M.I.T. Media Laboratory Perceptual Computing Technical Report No. 257
(To appear, ACM Multimedia Systems)

The ALIVE System:

Wireless, Full-body Interaction with Autonomous Agents

Pattie Maes, Trevor Darrell, Bruce Blumberg, Alex Pentland
MIT Media Laboratory
20 Ames Street
Cambridge, MA 02139
pattie/trevor/bruce/sandy@media.mit.edu
(617) 253-7442 (phone)
(617) 258-6264 (fax)

Revised, November 1995

Abstract

The cumbersome nature of wired interfaces often limits the range of application of virtual environments. In this paper we discuss the design and implementation of a novel system, called ALIVE, which allows unencumberedfull-body interaction between a human participant and a rich graphical world inhabited by autonomous agents. Based on results obtained with thousands of users, the paper argues that this kind of system can provide more complex and very different experiences than traditional virtual reality systems. The ALIVE system significantly broadens the range of potential applications of virtual reality systems; in particular, the paper discusses novel applications in the area of training and teaching, entertainment, and digital assistants or interface agents. We overview the methods used in the implementation of the exiting ALIVE systems.

KEYWORDS: Interactive Interface, Virtual Reality, Autonomous Agents, Computer Vision.

1 Introduction

Virtual Environments allow a user to be embedded into a computer generated environment. Most interfaces to these environments require the use of gloves, goggles, and/or a helmet, and offer limited interactions including only hand gestures and the ability to change viewpoint. (Notable exceptions are [21, 10, 15], discussed below.) The cumbersome nature of the equipment and the limited nature of the interaction in these interfaces has proven to restrict the range of applications that have used this type of technology.

In this paper we discuss the design and implementation of a novel system called ALIVE, an "Artificial Life Interactive Video Environment", that allows wireless full-body interaction between a human participant and a rich graphical world inhabited by autonomous agents. We use a single video camera to obtain a color image of a person, which we composite into a 3D graphical world. The resulting image is projected onto a large screen that faces the user and acts as a type of "magic mirror" (Figure 1): the user sees herself surrounded by objects and agents. No goggles, gloves, or wires are needed for interaction with the virtual world. Computer vision and audition techniques are used to extract information about the person, such as her 3D location, the position of various body parts as well as simple gestures and utterances. We combine active sensing and domain knowledge to achieve robust and real-time performance on these tasks.

The virtual world in our system is inhabited by inanimate objects as well as agents. Agents are modeled as autonomous behaving entities that have their own sensors and goals and that can interpret the actions of the participant and react to them in "interactive-time". Because of the presence of these semi-intelligent entities, the system does not only allow for the obvious direct-manipulation style of interaction, but also for a more powerful, indirect style of interaction in which gestures can have more complex meanings, which may vary according to the situation in which the agents and user find themselves.

Based on results obtained with real users in several public forums where the ALIVE system has been installed, this paper argues that (1) the "magic mirror" approach has several advantages over head-mounted display-based virtual reality systems for certain applications, and (2) virtual worlds including autonomous agents can provide more complex and very different experiences than traditional interactive virtual environments. The ALIVE system significantly broadens the range of potential applications of virtual reality systems; in particular, we discuss applications in the area of training and teaching, entertainment and participatory theater, and last but not least, telecommunication and interface agents.

The following sections discuss how autonomous agents are modeled in the ALIVE system, how vision algorithms are used to implement a wireless full-body interface, and how the agents we have devised interact with people. We conclude with a discussion of the experiences of users of our installations, issues for further study, and possible novel applications now possible with this new interface.

2 Modeling Agents

One important feature of the ALIVE system is that it allows the user to interact, in natural and believable ways, with autonomous semi-intelligent agents. Our agents have a set of internal needs and motivations, a set of sensors with which to perceive their environment, a repertoire of activities which they can perform, a motor system which allows them to move in and act on the virtual environment and a behavior system which chooses, in real-time, the set of activities to perform given the internal needs of the agent and the opportunities presented by the environment. The agent's state and geometry is updated according to the motor activities associated with the chosen behaviors and is re-rendered on every time-step. The user's location and hand and body gestures affect the behavior of the agents, and the user receives visual as well as auditory feedback about the agents' internal state and reactions.

