtual Reality System Incorporating Spatial Trading, Proc. the 15th International Conference on Distributed Computing Systems, Los Alamitos, CA, ACM, pp. 27-34.

[12] Benford S, Bowers J, Fahlen LE, Greenhalgh C, Mariani J, Rodden T (1995) Networked Virtual Reality and Cooperative Work, Presence: Teleoperators and Virtual Environments, Vol. 4, No. 4, pp. 364-386

[13] Capin TK, Noser H, Thalmann D, IPandzic IS, Magnenat Thalmann N (1997), Virtual Human Representation and Communication in VLNET Networked Virtual Environment, IEEE Computer Graphics and Applications, Vol. 17, No. 2, pp. 42-53.

[14] Magnenat Thalmann N, Thalmann D (1995) Digital Actors for Interactive Television, Proc. IEEE, Special Issue on Digital Television, Part 2, July 1995, pp. 1022-1031.

[15] Boulic R, Capin T, Huang Z, Kalra P, Lintermann B, Magnenat-Thalmann N, Moccozet L, Molet T, Pandzic I, Saar K, Schmitt A, Shen J, Thalmann D (1995) The Humanoid Environment for Interactive Animation of Multiple Deformable Human Characters, Proc. Eurographics '95, pp. 337-348.

[16] Thalmann D, Shen J, Chauvineau E, Fast Human Body Deformations for Animation and VR Applications, Proc. Computer Graphics International 96, IEEE Computer Society Press, pp. 166-174.

[17] Molet T, Boulic R, Thalmann D (1996) A Real-Time Anatomical Converter for Human Motion Capture, Proc. Eurographics Workshop on Computer Animation and Simulation, R. Boulic ed., Springer, Wien, 1996, pp. 79-94.

[18] Semwal SK, Hightower R, Stansfield S (1996) Closed Form and Geometric Algorithms for Real-Time Control of an Avatar, Proc. VRAIS 96, pp. 177-184.

[19] Badler NI, Phillips CB, Webber BL (1993) Simulating Humans: Computer Graphics Animation and Control, Oxford University Press.

[20] Capin TK, Pandzic IS, Magnenat-Thalmann N, Thalmann D (1995) Virtual Humans for Representing Participants in Immersive Virtual Environments, Proc. FIVE '95, London.

[21] Boulic R, Magnenat Thalmann N M,Thalmann D (1990) A Global Human Walking Model with Real Time Kinematic Personification, The Visual Computer, Vol. 6, No. 6, pp. 344-358.

[22] Noser H, Renault O, Thalmann D, Magnenat Thalmann N (1995) Navigation for Digital Actors based on Synthetic Vision, Memory and Learning, Computers and Graphics, Pergamon Press, Vol. 19, No. 1, pp. 7-19.