WILEY

Avatars in Networked Virtual Environments

Tolga K. Çapin
Igor S. Pandzic
Nadia Magnenat-Thalmann
Daniel Thalmann

Avatars in Networked Virtual Environments

Tolga K. Çapin
Computer Graphics Lab, EPFL, Lausanne, Switzerland

Igor S. Pandzic
MIRALab, University of Geneva, Switzerland

Nadia Magnenat-Thalmann
MIRALab, University of Geneva, Switzerland

Daniel Thalmann
Computer Graphics Lab, EPFL, Lausanne, Switzerland

JOHN WILEY & SONS, LTD
Chichester · New York · Weinheim · Brisbane · Singapore · Toronto

Other Wiley Editorial Offices

John Wiley & Sons, Inc., 605 Third Avenue,
New York, NY 10158-0012, USA

WILEY-VCH Verlag GmbH, Pappelallee 3,
D-69469 Weinheim, Germany

Jacaranda Wiley Ltd, 33 Park Road, Milton,
Queensland 4064, Australia

John Wiley & Sons (Canada) Ltd, 22 Worcester Road,
Rexdale, Ontario M9W 1L1, Canada

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01,
Jin Xing Distripark, Singapore 129809

Library of Congress Cataloging-in-Publication Data

Avatars in networked virtual environments / Tolga K. Çapin...[et al.]
p. cm
Includes bibliographical references and index
ISBN 0-471-98863-4 (alk. paper)
1. Interactive multimedia. 2. Virtual reality. I. Çapin, Tolga K.
QA76. 76. 159A99 1999
006. 7—dc21 99-13561
CIP

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN 0-471-98863 -4

Produced from PostScript files supplied by the author.
Printed and bound in Great Britain by Bookcraft (Bath) Ltd
This book is printed on acid-free paper responsibly manufactured from sustainable forestry, in which at least two trees are planted for each one used for paper production.

Avatars in Networked Virtual Environments

Preface

Telepresence is the future of multimedia systems and will allow participants to share professional and private experiences, meetings, games, parties. Networked virtual environments (NVEs) is a key technology to implement this telepresence. NVEs are systems that allow multiple geographically distant users to interact in a common virtual environment. One of the particularly important research challenges in NVEs is the user representation or avatar, the way that participants are graphically represented in the VE. This can range from a very simple block-like representation to highly realistic virtual humans with articulated bodies and faces. We believe that real-time realistic 3D avatars will be essential in the future, as better user representation can improve a user's sense of presence in the environment and their ability to communicate with other users. We also need autonomous virtual humans to populate the virtual worlds.

The objective of this book is to explain the techniques for integrating virtual humans into virtual environments. The first chapter introduces the basic concepts of NVEs. Chapter 2 surveys different factors and design decisions while developing networked virtual environment architectures, and it surveys previous work based on this perspective. Chapter 3 introduces our solution: the Virtual Life Network (VLNET), a flexible framework for virtual humans in networked collaborative virtual environments. Chapter 4 describes the design philosophy and elements of VLNET. As VLNET is the most advanced NVE in terms of realistic virtual humans, we emphasize this key issue. We also discuss the motivations and challenges behind including virtual humans (VH) in the NVE systems. Based on the VLNET system, Chapter 5 describes the different means of facial communication we have developed. It also looks at gesture and non-verbal communication. Chapter 6 is dedicated to all problems of handling virtual human data across the network. Chapter 7 considers the potential relation of the future MPEG-4 standard to NVEs, based on experience from our active participation in the MPEG-4 ad hoc group on face and body animation. It also discusses the standardization of virtual humans in VRML. Chapter 8 presents several applications, including a virtual tennis game. There is also an extensive study of the experimental results and achievements.

The research in VLNET has been financed by the Swiss Priority Programme for Information and Communication Structures (SPP ICS) of the Swiss National Science Foundation. Our participation in the work of the MPEG committee is in the framework of the European ACTS projects VIDAS and COVEN.

We are also grateful to all people of MIRALab, University of Geneva, and LIG, EPFL, for their cooperation and support during this work and the development of VLNET.

Tolga K. Çapin
Igor S. Pandzic
Nadia Magnenat Thalmann
Daniel Thalmann

Contents

1 Concepts in Networked Virtual Environments

1.1 Virtual Environments

Ivan Sutherland introduced the concept of inserting people in computer-generated worlds in 1965, and made the first realization in 1968 with a tracked head-mounted stereoscopic display drawing wireframe models. Since then, together with the development of computer graphics knowledge and technology, various systems containing virtual environments have been developed.

Steve Ellis has presented an important introduction to virtual environment concepts (Ellis 1991). He defines *virtualization* as 'the process by which a human viewer interprets a patterned sensory impression to be an extended object in an environment other than that in which it physically exists'. He further categorizes virtualization into three levels: *virtual space*, *virtual image* and *virtual environment*. An example of the virtual space is a flat surface on which the image is rendered, and the observer can visualize three-dimensional objects through this space. Virtual image refers to perception of the object in depth, for example through stereoscopic images displayed on helmet-mounted displays. The third level, virtual environment, embodies the participant as part of the virtual world, so that his/her viewpoint and actions in this world correspond to those of the physical (real) world.