A number of researchers have taken an Artificial Intelligence approach to animation in which the animated agents perform some actions in response to their perceived environment. Reynolds [28] was among the first to apply behavior-based animation to computer graphics (see also [7, 33]). This work, together with the work of Raibert [26], Girard [18], McKenna [24], and Badler [2] has focused primarily on modeling single behaviors (e.g. flocking, walking etc...), and on the problems of motion planning and motor control. Some researchers, including Badler [2], Zeltzer [36] and the Thalmanns [30] have begun to propose more general architectures for behavior-based animated characters. Most recently Tu and Terzopoulos [31] have modeled autonomous animated fish. Their model incorporates a "physics-based" model of fish locomotion and motor control, a perception system utilizing synthetic vision, and a behavior system which models a number of fish behaviors. While very impressive, the system's design is closely tied to the specifics of fish locomotion and behavior.

A number of other researchers have focused on building systems in which animated agents interact, in realtime, with the user. Joe Bates' Woggle world combines behavior models with ideas from traditional animation to explore what he calls "believable agents" [4]. Bates' agents can have fairly complex interactions, despite the user interaction being limited to moving a mouse. Fisher's [17] Menagerie system allows a user to interact with animated agents in real-time using goggles. In contrast with behaviors in the ALIVE system, these agents typically are engaged in a single high-level behavior, for example, flocking. A major difference between the ALIVE system and these others is the combination of both behavioral and motivational complexity in the creature model, and the use of a vision interface with which these behaviors can use the user's actual position, body pose, and hand gestures as sensory input.

The ALIVE system incorporates a behavior modeling tool kit for developing semi-intelligent autonomous agents and their interaction with one another and with the user [6] [5]. This tool kit is the result of our long-term research into architectures for adaptive autonomous agents [22]; it's goal is to produce agents that choose the set of activities on every time step that "make the most sense" given the agent's internal needs and motivations, past history, and the perceived environment with its attendant opportunities, challenges and changes.

Deciding on the right set of actions is complicated by a number of factors. For example, due to the problems inherent in sensing and perception, an agent's perception of its world is likely to be incomplete at best, and completely erroneous at worst. Moreover, there are typically competing goals which work at cross-purposes (e.g. moving toward food may move the agent away from water). Ideally, an agent will neither dither among competing activities nor pursue an unattainable goal to the exclusion of lower priority but attainable goals. External opportunities need to be weighed against internal needs in order to provide just the right level of opportunistic behavior. Actions may be unavailable or unreliable. To successfully produce competent autonomous action over extended periods of time, the agent's behavior system must provide solutions to these problems, as well as others. See [6] and [5] for a detailed discussion of how our archictecture addresses these issues.

When using the behavior tool kit to build an agent, the designer specifies:

- *The virtual sensors of the agent.* Virtual sensors are used by the agents to sense other agents in the world including the agent which acts as the virtual world's representation of the user. Agents in ALIVE also use synthetic vision for low level obstacle avoidance and navigation. See Reynolds [28], Renault [27]and Tu [31] for other examples of this approach.

- *The releasing mechanisms of the agent.* Releasing Mechanisms are entities which identify behaviorally significant stimuli from the agent's set of sensors and transduce values which represent the behavior-specific strength of the stimuli. For example, a dog may have a releasing mechanism for detecting when the user's hand is extended and down and its value may be inversely proportional to the distance of the dog's nose from the user's hand.

- *The motivations or internal needs of the agent.* Internal needs are modeled as variables which may vary over time. For example, a "dog" agent may have an internal need to receive attention from the user. Whenever the user pats the dog, this variable will temporarily decrease in value and as a result the dog will be less motivated to seek human attention.

