Challenges for the Research in Virtual Humans

Daniel Thalmann
Computer Graphics Lab, EPFL,
Lausanne, Switzerland
thalmann@lig.di.epfl.ch
http://ligwww.epfl.ch

Introduction

Virtual Humans imply many complex problems we have been solving for several years [1]. With the new developments of digital and interactive television [2] and multimedia products, there is also a need for systems that provide designers with the capability for embedding real-time simulated humans in games, multimedia titles and film animations. Virtual Humans have also a key role to play in these shared Virtual Environments and true interaction with them is a great challenge. Although a lot of research has been going on in the field of Networked Virtual Environments, most of the existing systems still use simple embodiments for the representation of participants in the environments. More complex virtual human embodiment increases the natural interaction within the environment. The users' more natural perception of each other (and of autonomous actors) increases their sense of being together, and thus the overall sense of shared presence in the environment.

Fig.1 Virtual Humans

But, the modelling of Virtual Humans is an immense challenge as it requires to solve many problems in various areas. Table 1 shows the various aspects of research in Virtual Human Technology. Each aspects will be detailed and the problems to solve will be identified.

Face and body representation
Avatar functions
Motion control
High-level behavior
Interaction with objects
Intercommunication
Interaction with user
Collaborative Virtual Environments
Crowds
Rendering

Table 1. Aspects of research in Virtual Humans

Face and body representation

Human modelling is the first step in creating Virtual Humans. For head, although t is possible to create them using an interactive sculpting tool, the best way is to reconstruct them from reality. Three methods have been used for this:

1) Reconstruction from 2D photos [3]
2) Reconstruction from a video sequence [4]
3) Construction based on the laser technology

The methods could be used for body modelling, but the main problem is still with the body deformations which has been addressed by many researchers, but is still not 100% solved.

Concerning facial expressions in Networked VEs, four methods are possible: video-texturing of the face, model-based coding of facial expressions, lip movement synthesis from speech and predefined expressions or animations. Believable facial emotions are still very hard to obtain.

Main problem to solve: realistic body and face construction and deformations

Avatar functions

The avatar representation fulfils several important functions:

1) the visual embodiment of the user
2) means of interaction with the world
3) means of sensing various attributes of the world

It becomes even more important in multi-user Networked Virtual Environments [5], as participants' representation is used for communication. This avatar representation in NVEs has crucial functions in addition to those of single-user virtual environments [6 7]:

1) perception (to see if anyone is around)
2) localisation (to see where the other person is)
3) identification (to recognise the person)
4) visualisation of others' interest focus (to see where the person's attention is directed)
5) visualisation of other's actions (to see what the other person is doing and what is meant through gestures)
6) social representation of self through decoration of the avatar (to know what the other participants' task or status is)

Using articulated models for avatar representation fulfils these functionalities with realism, as it provides the direct relationship between how we control our avatar in the virtual world and how our avatar moves related to this control, allowing the user to use his/her real world experience. We chose to use complex virtual human models aiming for a high level of realism, but articulated "cartoon-like" characters could also be well suited to express ideas and feelings through the nonverbal channel in a more symbolic or metaphoric way.

Main problem to solve: easy way of directing an avatar

Motion control

The main goal of computer animation is to synthesize the desired motion effect which is a mixing of natural phenomena, perception and imagination. The animator designs the object's dynamic behavior with his mental representation of causality. He/she imagines how it moves, gets out of shape or reacts when it is pushed, pressed, pulled, or twisted. So, the animation system has to provide the user with motion control tools able to translate his/her wishes from his/her own language.

In the context of Virtual Humans, a Motion Control Method (MCM) specifies how the Virtual Human is animated and may be characterized according to the type of information it privileged in animating this Virtual Human. For example, in a keyframe system for an articulated body, the privileged information to be manipulated is the angle. In a forward dynamics-based system, the privileged information is a set of forces and torques; of course, in solving the dynamic equations, joint angles are also obtained in this system, but we consider these as derived information. In fact, any MCM will eventually have to deal with geometric information (typically joint angles), but only geometric MCMs explicitly privilege this information at the level of animation control.

Many MCMs have been proposed: motion capture, keyframe, inverse kinematics, dynamics, walking models, grasping models, etc.. But, no method is perfect and only combination of blending of methods can provide good and flexible results.

Main problem to solve: flexible reuse, combination, and parameterisation of existing movements

High-level behavior

Autonomous Virtual Humans should be able to have a behaviour, 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 [8]: visual [9 10], tactile and auditory sensors. Based on the perceived information, the actor's behavioural 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.

