"Towards Intelligent Virtual Actors"

Daniel THALMANN
Computer Graphics Lab, EPFL
CH 1015 Lausanne Switzerland
thalmann@lig.di.epfl.ch

In this paper, we try to define the concept of virtual actors and show their evolution from clones and guided actors to autonomous, perceptive, and emotional actors. Examples of behaviors with limited intelligence are presented: knowledge-based navigation, locomotion, grasping and game playing.

keywords: *virtual actor, autonomous actor, perceptive actor, emotional actor, intelligent actor*

1. Introduction

Since a few years, the terms of virtual, synthetic or digital actors have been extensively used in order to describe very different kinds of entities. The term “actor” is well-known in the world of cinema and everybody knows it. A Hollywood-like actor is first a human personality. In the context of software, an actor as introduced by Hewitt [1] in 1971 is essentially an object that can send or receive messages. According to Greif and Hewitt [2], actors are defined by their behavior, whatever their shape: a cube that is able to jump by itself and avoid obstacles without intervention of the animator may be also defined as a virtual actor, because it is virtual and because it is able to behave itself.

Now, the advent of the hardware new technogies and the integration of techniques of Computer Animation, Artificial Intelligence, and Artificial Life allow to create truly synthetic or virtual actors that correspond to our two views of actors: Hollywood-like actors and Hewitt-like actors. This means actors which have the appearance of a cinema actor and which behaves as a true actor playing a role.

For the modeling of actor *behaviors*, the ultimate objective is to build *intelligent autonomous* virtual humans with *adaptation, perception* and *memory*. These virtual humans should be able to act *freely* and *emotionally*. They should be *conscious* and *unpredictable*. But can we expect in the near future to represent in the computer the concepts of behavior, intelligence, autonomy, adaptation, perception, memory, freedom, emotion, consciousness, and unpredictability ? First, we will try to define these terms. More details may be found in [3].

- **Behavior** for virtual humans may be defined as a manner of conducting themselves. it is also the response of an individual, group, or species to its environment.
- **Intelligence** may be defined as the ability to learn or understand or to deal with new or trying situations.
- **Autonomy** is generally defined as the quality or state of being self-governing.
- **Adaptation**: an artificial organism is adaptive as long as it may "survive" in more or less unpredictable and dangerous environments.
- **Perception** is defined as the awareness of the elements of environment through physical sensation.

- **Memory** is generally defined as the power or process of reproducing or recalling what has been learned and retained especially through associative mechanisms.
- **Emotion** may be defined as the affective aspect of consciousness; this is a state of feeling, a psychic and physical reaction subjectively experienced as strong feeling and physiologically involving changes that prepare the body for immediate vigorous.
- **Consciousness** may be defined as the quality or state of being aware especially of something within oneself or the state of being characterized by sensation, emotion, volition, and thought.
- **Freedom** for a virtual actor may be defined as the extent that his future behaviour is unpredictable to somebody.

In the next sections, we will introduce different types of virtual actors, starting from the most primitive ones to the most complex ones in terms of capability of behaviors.

2. Clones or Virtual actors as self representations of the user

The virtual actor is required to have a natural-looking body and be animated correlated to the actual body. It corresponds to a real-time form of traditional rotoscopy. Traditional rotoscopy in animation consists of recording the motion by a specific device for each frame and using this information to generate the image by computer. Using the terminology we introduced previously [4], we call *real-time rotoscopy method* a method consisting of recording input data from a VR device in real-time allowing to apply at the same time the same data to a graphics object on the screen. As shown in Figure 1, a popular way of animating such an actor is the use of sensors like the Flock of Birds.