- *The behaviors of the agent.* The behavior system of the agent is organized as a distributed collection of goal-directed, self-interested entities called behaviors or activities (e.g. "find-food", "chase-hamster"). Behaviors rely on input from relevant motivational variables and releasing mechanisms to arrive at a "value" which represents the importance or relevance of the behavior given the agent's state and environment. They compete on the basis of this value for control of the agent's body (i.e. for the prioritized privilege to issue motor commands to the agent's motor system). Behaviors may be organized into groups of competing behaviors called behavior groups, which in turn may be organized into loose hierarchies with the top of the hierarchies representing more general behaviors (e.g. "find-food") and the leaves representing more specific behaviors (e.g. "chew"). The lowest-level behaviors control the motor system, for example, making the dog move a little to the left or right, or making it bark in a certain way.

- *The motor skills of the agent.* The primitive actions of the agent such as "walking", "sitting", "wagging tail" etc... are called motor skills. Motor skills rely on kinematic and inverse-kinematic modeling to modify the underlying geometry. Motor skills and the motor system in general are defined in such a way that only motor skills which are complementary may run concurrently. Thus, the dog may "walk", "look-at", "wag-tail" and "growl" all at the same time, but may not "walk" and "sit" concurrently. In addition, the motor system supports blending of motor preferences by the behavior system. For example, one behavior may express a preference to move forward, and another to use a "bound", and the result is the agent moves forward using a bounding gait.

The ALIVE system creates a “special” 3D agent in the environment to represent the user. The position and state of that agent are based on the information computed by the vision system on the basis of the camera image of the user, as well as auditory cues [9]. Thus, the artificial agents can sense the user using the same virtual sensors that they use to detect objects and other agents. The person agent is rendered in the final image using the live video image of the actual user, or using a combination of graphics and texture-mapped video in situations where bandwidth is limited.

3 A Vision-Based Interface

We wish to create a non-intrusive interface to virtual worlds, while allowing a rich and intuitive set of gestures to be used in controlling and navigating that world. Passive computer vision techniques hold great promise as an interface tool, especially as available computational and video processing power increases. We are currently at the point where simple, real-time computer vision techniques can provide the types of information that had previously only been possible with wired sensors, as well as providing information that has heretofore been unavailable.

A wireless sensor, such as vision, has several additional advantages over tethered goggles-and-gloves interfaces. It provides a safer solution because the user can still see where she is moving, and thus can avoid bumping into things, or tripping over wires. Second, the user enjoys greater behavioral and expressive freedom. We observed that users of the ALIVE system feel very uninhibited (we have have seen users doing cartwheels, jumping jacks, etc). Finally, the user ends up concentrating more on the environment itself, rather than on the complex and unfamiliar equipment being used to interact with that environment.

Other systems, such as the Visual Portal [15] and the CAVE [10] system have solved many of the limitations of traditional goggle-based environments through the use of wireless batons and other sensors, thus avoiding both the problems of a tethered display and viewpoint estimation (head angle). Our system has the advantage that it is completely unencumbered, and works on users with no special tools or marks. We also adopt a mirror paradigm, where the user explicitly sees a representation of him/herself and his/her relationship to other objects in the world.

The novel vision-based interface presented here was inspired by the pioneering work of Myron Krueger's Videoplace system [21]. ALIVE and Videoplace differ primarily in three respects. The first is that Videoplace focuses on 2D rather than 3D worlds and interaction. A second difference is our emphasis on modeling agents. Most of Krueger's worlds allow users to interact with other users, a notable exception being the "critter", a 2D animated sprite. Finally, the ALIVE vision system is able to recognize hand and body gestures as patterns in space and time.

Another system that bears similarities to ALIVE is the Mandala system [32] which composites the user's color image with a virtual world that is sometimes video-based and sometimes computer animated. Unlike ALIVE, the Mandala system only supports 2D and requires a chromakey background or specially-colored manipulation objects; it does not attempt to recognize parts of the user's figure nor does it do any gesture recognition. Other systems have been developed for vision-based interactive graphics but have generally been restricted to off-line analysis of either face or limb motion [14, 34, 29]. (But see [13] for a real-time facial analysis system.)