A high level behavior uses in general sensorial input and special knowledge. A way of modeling behaviors is the use of an automata approach. Each actor has an internal state which can change each time step according to the currently active automata and its sensorial input. Abstraction mechanisms to simulate intelligent behaviours have been discussed in the AI (Artificial Intelligence) and AA (Autonomous Agents') literature. Several methods have been introduced to model learning processes, perceptions, actions, behaviours, etc, in order to build more intelligent and autonomous virtual agents.

Main problem to solve: development of very complex believable behaviors

Interaction with objects

The necessity to model interactions between an object and a virtual human agent (here after just referred to as an agent), appears in most applications of computer animation and simulation. Such applications encompass several domains, as for example: virtual autonomous agents living and working in virtual environments, human factors analysis, training, education, virtual prototyping, and simulation-based design. A good overview of such areas is presented by Badler [11]. An example of an application using agent-object interactions is presented by Johnson et al [12], whose purpose is to train equipment usage in a populated virtual environment.

Another interesting way is to model general agent-object interactions based on objects containing interaction information of various kinds: intrinsic object properties, information on how-to-interact with it, object behaviors, and also expected agent behaviors. The smart object approach, introduced by Kallmann and Thalmann [13 14] extends the idea of having a database of interaction information. For each object modeled, we include the functionality of its moving parts and detailed commands describing each desired interaction, by means of a dedicated script language. A feature modeling approach [15] is used to include all desired information in objects. A graphical interface program permits the user to interactively specify different features in the object, and save them as a script file.

Main problem to solve: make the Virtual Human learn how to interact with objects

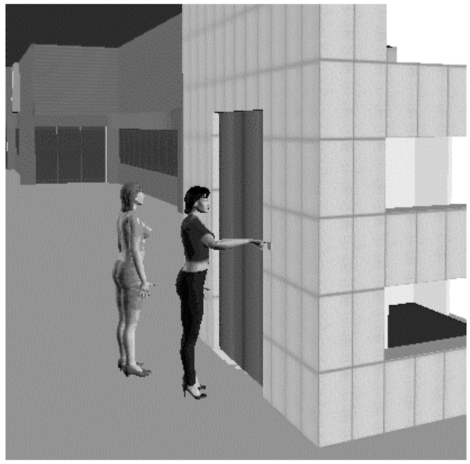

Fig.2. Interaction with objects

Intercommunication

Behaviours may be also dependent on the emotional state of the actor. A non-verbal 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. This non-verbal communication is essential to drive the interaction between people without contact or with contact.

What gives its real substance to face-to-face interaction in real life, beyond the speech, is the bodily activity of the interlocutors, the way they express their feelings or thoughts through the use of their body, facial expressions, tone of voice, etc. Some psychological researches have concluded that more than 65 percent of the information exchanged during a face-to-face interaction is expressed through nonverbal means [16]. A VR system that has the ambition to approach the fullness of real-world social interactions and to give to its participants the possibility to achieve a quality and realistic interpersonal communication has to address this point; and only realistic embodiment makes nonverbal communication possible.

Fig.3. Intercommunication

Interaction with user

The real people are of course easily aware of the actions of the Virtual Humans through VR tools like Head-mounted displays, but one major problem to solve is to make the virtual actors conscious of the behaviour of the real people. Virtual actors should sense the participants through their virtual sensors. Such a perceptive actor would be independent of each VR representation and he could in the same manner communicate with participants and other perceptive actors. Perceptive actors and participants may easily be. For virtual audition, we encounter the same problem as in virtual vision. The real time constraints in VR demand fast reaction to sound signals and fast recognition of the semantic it carries. For the interaction between virtual humans and real ones, gesture recognition will be a key issue. As an example, Boulic et al. [[17]] produced a fighting between a real person and an autonomous actor. The motion of the real person is captured using a Flock of Birds. The gestures are recognised by the system and the information is transmitted to the virtual actor who is able to react to the gestures and decide which attitude to do.

Main problem to solve: development of a complete real-time vision-based recognition

Specific problems of Networked Virtual Environments

Inserting virtual humans in the NVE is a complex task [5]. The main issues are:

1) selecting a scalable architecture to combine these two complex systems,
2) modeling the virtual human with believable appearance for interactive manipulation,
3) animating it with minimal number of sensors to have maximal behavioral realism,
4) investigating different methods to decrease the networking requirements for exchanging complex virtual human information.

Particularly, controlling the virtual human with limited input information is one of the main problems. For example, a person using a mouse will need extra input techniques or tools to exploit the functionalities of his embodiment. In this paper, we survey these tools that help a user with desktop VR configuration, we did not consider full tracking of the body using magnetic trackers, although this approach can be combined with limited tracking of the participant's arms.

Main problems to solve: controlling a realistic virtual human with limited input information, minimizing the information to be transmitted.