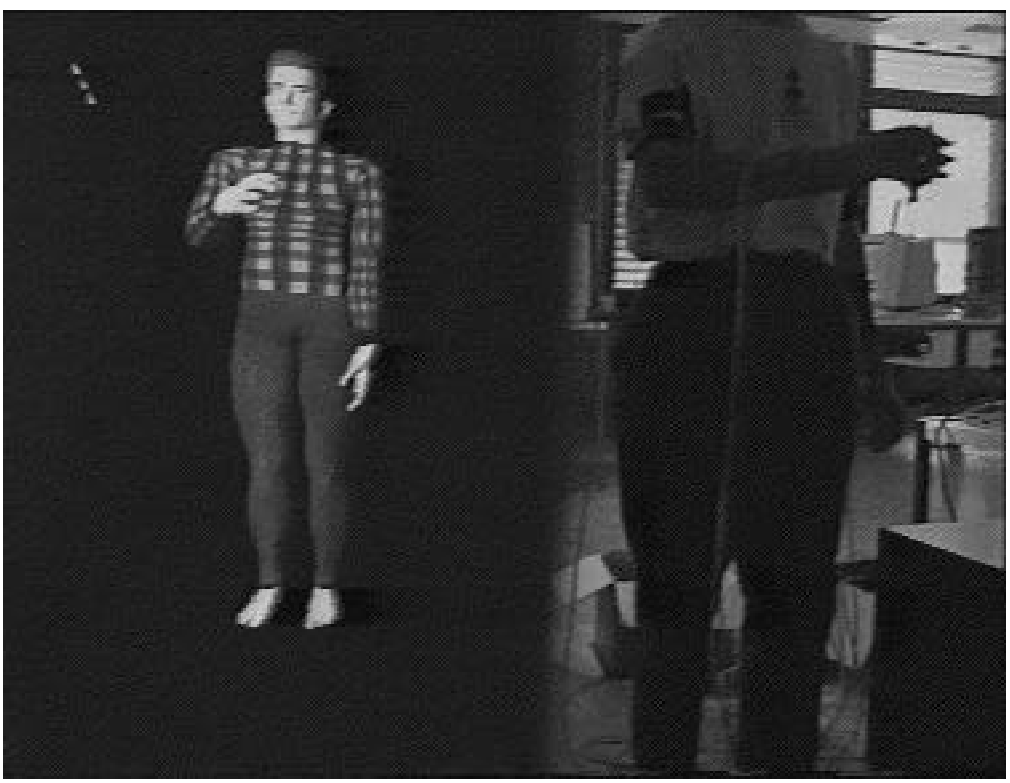

Fig.1. Animating a Virtual Actor using a Flock of Birds

In Virtual Reality, representing participants by a virtual actor is an important factor for presence. This becomes even more important in multi-user environments, where effective interaction among participants is a contributing factor to presence. This is the Self-representation in the Virtual World. Even with limited sensor information, a virtual human frame can be constructed in the Virtual World, that reflects the activities of the real body. Slater and Usoh [5] indicates that such a body, even if crude, heightens the sense of presence.

3. Guided actors

Guided actors are actors which are driven by the user but which do not correspond directly to the user motion. They are based on the concept of *real-time direct metaphor* [4] a method consisting of recording input data from a VR device in real-time allowing to produce effects of different nature but corresponding to the input data. There is no analysis of the meaning of the input data. To understand the concept, we may take an example of traditional metaphor: *the puppet control*. A puppet may be defined as a doll with jointed limbs moved by wires or strings. Similarly glove-puppets are dolls of which the body can be put on the hand like a glove, the arms and head being moved by the fingers of the operator. In both cases, human fingers are used to drive the motion of the puppet.

The best example of actor guidance is guided navigation like implemented in the VLNET system [6] (see Fig.2).

Fig.2. A guided actor in the VLNET system

The animator participant uses the input devices to update the position of the virtual actor. This local control is used by computing the incremental change in the actor position, and estimating the rotation and velocity of the center of body. The walking motor [7] uses the instantaneous velocity of motion, to compute the walking cycle length and time, by which it computes the necessary joint angles. The motor is based on the walking model, guided by the user interactively or automatically generated by a trajectory. This model can also include kinematics personification depending on the individuality of the user, and it is based on the mathematical parameterization coming from biomechanical experimental data. The sensor information for walking can be obtained from various types of input devices like DataGlove, or SpaceBall. Another good example for local control for walking is the virtual treadmill as proposed by Slater et al. [8].