We have developed a set of vision routines for perceiving body actions and gestures performed by a human participant in an interactive system. Vision routines in ALIVE acquire the image of the user, compute a figure/ground segmentation, and find the location of head, hands, and other salient body features. We use only a single, calibrated, wide field-of-view camera to determine the 3-D position of these features. We do assume that the background is fixed, although it can be arbitrarily complex, and that the person is normally facing the camera/screen. With the most recent version of our system, the integration of the person and and localization of his/her head or hand features in the world are performed using the following modules: figure-ground processing, scene projection, hand tracking, and gesture interpretation.

To detect appropriate hand/face features and composite the user's image onto the magic mirror, the vision system must isolate the figure of the user from the background (and from other users, if present). This is accomplished by use of spatially-local pattern recognition techniques to characterize changes in the scene, followed by connected-components and morphological analysis to extract objects.

We assume the background to be an arbitrary, but static, pattern. Mean and variance information about the background pattern are computed, and these statistics are used to determine space-variant criteria for pixel class membership. We will omit mathematical details of our segmentation algorithm, for further reference see [35]. In general, we use a hierarchical color classification is used to compute figure/ground segmentation, using a Gaussian model of each background pixel's color and an n-class adaptive model of foreground (person) colors. The classification takes care to identify possible shadow regions, and to normalize these region's brightness before the figure/ground classification. The classification also makes use of Markov neighborhood statistics in setting the priors for each pixel's classification.

Once each pixel has been identified as most likely belonging to the user, we use connected components and morphological analysis to delineate the foreground region.

This analysis begins with a seed point at the centroid location of the person in the previous frame; if this fails to grow a sufficiently large region, random seed points are selected until a stable region is found. Finally, we compute the contour of the extracted region by chain-coding the connected foreground region.

When the figure of the user has been isolated from the background, we compute an estimate of its 3D location in the world. If we assume the user is indeed sitting or standing on the ground plane, and we know the calibration of the camera, then we can compute the location of the bounding box in 3D. Establishing the calibration of a camera is a well-studied problem, and several classical techniques are available to solve it in certain broad cases [3, 19]. Typically these methods model the camera optics as a pinhole perspective optical system, and establish its parameters by matching known 3D points with their 2D projection.

Knowledge of the camera geometry allows us to project a ray from the camera through the 2D projection of the bottom of the bounding box of the user. Since the user is on the ground plane, the intersection of the projected ray and the ground plane will establish the 3D location of the user's feet. The 2D dimensions of the user's bounding box and its base location in 3D constitute the low-level information about the user that is continuously computed and made available to all agents in the computer graphics world. The contour is projected from 2-D screen coordinates into 3-D world coordinates, based on the computed depth location of the person. This is then used to perform video compositing and depth clipping to combine the user's video image with computer graphics imagery.

One of the most salient cues used by the agents in our world is the location of the user's hands. We have implemented a feature localization algorithm that determines hand locations by searching within a window along the side of the contour for extremal horizontal and vertical points. If the highest point in the window is above the shoulder of the user, we label that the hand, otherwise the horizontal extremal point is used. The highest point within a window of the contour located above the centroid of the foreground region is labeled the head. These feature localiztion algorithms are not infallible, but we have found they work well in a wide range of conditions, especially if combined with color space classification to identify the location of flesh tones[35].

ALIVE improves on earlier systems in which only the 2D position of the user's hand was used to determine activation of objects such as virtual buttons. The improvements avoid inadvertent manipulation of objects, such as unintended activation of buttons. The system uses combination of clues including 2D position of the hands, Z position of the user's body, and gesture information to make sure that the user's intention is to actually manipulate an object. For example, in order for the button to be pushed, the user has to perform a "pointing gesture", have the hand over the button in 2D and be in the correct Z-plane.

Both the absolute position of hands, and whether they are performing characteristic gesture patterns, are relevant to the agents in the virtual world. We use pattern recognition strategies to detect and classify these characteristic gesture patterns. Static gestures, such as pointing, are computed directly from the hand feature location. To recognize dynamic gestures, we use a high-resolution, active camera to provide a foveated image of the hands (or face) of the user. The camera is guided by the computed feature location, and provides images which can be used successfully in a spatio-temporal gesture recognition method[12].