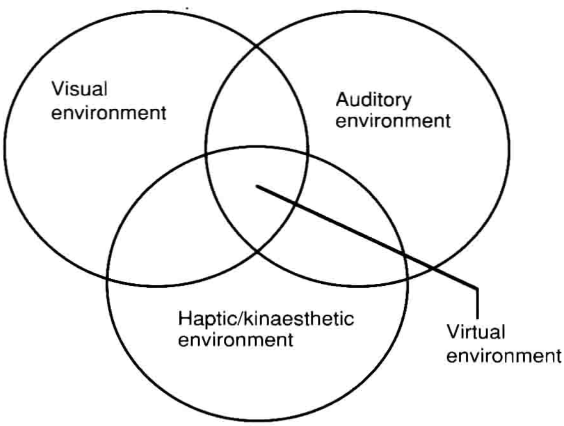

Figure 1-1. Elements of a virtual environment

In a virtual environment, the sensory impressions are delivered to the senses of the human participant through computer-generated displays. Ideally, a virtual environment has to provide three elements to immerse a participant: visual, auditory, haptics/kinesthetics. As shown in Figure 1-1, the virtual environment is an intersection of these elements. This intersection is important because it shows a fully interactive environment including all elements. However, this definition is general, and smaller virtual environments can be created by using subsets of the three components.

Kalawsky (1993) defined virtual environments as 'synthetic sensory experiences that communicate physical and abstract components to a human operator or participant. The synthetic sensory experience is generated by a computer system that one day may present an interface to the human sensory systems that is indistinguishable from the real physical world'.

Virtual environment refers to a technology which is capable of shifting a subject into a different environment without physically moving him/her. To this end, the inputs into the subject's sensory organs are manipulated in such a way that the perceived environment is associated with the desired virtual environment, not with the physical environment. The manipulation process is controlled by a computer model that is based on the physical description of the virtual environment. Consequently, the technology is able to create almost arbitrarily perceived environments.

There are other definitions of a virtual environment. However, most definitions agree that support of a totality of senses is necessary to make the participant feel present in the virtual surrounding. This book deals with virtual environments, where the participants are represented by 3D virtual human embodiments within the computer-generated world. This can be achieved by high-end configurations with magnetic trackers attached to the body, but also by desktop configurations where the participant sees their virtual embodiment interacting with the virtual objects.

1.2 Networked Virtual Environments

Until recently, the majority of virtual worlds have been single-user systems (Singh and Serra 1994). Compared with traditional systems, single-user systems provide tremendous benefit to the user but their utility in the real world is quite limited. This is because they do not support collaboration among a group of users. Networking coupled with highly interactive technology of virtual worlds will dominate the world of computers and information technology. It will not be enough to produce slick single-user, standalone virtual worlds. These systems will have to connect people, systems, information streams and technologies with one another. The information that is currently shared through file systems or through other 'static' media will have to be exchanged through the network. This information has to reside 'in the net' where it is easy to get at. Developing virtual worlds that support collaboration among a group of users is a complex and time-consuming task. In order to develop such virtual worlds, the developer has to be proficient in network programming, object management, graphics programming, device handling and user interface design. Even after gaining expertise in such diverse specializations, developing network-based virtual worlds takes a long time since network-based programs are inherently more difficult to write and debug than standalone programs.

Trends towards networked applications and computer supported collaborative work, 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) (Durlach and Mavor 1995). 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.

Figure 1-2 schematically presents the basic principle of the NVE. Each workstation has a copy of the virtual environment.

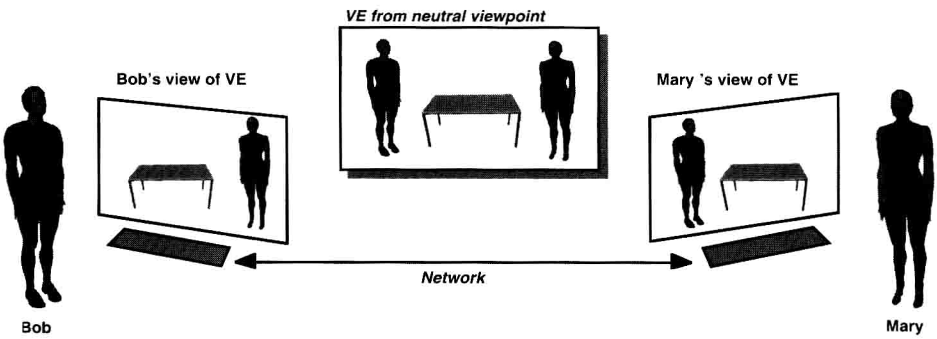

Figure 1-2. Principles of networked virtual environments

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 systems are suitable for numerous collaborative applications ranging from games to medicine (Doenges et al. 1997; Zyda and Sheehan 1997), 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 have been an active area of res ;ch for several years now, and a numb · of working systems exist (Barrus et al. 1996 ,arlsson and Hagsand 1993; Macedoni 1994; Ohya et al. 1995; Singh et al. 199 ,yda et al. 1993). They differ largely in net solutions, number of users supported, interaction capabilities and application scope (ia and Zyda 1997), but share the same basic principle.