4. Autonomous and perceptive actors

4.1 The Behavior of the actors

Autonomous actors [9 10] are able to have a behavior, which means they must have a manner of conducting themselves. Behavior is not only reacting to the environment but should also include the flow of information by which the environment acts on the living creature as well as the way the creature codes and uses this information. If we consider the synthetic environment as made of 3D geometric shapes, one solution to this problem is to give the actor access to the exact position of each object in the complete environment database corresponding to the synthetic world. This solution could work for a very "small world", but it becomes impracticable when the number of objects increases. Moreover, this approach does not correspond to reality where people do not have knowledge about the complete environment. Another approach has been proposed by Reynolds [11]: the synthetic actor has knowledge about the environment located in a sphere centered on him. Moreover, the accuracy of the knowledge about the objects of the environment decreases with the distance. This is of course a more realistic approach, but as mentioned by Reynolds, an animal or a human being has always around him areas where his sensitivity is more important. Consider, for example, the vision of birds, they have a view angle of 300º and a stereoscopic view of only 15º. The sphere model does not correspond to the sensitivity area of the vision. Reynolds goes one step further and states that if actors can see their environment, they will improve their trajectory planning. This means that the vision is a realistic information flow. Unfortunately, what is realistic to do for an human being walking in a corridor seems unrealistic to do for a computer. However, using hardware developments, it is possible to give a geometric description of 3D objects together with the viewpoint and the interest point of a synthetic actor in order to get the vision on the screen. This vision may then be interpreted like the synthetic actor vision. More generally, in order to implement perception, virtual humans should be equipped with virtual sensors.

4.2 Perception through Virtual Sensors

Perception is defined as the awareness of the elements of environment through physical sensation. In order to implement perception, virtual actors should be equipped with visual, tactile and auditory sensors. These sensors should be used as a basis for implementing everyday human behaviour such as visually directed locomotion, handling objects, and responding to sounds and utterances. For synthetic audition, in a first step, we model a sound environment where the synthetic actor can directly access to positional and semantic sound source information of a audible sound event. Simulating the haptic system corresponds roughly to a collision detection process. But, the most important perceptual subsystem is the vision system. A vision based approach for virtual humans is a very important perceptual subsystem and is for example essential for navigation in virtual worlds. It is an ideal approach for modeling a behavioral animation and offers a universal approach to pass the necessary information from the environment to the virtual human in the problems of path searching, obstacle avoidance, and internal knowledge representation with learning and forgetting. In the next sections, we describe our approach for the three types of virtual sensors: vision, audition and haptic.

4.3 Virtual vision

Although the use of vision to give behavior to synthetic actors seems similar to the use of vision for intelligent mobile robots [12 13], it is quite different. This is because the vision of the synthetic actor is itself a synthetic vision. Using a synthetic vision allow us to skip all the problems of pattern recognition and distance detection, problems which still are the most difficult parts in robotics vision. However some interesting work has been done in the topic of intelligent mobile robots, especially for action-perception coordination problems. For example, Crowley [14], working with surveillance robots states that "most low level perception and navigation tasks are algorithmic in nature; at the highest levels, decisions regarding which actions to perform are based on knowledge relevant to each situation". This remark gives us the hypothesis on which our vision-based model of behavioral animation is built.

We first introduced [15] the concept of synthetic vision as a main information channel between the environment and the virtual actor. Reynolds [16] more recently described an evolved, vision-based behavioral model of coordinated group motion, he also showed [17] how obstacle avoidance behavior can emerge from evolution under selection pressure from an appropriate measure using a simple computational model of visual perception and locomotion. Tu and Terzopoulos [18] use a kind of synthetic vision for artificial fishes.

In [15], each pixel of the vision input has the semantic information giving the object projected on this pixel, and numerical information giving the distance to this object. So, it is easy to know, for example, that there is a table just in front at 3 meters. With this information, we can directly deal with the problematic question: "what do I do with such information in a navigation system?" The synthetic actor perceives his environment from a small window of typically 30x30 pixels in which the environment is rendered from his point of view. As he can access z buffer values of the pixels, the color of the pixels and his own position he can locate visible objects in his 3D environment. This information is sufficient for some local navigation.

We can model a certain type of virtual world representation where the actor maintains a low level fast synthetic vision system but where he can access some important information directly from the environment without having to extract it from the vision image. In vision based grasping for example, an actor can recognize in the image the object to grasp. From the environment he can get the exact position, type and size of the object which allows him to walk to the correct position where he can start the grasping procedure of the object based on geometrical data of the object representation in the world. This mix of vision based recognition and world representation access will make him fast enough to react in real time. The role of synthetic vision can even be reduced to a visibility test and the semantic information recognition in the image can be done by simple color coding and non shading rendering techniques. Thus, position and semantic information of an object can be obtained directly form the environment world after being filtered.