4 ALIVE Environments

We have combined the behavior modeling and vision techniques described above to construct a system for video-based interaction with artificial agents. In our system the user moves around in a real-world space of approximately 16 by 16 feet. A video camera at the front of the space captures the user's image, which is then segmented from the background. The user's image is composited with the 3D animated world and the resulting image displayed on a large projection screen (10 by 10 feet) that faces the user.

Several virtual worlds have been implemented for this environment. The first ALIVE installation had two worlds, one inhabited by a Puppet character and one by Hamster/Predator creatures. The behavior of these agents were fairly simple, and was constructed using the behavior toolkit described above.

In the Puppet world, the Puppet would follow the user around (in 3D) and try to hold the user's hand, and would imitate some of the actions of the user (sitting down, jumping, etc). It would be sent away when the user pointed away and come back when the user waved. The Puppet employed facial expressions to convey its internal state. For example, it would pout when the user sent it away and smile when the user motioned it to come back. It giggled when the user would touch its belly.

In a typical experience with the Hamster/Predator world, the Hamster avoided objects and would follow the user and beg for food, if none was available. The Hamster would roll over to have its body scratched if the user bent over. If the user has been patting the Hamster for a while, its need for attention would be fulfilled and some other activity would take over (e.g., looking for food). The user was able to feed the Hamster by picking up food from a virtual table and putting it on the floor. The user was also able to let the Predator out of its cage and into the Hamster's world. The predator would then try to chase and eat the Hamster. The Predator viewed the person as a predator and would attempt to avoid and flee from the person. Both the Predator and the Hamster were successful at arbitrating among their multiple internal needs (avoiding the predator, finding food, not running into obstacles, etc.).

In the most recently implemented ALIVE world the user can interact with a virtual dog, as shown in Figures 3-6. The Dog has a more sophisticated repertoire of behaviors than the previous characters, including behaviors for playing, feeding, drinking, receiving attention and sleeping. The Dog also uses auditory input, consisting of simple verbal commands that are recognized using a commercial speech recognition system, and produces auditory output, consisting of a wide variety of prerecorded samples which are played at appropriate times. The dog has both interactive behaviors and autonomous action; while it's primary goal is to play with the user internal motivations (e.g., thirst) will occasionally override.

In the ALIVE worlds, the user can interact with the agents using gestures that are interpreted by the agents depending on their current state. An important contribution of the agent system is that behavior and gestures of the user are interpreted differently by agents in the world depending on their current context and past history. Users can use a direct manipulation interaction style [20] for objects which have no associated behaviors.

For example, in the current implementation the waving gesture elicits a response from the Puppet agent, but not from the Dog agent. And the response given by the Puppet is state dependent; it will return to the user when waved at only when it has been sent away or otherwise ignored by the user. In addition, when the user performs a pointing gesture, and thereby sends the Puppet away, the Puppet will go to a different place depending on where the user is standing (as well as the direction in which she is pointing in). If the user comes towards the Puppet after it has been "sent away", this gesture is interpreted to mean that the user no longer wants the Puppet to "go away", and so the Puppet will smile and return to the user. In this manner, the gestures employed by the user can have rich meaning that varies on the previous history, the agents internal needs and the current situation.

5 User Experiences

The ALIVE system has been installed in several public exhibitions,[1] which has allowed us to gather experience with thousands of users in the system. The overwhelming majority report that they enjoy interacting with the system and consider the actions and reactions of objects and agents believable. We feel several interesting lessons can be learned from these users experiences:

- The magic mirror paradigm proves natural and easy to use. Unlike viewer-centered virtual reality systems, the user does not get easily disoriented. They know at all times where they are in the artificial world and can observe the actions of the other agents as well as their own. The user is able to see the whole world as in a third person perspective; the mirror-like paradigm makes it easier for the user to map her actions onto those of her image than is the case in a normal third person perspective, because the mirror mapping from user to user's image is a very simple one which people are familiar with. However, some users do report confusion when moving forwards or backwards with respect to the magic mirror. This may be due to the fact that movement front or back is displayed as movement down or up on the magic mirror. This appears counter-intuitive to some people (i.e. they appear to associate front with up, and back with down). A possible solution to this problem is to put the camera closer to eye level, thereby reducing the perspective distortion (in ALIVE the camera is located on top of the screen, pointing down in a 45 degree angle).