Several aspects of NVE s 's have been subject to thorough research with interesting results: scalability and network opologies (Macedonia et al. 1994; Singh et al. 1995; Funkhouser 1996), efficient space structuring (Barrus et al. 1996; Benford et al. 1995), real-time simulation (Rohlf and Helman 1994), feeling of presence in NVEs (Benford et al. 1995; Welch et al. 1996, Hendrix and Barfield 1996; Tromp 1995).

There is also an increasing interest in shared graphical spaces that allo le t interact with each other in remote locations, or with programs and virtual o lere a some different perspectives:

- *Science fiction vision:* William Gibson's *Neuromancer* (1984), a popular science fiction novel, envisions the computer network as an immersive virtual space, wit people interacting with each other. Another science fiction novel is *Snow Crash* by Neal Stephenson (1992); this involves a computer-generated 3D large-scale space and people performing business with their virtual embodiments.
- *MUDs:* A MUD is defined as a networked, multi-participant, user-ext environment in which the user interface is entirely textual. Participants comm with each other and input commands for their actions (e.g. go east, smile) by te messages (Curtis 1992). Several MUDs involve a large number of participants, later systems such as Internet Relay Chat (IRC) have received a wider popularity, IRC, participants create channels for various subjects and distant participants can j the same channel for chatting. An improvement to these systems is Palace fr Palace Corporation, which is a multi-user virtual environment where the particip representations are 2D.
- *Successful virtual reality applications:* There have been an increasing number successful applications that make use of virtual reality. Most of these application benefit from sharing the experience with other people.

Until recently, networked graphics applications were prototype systems, demonstrating the effectiveness of the technology. However, a current effort is to provide real applications, manifested by the 3D graphics interchange standardization efforts such as VRML 2.0 and MPEG-4. The main contributors to these standards are companies hoping to diffuse their application content.

1.3 Virtual Environments for Interaction

Increasing hardware and network performance together with the software technology make it possible to define more complex interfaces for networked collaborative applications. Exploiting a virtual environment is an increasingly popular method as a natural interface for this purpose. An NVE can provide a more natural shared system, by supporting interactive human collaboration and integrating different media in real time in a single 3D surrounding. It provides a powerful system as it supports awareness of and interaction with other users; and it provides an appropriate mechanism for interaction with t[illegible] by supporting visual mechanisms for actions, data sharing and protection.

Providing a behavioural realism is a significant requirement for systems that are based on human collaboration, such as computer supported cooperative work (CSCW) systems. Networked CSCW systems also require that the shared environment supports an appropriate interface for gestural communication, awareness of other users in the environment, mechanisms for different modes of interaction (synchronous vs. asynchronous, the ability to work at different times in the same environment), mechanisms for customized tools for data visualization, protection and sharing.

Virtual environments can support powerful mechanisms for networked CSCW systems, as they provide a more natural interface to the environment. This can be accomplished in several ways:

- Representing the users and special-purpose service programs by 3D virtual humans in the virtual environment.
- Mechanisms for the participants to interact with each other in the natural interface via facial interaction and body gestures of their virtual embodiments.
- Mechanisms for the participants to interact with the rest of the virtual environment through complex and realistic behaviours such as walking for navigation, grasping, etc.
- User-customized tools for editing the picked objects, depending on the object type (e.g. images, free-form surfaces).

1.4 Virtual Humans in NVEs

Realism not only includes believable appearance and simulation of the virtual world, but also implies the natural representation of participants. This representation fulfils several functions:

- The visual embodiment of the user
- The means of interacting with the world
- The means of feeling various attributes of the world using the senses

The realism in participant representation involves two elements: believable appearance and realistic movements. Realism becomes even more important in multi-user networked virtual environments, as participants' representation is used for communication. The local program of the participants typically stores 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. In particular, orientation of the embodiments may convey a special intention related to non-verbal 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 concentrates, this may be critical to supporting interaction. For CSCW applications, it may make it easy to focus the discussion. Or, for non-verbal communication, the gaze direction helps to control turn-taking in conversation, as well as modifying, strengthening or weakening what is said verbally.
- *Visualization of others' actions (to see what the other person is doing and what they mean through gestures)*: an 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). Figure 1-3 shows the importance of individual and general views.
- *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.

Although networked virtual environments 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 (Greenhalgh and Benford 1995), non-articulated human-like or cartoon-like avatars (Benford et al. 1995) to articulated body representations using rigid body segments (Barrus et al. 1996; Carlsson and Hagsand 1993; Pratt et al. 1997). Ohya et al. (1995) report the use of human representations with animated bodies and faces in a virtual teleconferencing application, as described in Chapter 2.

Using virtual human figures for avatar representation fulfils these functionalities with realism, as it provides a direct relationship between how we control our avatar in the virtual world and how our avatar moves related to this control. Even with limited sensor information, a virtual human frame that reflects the activities of the user can be constructed in the virtual world. Slater and Usoh (1994) indicate that using a virtual body, even if simple, increases the sense of presence in the virtual world.