4.4 Virtual Audition

In real life, the behavior of persons or animals is very often influenced by sounds. For this reason, we developed a framework for modeling a 3D acoustic environment with sound sources and microphones. Now, our virtual actors are able to hear [19]. Any sound source (synthetic or real) should be converted to the AIFF format and processed by the sound renderer. The sound renderer takes into account the real time constraints. So it is capable to render each time increment for each microphone in "real time" by taking into account the final propagation speed of sound and the moving sound sources and microphones. So, the Doppler effect, for example, is audible.

4.5 Virtual haptic

One of our aims is to build a behavioral model based on tactile sensory input received at the level of skin from the environment. This sensory information can be used in tasks as touching objects, pressing buttons or kicking objects For example at basic level, human should sense physical objects if any part of the body touches them and gather sensory information. This sensory information is made use of in such tasks as reaching out for an object, navigation etc. For example if a human is standing, the feet are in constant contact with the supporting floor. But during walking motion each foot alternately experiences the loss of this contact. Traditionally these motions are simulated using dynamic and kinematic constraints on human joints. But there are cases where information from external environment is needed. For example when a human descends a stair case, the motion should change from walk to descent based on achieving contact with the steps of the stairway. Thus the environment imposes constraints on the human locomotion. We propose to encapsulate these constraints using tactile sensors to guide the human figure in various complex situations other than the normal walking.

As already mentioned, simulating the haptic system corresponds roughly to a collision detection process. In order to simulate sensorial tactile events, a module has been designed to define a set solid objects and a set of sensor points attached to an actor. The sensor points can move around in space and collide with the above mentioned solid objects. Collisions with other objects out of this set are not detected. The only objective of collision detection is to inform the actor that there is a contact detected with an object and which object it is. Standard collision detection tests rely on bounding boxes or bounding spheres for efficient simulation of object interactions. A very important, but special case is the contact between the hands and objects during the grasping process. Our approach [20] is based on multi-sensors. These sensors are considered as a group of objects attached to the articulated figure. A sensor is activated for any collision with other objects or sensors. Here we select sphere sensors for their efficiency in collision detection (Figure 3). The sphere multi-sensors have both tactile and length sensor properties. Each sphere sensor is fitted to its associated joint shape with different radii. This configuration is important in our method because when a sensor is activated in a finger, only the articulations above it stop moving, while others can still move. By doing this way, all the fingers are finally positioned naturally around the object.

4.6 Perception-based actions

Synthetic vision, audition and tactile allow the actor to perceive the environment. Based on this information, his behavioral mechanism will determine the actions he will perform. Actions may be at several degrees of complexity. An actor may simply evolve in his environment or he may interact with this environment or even communicate with other actors. However, using a kind of reasoning, the actor may behave with a certain degree of intelligence, as discussed in Section 5.

Actions are performed using the common architecture for motion control developed in the European projects HUMANOID [21] and HUMANOID-2. HUMANOID has led to the development of a complete system for animating virtual actors for film production. HUMANOID-2 currently extends the project for realtime applications and behavioral aspects. The heart of the HUMANOID software is the motion control part which includes 5 generators: keyframing, inverse kinematics, dynamics, walking and grasping and high-level tools to combine and blend them. The walking model used is based on biomechanical studies of specific motion pattern [7].

4.7 Emotional actors

Behaviors may be also dependent on the emotional state of the actor. We also developed a model of nonverbal communication [22]. The believability of virtual actors is improved by their capability to interpret and use a nonverbal language. 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. Usually, people don't use consciously nonverbal communication, but they instinctively understand it to a considerable extent and respond to it without any explicit reasoning. These nonverbal communication is essential to drive the interaction between people wit contact (Fig.3) or without contact (Fig.4).