[1] including Tomorrow's Realities, SIGGRAPH 93, Anaheim; AAAI-94 Art Exhibition, Seattle; ARTEC-95, Nagoya, Japan; and Interactive Communities, SIGGRAPH 95, Los Angeles.

- Successful gestures are intuitive with respect to the domain and provide immediate feedback. Users have less difficulty interacting with the agents when they can use simple gestures that are natural and appropriate given the domain, e.g petting for creatures or pointing and waving for the virtual puppet. In addition, interaction is improved when the user receives feedback from the agents, either in terms of movement and/or facial and body expression (the Hamster rolls over when being patted, the Puppet smiles when being tickled, etc). In the ALIVE system, whenever an agent recognizes a gesture by the user, it provides some distinguishable visual feedback to the user. This helps the user get an understanding for the space of recognized gestures.

- Users are more tolerant of imperfections in an agent's perception (such as lags and occasional incorrect or missed recognition) as opposed to that of objects in the virtual world. People expect virtual inanimate objects to "work" like an object, i.e. the reaction of the object has to be immediate, predictable and consistent. On the other hand, people assume that animal or human-like agents have perception and state, and thus are able to accept that the agent may not have sensed something. As a result, gestures that are hard to recognize, such as waving, can be used successfully in the context of agents (an agent might not have "seen" the user waving), but the same gesture would cause the user frustration if used in the context of some inanimate object, e.g., a switch.

6 Discussion and Conclusions

The "magic mirror" interface introduced by the ALIVE system offers several advantages over tethered goggle and glove interfaces. One direct benefit is that it constitutes a safer and more convenient solution, because the user is free from goggles, gloves, and wired sensors. More importantly, it constitutes a more universal solution to virtual reality since users of all shapes and sizes can use one and the same system. Because of the natural gesture-based interaction, users do not need to be given detailed instructions beforehand about how to use the system. All of these advantages make the magic mirror type system more accessible to a broad group of users, including children, the physically handicapped, etc.

The most significant characteristic of the magic mirror approach is that the user can see herself in the virtual world. Typical goggles-and-gloves systems only allow the user a first person point of view: the user only sees her hands. ALIVE allows a third person point of view, so the user tends to be less disoriented about her own position and orientation in the space. This third person perspective is essential to applications that require a user to see her own body. For example, when using a virtual environment to teach a user a physical skill such as dance or golf, the user can benefit from being able to see (and then correct) her own position and movements.

The presence of agents in the virtual environment opens up a whole range of novel applications for virtual environments. The ALIVE system demonstrated here represents only the beginning of a set of novel applications that could be explored. Work is currently underway at the MIT Media Laboratory to build prototypes for the following applications:

- Entertainment agents.
  The ALIVE worlds described above are simple examples of entertaining agents. We are currently investigating ALIVE for interactive story telling applications in which the user plays one of the characters in the story and all other characters are artificial agents that collaborate to make the story move forwards. Another obvious entertainment application of ALIVE is video games. We have hooked up the ALIVE vision-based interface to existing video game software, so as to let the user control a game with her full body. In addition, we are investigating ways in which autonomous video game characters can learn and improve their competence over time, so as to keep posing a challenge to a video game player.

- Agents as personal teachers and trainers.
  The magic mirror metaphor is ideally suited to teach a user physical skills, because the user can see herself performing the movements and actions to be mastered. Furthermore, the autonomous agents in the world can be modeled to act as a personal trainer, able to demonstrate to the user how to perform an action and to provide personalized and timely feedback to the user on the basis of the sensory information about the user's gestures and body positions. For example, one could build a virtual aerobics teacher that would give feedback to the user to tell her she has to lift her knees higher or 'move those feet a little faster'.

- Interface agents or personal digital assistants.
  The ALIVE system can be successfully combined with the concept of a personal digital assistant. These "interface agents" help a user with daily tasks such as remembering where things were put, filtering electronic news and mail, scheduling meetings, etc [23]. A system such as ALIVE can be used to visualize the interface agent's states and activities. In addition, in an enhanced reality setup, the interface agent could point to real objects as well as virtual objects in its interactions with the user. For example, the agent could point at a drawer of a real filecabinet and remind the user that she put a particular document in that drawer.