Crowds

An accepted definition of crowd is that of a large group of individuals in the same physical environment, sharing a common goal (e.g. people going to a rock show or a football match). The individuals in a crowd may act in a different way than when they are alone or in a small group [18]. Although sociologists are often interested in crowd effects arising from social conflicts or social problems [19] the normal behavior of a crowd can also be studied when no changes are expected.

Our goal [20] is to simulate the behavior of a collection of groups of autonomous virtual humans in a crowd. Each group has its general behavior [21] specified by the user, but the individual behaviors are created by a random process through the group behavior. This means that there is a trend shared by all individuals in the same group because they have a pre specified general behavior.

Main problem to solve: define collective behaviors while keeping individualities

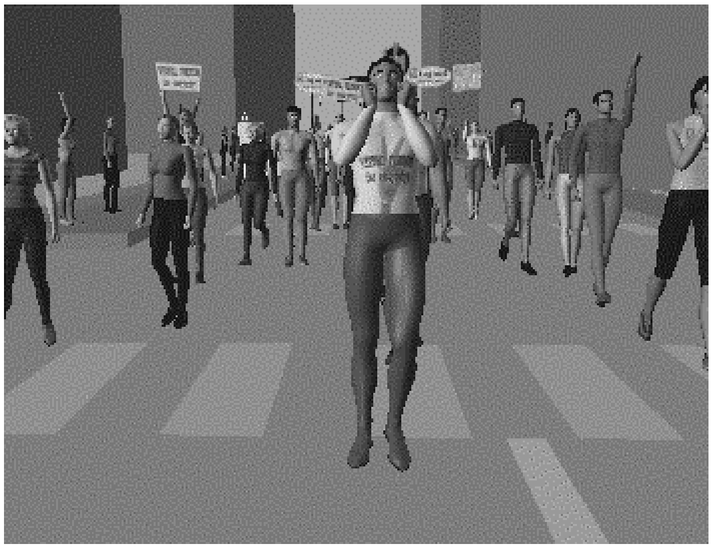

Fig.4. Crowds

Rendering

Rendering and animating in real-time a multitude of articulated characters presents a real challenge and few hardware systems are up to the task. Up to now little research has been conducted to tackle the issue of real-time rendering of numerous virtual humans. However, due to the growing interest in collaborative virtual environments the demand for numerous realistic avatars is becoming stronger.

There exist various techniques to speed up the rendering of a geometrical scene. They roughly fall into three categories: culling, geometric level-of-detail and image-based rendering which encompasses the concept of image caching. They all have in common the idea of reducing the complexity of the scene while retaining its visual characteristics.

Geometric level of detail (LOD) attempts to reduce the number of rendered polygons by using several representations of decreasing complexity of an object. At each frame the appropriate model or resolution is selected. Typically the selection criterion is the distance to the viewer although the object motion is also taken into account (motion LOD) in some cases. The major hindrance to using LOD is related to the problem of multi-resolution modeling, that is to say the automatic generation from a 3D object of simpler, coarser 3D representations that bear as strong a resemblance as possible to the original object.

Because 3D chips were not affordable or did not even exist in the 80s, video game characters, human-like or not, were then represented with 2D sprites. A sprite can be thought of as a block of pixels and a mask. The pixels give the color information of the final 2D image while the mask corresponds to a binary transparency channel. Using sprites, a human figure could easily be integrated into the decor. As more computing power was available in the 90s, the video game industry shifted towards 3D. However, the notion of sprites can also be used in the context of 3D rendering. Aubel and Thalmann [22] propose an approach to accelerated rendering of moving, articulated characters, which could easily be extended to any moving and/or self-deforming object. The method is based on impostors, a combination of traditional level-of-detail techniques and image-based rendering and relies on the principle of temporal coherence. It does not require special hardware (except texture mapping and Z-buffering capabilities, which are commonplace on high-end workstations nowadays) though fast texture paging and frame buffer texturing is desirable for optimal performance.

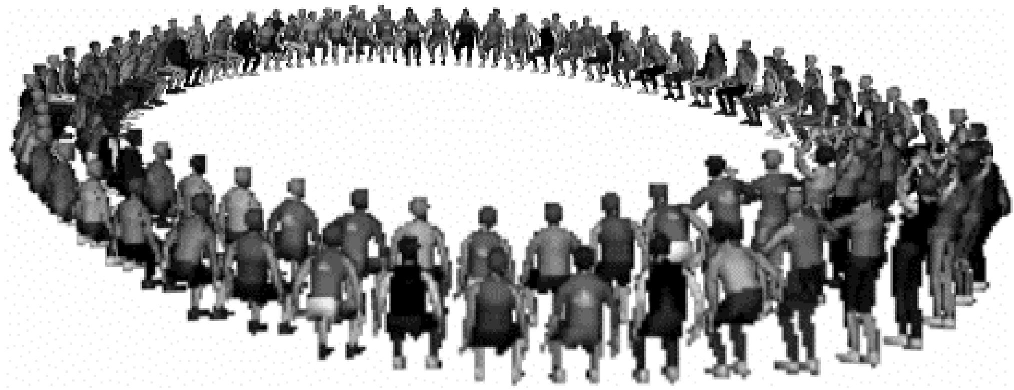

Fig.5. Use of Impostors

Main problems to solve: introduction of LOD for animation, integration with impostor technology