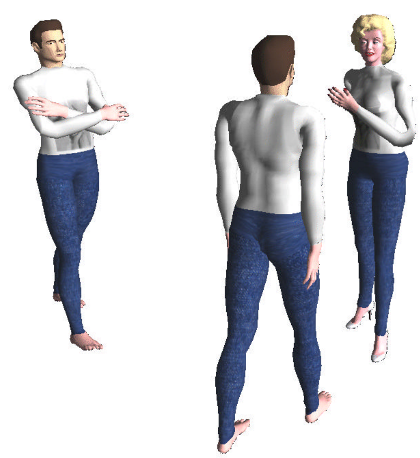

Fig.3. Nonverbal intercommunication

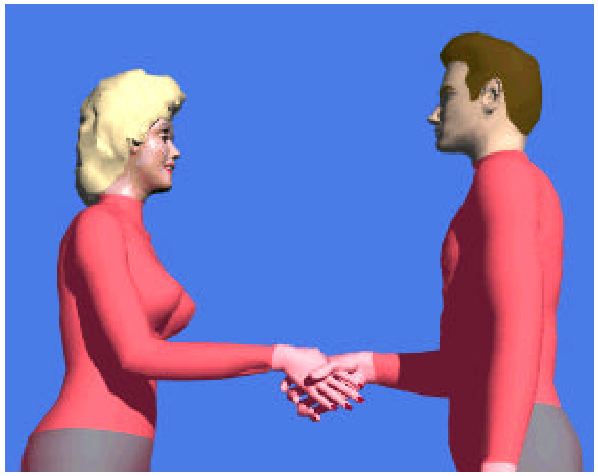

Fig.4. Interaction between two actors

5. Towards "Intelligent" actors

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. In the following we use behavior and automata as synonyms. To control the global behavior of an actor we use a stack of automata. At the beginning of the animation the user provides a sequence of behaviors (the script) and pushes them on the actor's stack. When the current behavior ends the animation system pops the next behavior from the stack and executes it. This process is repeated until the actor's behavior stack is empty. Some of the behaviors use this stack too, in order to reach subgoals by pushing itself with the current state on the stack and switching to the new behavior allowing them to reach the subgoal. When this new behavior has finished the automata pops the old interrupted behavior and continues. This behavior control using a stack facilitates to an actor to become more autonomous and to create his own subgoals while executing the original script.

We present three behaviors, which have some "intelligent" aspects: knowledge-based navigation, knowledge-based locomotion, knowledge-based grasping, and knowledge-based tennis. They are based on a kind of reasoning on the data obtained from the perception process and some basic knowlege. In the future, it will be possible to increate the complexity of situations and reasoning.

5.1 Knowledge-based navigation

The task of a navigation system is to plan a path to a specified goal and to execute this plan, modifying it as necessary to avoid unexpected obstacles [14]. This task can be decomposed into global navigation and local navigation. The global navigation uses a prelearned model of the domain which may be a somewhat simplified description of the synthetic world and might not reflect recent changes in the environment. This prelearned model, or map, is used to perform a path planning algorithm. The local navigation algorithm uses the direct input information from the environment to reach goals and sub-goals given by the global navigation and to avoid unexpected obstacles. The local navigation algorithm has no model of the environment, and doesn't know the position of the actor in the world.

The global navigation needs a model of the environment to perform path-planning. This model is constructed with the information coming from the sensory system. Most navigation systems developed in robotics for intelligent mobile robots are based on the accumulation of accurate geometrical descriptions of the environment. Kuipers et al. [23] give a nearly exhaustive list of such methods using quantitative world modeling. In robotics, due to low mechanical accuracy and sensory errors, these methods have failed in large scale area. We don't have this problem in Computer Graphics because we have access to the world coordinates of the actor, and because the synthetic vision or other simulations of perception systems are more accurate. We develop a 3D geometric model, based on grid, implemented as an octree [24]. With an octree we can easily construct enclosing objects by choosing the maximum depth level of the subdivision of space. Detailed objects like flowers and trees do not need to be represented in complete detail in the problem of path searching. It is sufficient to represent them by some enclosing cubes corresponding to the occupied voxels of the octree. The octree adapts itself to the complexity of the 3D environment, as it is a dynamic data structure making a recursive subdivision of space. The octree has to represent the visual memory of an actor in a 3D environment with static and dynamic objects. Objects in this environment can grow, shrink, move or disappear.

In the last few years, research in robot navigation has tended towards a more qualitative approach to world modeling, first to overcome the fragility of purely metrical methods, but especially, because humans do not make spatial reasoning on a continuous map, but rather on a discrete map [25]. Kuipers et al.[23] present a topological model as the basic element of the cognitive map. This model consists of a set of nodes and arcs, where nodes represent distinctively recognizable places in the environment, and arcs represent travel edges connecting them. Travel edges corresponding to arcs are defined by local navigation strategies which describe how a robot can follow the link connecting two distinct places. These local navigation strategies correspond to the Displacement Local Automata (DLA) implemented in the local navigation part of our system. These DLAs work as a black box which has the knowledge to create goals and sub-goals in a specific local environment. They can be thought of as low-level navigation reflexes which use vision, reflexes which are automatically performed by the adults. There are three families of DLA: the DLAs creating the global goal (follow_the_corridor, follow_the_wall, follow_the_visual_guide), the DLAs creating the local goal (avoid_obstacle, closest_to_goal), and the DLAs effectively moving the actor (go_to_global_goal). The DLAs creating goals only use the vision as input. All these DLAs have access to a library of routines performing high level operations on the vision. A detailed algorithm of the use of vision to find avoidance goal is described by Renault et al. [15].

To illustrate the possibilities of the navigation system, we have developed several examples. For example, an actor is placed inside a maze with an impasse, a circuit and some animated flowers. The actor's first goal is a point outside the maze. After some time, , the actor succeeds in finding his goal. When he had completely memorized the impasse and the circuit, he avoided them. After reaching his first goal, he had nearly complete visual octree representation of his environment and he could find again his way without any problem by a simple reasoning process.

5.2 Knowledge-based locomotion

When the actor evolves in his environment, a simple walking model is not sufficient, the actor has to adapt his trajectory based on the variations of terrain by bypassing, jumping or climbing the obstacles he meets. The bypassing of obstacles consists in changing the direction and velocity of the walking of the actor. Jumping and climbing correspond to more complex motion. These actions should generate parameterized motion depending on the height and the length of the obstacle for a jump and the height and location of the feet for climbing the obstacle. These characteristics are determined by the actor from his perception. Figure 5 shows an actress walking in a landscape.

Fig. 5. Actress walking

The actor can be directed by giving his linear speed and his angular speed or by giving a position to reach. In the first case, the actor makes no perception (virtual vision). He just walks at the given linear speed and turns at the given angular speed. In the second case, the actor makes use of virtual vision enabling him to avoid obstacles. The vision based navigation can be local or global. With a local navigation, the actor goes straight on to his goal and it is possible that he cannot reach it. With a global navigation, the actor first tries to find a path to his goal and if the path exists, the actor follows it until he reaches the goal position or until he detects a collision by his vision. During global navigation the actor memorizes his perceived environment by voxelizing it, based on his virtual vision.

5.3 Knowledge-based grasping

In this problem, our strategy is the following one:

1. Based on a virtual sensors, typically the vision, the actor decides which object to grasp.
2. The actor may have to walk in order to near the objetc
3. Inverse kinematics is used to find the final arm posture
4. Based on a grasp taxonomy [26], an automatic system decides the way the actor grasps the object. For example, the system decides to use a pinch when the object is too small to be grasped by more than two fingers or to use a two hands when the object is large.

5. Using multi-sensors, the fingers a adjusted in order to have an accurate grasping.
Figure 6 shows different types of grasping.

Fig. 6. Examples of grasping different objects

5.4 Vision-based Tennis Playing

Tennis playing (Fig. 7) is a human activity which is severely based on the vision of the players. In our model, we use the vision system to recognize the flying ball, to estimate its trajectory and to localize the partner for game strategy planning. The geometric characteristics of the tennis court however, make part of the players knowledge. For the dynamics simulation of the ball, gravity, net, ground and the racquet we use physics-based calculations. The tracking of the ball by the vision system is controlled by a special automata. A prototype of this automata is already able to track the ball, to estimate the collision time and collision point of ball and racquet and to perform successfully a hit with given force and a given resulting ball direction.

In robotics image recognition and extraction of semantic information from images is a big problem. In virtual environments using synthetic vision this topic is easier to handle as we don't encounter real world's complexity. In navigation problems all perceived objects are only obstacles. Very often only simple object discrimination is needed. In tennis playing, for example, the actors need to differentiate with their vision system between the ball, the other actor and the rest of the environment. To get this semantic information from the vision we use color coding. By associating a certain color to an object. and giving this knowledge to an actor he can localize these objects by the pixel color of his vision image. The actor knows that a certain object is made of a specific material. When it scans the image it looks for the corresponding pixels and calculates its average position and its approximate size. Thus each actor can extract some limited semantic information from the image.