Acknowledgments

The author would like to thank people from LIG for their contributions, especially Soraia R. Musse, Marcelo Kallmann, Mireille Clavien

References

1. N. Magnenat Thalmann, D. Thalmann, Complex Models for Animating Synthetic Actors, IEEE Computer Graphics and Applications, Vol.11, No5, 1991, pp.32-44.
2. N. 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.
3. W.S.Lee, N.Magnenat-Thalmann, Head Modeling from Pictures and Morphing in 3D with Image Metamorphosis Based on triangulation, in: Modelling and motion Capture Techniques for Virtual Environments, Lecture Notes in Artificial Intelligence, 1537, Springer, 1998.

4. P. Fua, R. Plankers and D. Thalmann, From Synthesis to Analysis: Fitting Human Animation Models to Image Data, Proc. CGI 99, IEEE Computer Society Press, 1999
5. T. K. Capin, I.S. Panzic, N. Magnenat-Thalmann, D. Thalmann Avatars in Networked Virtual Environments, John Wiley and Sons, 1999.
6. T. K. Capin, I.S. Panzic, N. Magnenat-Thalmann, D. Thalmann, Virtual Human Representation and Communication in VLNET Networked Virtual Environment, IEEE Computer Graphics and Applications, March 1997.
7. S. D. Benford et al., « Embodiments, Avatars, Clones and Agents for Multi-user, Multi-sensory Virtual Worlds », Multimedia Systems, Berlin, Germany: Springer-Verlag, 1997.
8. D. Thalmann, Virtual Sensors: A Key Tool for the Artificial Life of Virtual Actors, Proc. Pacific Graphics '95, Seoul, Korea, August 1995, pp.22-40.
9. O. Renault, N. Magnenat Thalmann, D. Thalmann (1990) A Vision-based Approach to Behavioural Animation, The Journal of Visualization and Computer Animation, Vol 1, No 1, pp 18-21.
10. X. Tu, D. Terzopoulos (1994) Artificial Fishes: Physics, Locomotion, Perception, Behavior, Proc. SIGGRAPH '94, Computer Graphics, pp.42-48.
11. N. N. Badler, "Virtual Humans for Animation, Ergonomics, and Simulation", IEEE Workshop on Non-Rigid and Articulated Motion, Puerto Rico, June 97.
12. W. L. Johnson, and J. Rickel, "Steve: An Animated Pedagogical Agent for Procedural Training in Virtual Environments", Sigart Bulletin, ACM Press, vol. 8, number 1-4, 16-21, 1997.
13. M. Kallmann, D. Thalmann, Modeling Objects for Interaction Tasks, Proc. Eurographics Workshop on Animation and Simulation, Springer, 1998
14. M.Kallmann, D.Thalmann, A Behavioral Interface to Simulate Agent-Object Interactions in Real-Time, Proc. Computer Animation 99, IEEE Computer Society Press (to appear)
15. J.J. Shah, and M. Mäntylä, "Parametric and Feature-Based CAD/CAM", John Wiley & Sons, inc. 1995, ISBN 0-471-00214-3.
16. M. Argyle, Bodily Communication, New York: Methuen & Co., 1988.
17. L. Emering, R. Boulic, D. Thalmann, Interacting with Virtual Humans through Body Actions, IEEE Computer Graphics and Applications, 1998 , Vol.18, No1, pp.8-11.
18. M. E. Roloff. "Interpersonal Communication - The Social Exchange Approach". SAGE Publications, v.6, London. 1981.
19. J. S. McClelland. The Crowd and The Mob. Book printed in Great Britain at the University Press, Cambridge. 1989.
20. S.R. Musse, C. Babski, T. Capin, D. Thalmann, Crowd Modelling in Collaborative Virtual Environments, ACM VRST /98, Taiwan
21. S.R. Musse, D. Thalmann, A Model of Human Crowd Behavior: Group Inter-Relationship and Collision Detection Analysis. Proc. Workshop of Computer Animation and Simulation of Eurographics'97, Sept, 1997. Budapest, Hungary.
22. A.Aubel, R. Boulic, D.Thalmann, Animated Impostors for Real-time Display of Numerous Virtual Humans, Proc. Virtual Worlds 98, Paris, September 1998.