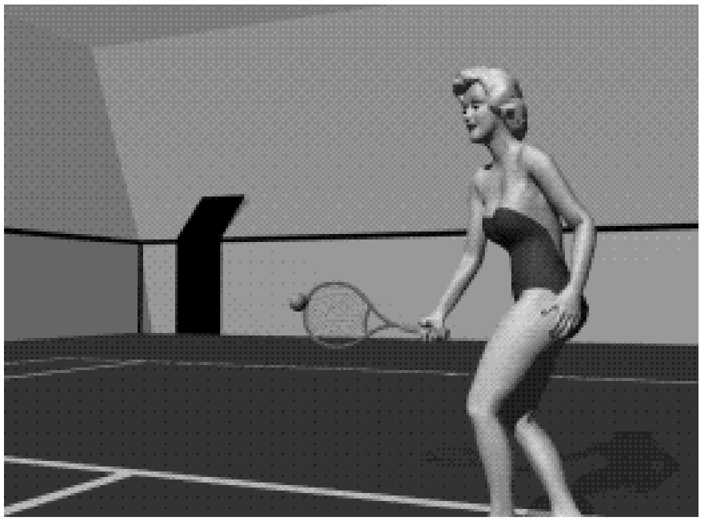

Figure 7. Marilyn playing tennis

Once the actor has recognized the ball, it follows it with his vision system and adjusts at each frame his field of view. To play tennis each partner has to estimate the future racket-ball collision position and time and to move as fast as possible to this point. At each frame (1/25 sec) the actor memorizes the ball position. So, every n-th frame the actor can derive the current velocity of the ball. From this current velocity and the current position of the ball it can calculate the future impact point and impact time. We suppose that the actor wants to hit the ball at a certain height h. In the next phase the actor has to play the ball. Now he has to determine the racket speed and its orientation to play the ball to a given place. Before playing the ball the actor has to decide where to play. In our simulation approach he looks where his partner is placed and then he plays the ball in the most distant corner of the court.

All the above features are coordinated by a specialized "tennis play" automata. First an actor goes to his start position. There he waits until his partner is ready. Then he looks for the ball, which is thrown into the game. Once the vision system has found the ball, it always follows it by adjusting the field of view angle. If the ball is flying towards the actor, it starts estimating the impact point. Once the ball has passed the net, the actor localizes his partner with his vision system during one frame. This information is used for the game strategy. After playing the ball, the actor goes back to his start point and waits until the ball comes back to play it again.

As an example of “intelligent” behavior, if during the impact point estimation the actor estimates the impact point to be outside his court he decides to let the ball and to move directly back to his start position waiting there for a new game.

6. Conclusion

In this paper, we have shown an evolution of virtual actors towards perceptive, autonomous, emotional, and even intelligent actors. The intelligence of virtual actors is constrained and limited to the results obtained in the development of new methods of Artificial Intelligence. However, the representation under the form of virtual actors is a way of visually evaluating the progress.

7. Acknowledgments

The author is grateful to the people who contributed to this work, in particular Pascal Bécheiraz, Ronan Boulic, Tolga Çapin, Zhyong Huang, Joël François, and Hansrudi Noser,. The research was supported by the Swiss National Science Research Foundation, the Federal Office for Education and Science, and is part of the Esprit Project HUMANOID-2.

8. References

1. Hewitt C. (1971) “Description and Theoretical Analysis of PLANNER: a Language for Proving Theorems and Manipulating Models for a Robot”, PhD diss., MIT.
2. Greif I, Hewitt C. (1975) “Actors Semantic of PLANNER-73”, Proc. ACM SIGPLAN/SIGACT Conf., pp.67-77.
3. Magnenat Thalmann N., Thalmann D. (1994) “Creating Artificial Life in Virtual Reality” in: (Magnenat Thalmann and Thalmann, eds) Artificial Life and Virtual Reality, John Wiley, Chichester, 1994, pp.1-10
4. Thalmann D., "Using Virtual Reality Techniques in the Animation Process" in: Virtual Reality Systems (Earnshaw R, Gigante M, Jones H eds), Academic Press, pp.143-159.
5. Slater and Usoh (1994) "Body Centred Interaction in Immersive Virtual Environments", in: (Magnenat Thalmann and Thalmann, eds) Artificial Life and Virtual Reality, John Wiley, Chichester, 1994, pp.1-10
6. I. Pandzic, T. Çapin, N. Magnenat Thalmann, D. Thalmann (1995) VLNET: A Networked Multimedia 3D Environment with Virtual Humans, Proc. Multi-Media Modeling MMM ‘95, World Scientific, Singapore, pp.21-32.
7. Boulic R., Magnenat-Thalmann N. M.,Thalmann D. (1990) "A Global Human Walking Model with Real Time Kinematic Personification", The Visual Computer, Vol.6(6),1990.
8. Slater M., Steed A., Usoh M. (1993) "The Virtual Treadmill: A Naturalistic Method for Navigation in Immersive Virtual Environments", Proceedings of the First Eurographics Workshop on Virtual Environments.
9. Thalmann D. (1994) "Automatic Control and Behavior of Virtual Actors", Interacting with Virtual Environments, MacDonald L., Vince J. (Ed).
10. Magnenat Thalmann N., Thalmann D. (1995) "Digital Actors for Interactive Television", Proc. IEEE.
11. Reynolds C. (1987) “Flocks, Herds, and Schools: A Distributed Behavioral Model”, Proc.SIGGRAPH '87, Computer Graphics, Vol.21, No4, pp.25-34
12. Horswill I. (1993) ”A Simple, Cheap, and Robust Visual Navigation System”, in: From Animals to Animats 2, Proc. 2nd Intern. Conf. on Simulation of Adaptive Behavior, MIT Press, pp.129-136.

13. Tsuji S, Li S. (1993) “Memorizing and Representing Route Scenes”, in: Meyer J.A. et al. (eds) From Animals to Animats, Proc. 2nd International Conf. on Simulation of Adaptive Behavior, MIT Press, pp.225-232.
14. Crowley J.L. (1987) “Navigation for an Intelligent Mobile Robot”, IEEE journal of Robotics and Automation, Vol. RA-1, No. 1, pp 31-41.
15. Renault O., Magnenat Thalmann N., Thalmann D. (1990) ”A Vision-based Approach to Behavioural Animation”, The Journal of Visualization and Computer Animation, Vol 1, No 1, pp 18-21.
16. Reynolds C.W. (1993) ”An Evolved, Vision-Based Behavioral Model of Coordinated Group Motion”, in: Meyer J.A. et al. (eds) From Animals to Animats, Proc. 2nd International Conf. on Simulation of Adaptive Behavior, MIT Press, pp.384-392.
17. Reynolds C.W. (1994) “An Evolved, Vision-Based Model of Obstacle Avoidance Behavior”, in: C.G. Langton (ed.), Artificial Life III, SFI Studies in the Sciences of Complexity, Proc. Vol. XVII, Addison-Wesley.
18. Tu X., Terzopoulos D. (1994) “Artificial Fishes: Physics, Locomotion, Perception, Behavior”, Proc. SIGGRAPH '94, Computer Graphics, pp.42-48.
19. Noser H., Thalmann D. (1995) “Synthetic Vision and Audition for Digital Actors”, Proc. Eurographics ‘95.
20. Huang Z., Boulic R., Magnenat Thalmann N., Thalmann D. (1995) ”A Multi-sensor Approach for Grasping and 3D Interaction”, Proc. CGI ‘95
21. Boulic R., Capin T., Kalra P., Lintermann B., Moccozet L., Molet T., Huang Z., Magnenat-Thalmann N., Saar K., Schmitt A., Shen J. and Thalmann D. (1995) “A system for the Parallel Integrated Motion of Multiple Deformable Human Characters with Collision Detection”, EUROGRAPHICS' 95, Maastricht.
22. Bécheiraz P., Thalmann, D. (1996) “A Model of Nonverbal Communication and Interpersonal Relationship between Virtual Actors”, Proc. Computer Animation ‘96, IEEE Computer Society Press.
23. Kuipers B., Byun Y.T. (1988) “A Robust Qualitative Approach to a Spatial Learning Mobile Robot”, SPIE Sensor Fusion: Spatial Reaoning and Scene Interpretation, Vol. 1003.
24. 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, No1, pp.7-19.
25. Sowa JF (1964) ”Conceptual Structures”, Addison-Wesley.
26. Mas S.R., Thalmann D. (1994) ”A Hand Control and Automatic Grasping System for Synthetic Actors”, Proceedings of Eurographic'94, pp.167-178.