In conclusion, the ALIVE system allows a user to interact in an unobtrusive, natural way with a virtual environment inhabited by autonomous agents. The "magic mirror" interface introduced by the system has proven practical and universal: users do not need any training nor do they need to be equipped in order to successfully interact with the world. They can use simple, natural gestures, that are interpreted by the autonomous agents inhabiting the environment. The use of unobtrusive, full-body sensing as well as the presence of agents in the environment opens up a range of novel applications for virtual environments. The technology developed in the ALIVE project introduces possibilities for modeling virtual environment systems containing agents that act as personal entertainers, teachers and assistants.

Acknowledgments

Graduate students Michael P. Johnson, Thad Starner and Leonard Foner as well as undergraduate students Johnny Yoon, Cecile Pham and Kenneth Russell contributed significantly to the implementation of ALIVE. Jeremy Brown was responsible for the initial implementation of the Puppet character. This work was funded by SEGA of America, the Office of Naval Research, and the National Science Foundation.

References

[1] Aloimonos Y., ed., (1993) Active Perception, Lawrence Erlbaum Associates

[2] Badler N.I., Phillips C.B. and Webber B.L., (1993) Simulating Humans, Computer Graphics Animation and Control, Oxford University Press

[3] Ballard, D., and Brown, C., (1982) Computer Vision, Prentice-Hall, Englewood

[4] Bates J., Altucher J., Hauptman A., Kantrowitz M., Loyall A.B., Murakami K., Olbrich P., Popovic Z., Reilly W.S., Sengers P., Welch W., Weyhrauch P. and Witkin A.,(1993) Edge of Intention, SIGGRAPH-93 Visual Proceedings, Machine Culture, ACM SIGGRAPH, pp. 113-114.

[5] Blumberg B., (1994), Action-Selection in Hamsterdam: Lessons from Ethology, Proceedings of the Third International Conference on the Simulation of Adaptive Behavior, MIT Press, Brighton, August 1994.

[6] Blumberg, B and Tinsley Galyean. (1995) Multi-level Direction of Autonomous Creatures for Real-Time Virtual Environ ments. Proceedings of SIGGRAPH 95 .

[7] Broadwell, P, Myers, R., and Schaufler, R., (1985) Plasm: A Fish Sample, Siggraph 85 Art Show, Siggraph Visual Proceedings, ACM Press. Also available as "gold", IndyZone2 CDROM, Summer 1994, Silicon Graphics Inc.

[8] Burden R. and Faires J.,(1989) Numerical Analysis, PWS-Kent Publishing Co. Boston.

[9] Casey, M., Gardner, W., and Basu, S., (1995) "Vision Steered Beam-forming and Transaural Rendering for the Artificial Life Interactive Video Environment (ALIVE) ". Proc. Audio Eng. Soc. Convention 1995.

[10] Cruz-Neira, C., Sandin, D.J., DeFanti, T.A., Kenyon, R., and Hart, J.C., (1992) "The CAVE, Audio Visual Experience Automatic Virtual Environment", Communications of the ACM, June 1992, pp. 64-72

[11] Darrell T. and Pentland A., (1993) "Space-Time Gestures," In: IEEE Conference on Vision and Pattern Recognition, NY, NY, June 1993.

[12] Darrell, T., and Pentland, A., (1995) "Attention-driven Expression and Gesture Analysis in an Interactive Environment," Int'l Workshop on Face and Gesture Recognition, Zurich, Switzerland, June 26-28.

[13] Darrell, T., Essa, I., and Pentland, A., (1994) “Correlation and Interpolation Networks for Real-Time Expression Analysis/Synthesis”, Advances in Neural Information Processing Systems-7, Tesauro, Touretzky, and Leen, eds., MIT Press. Conference",

[14] Essa, I., and Pentland, A., (1995) Facial Expression Recognition using a Dynamic Model and Motion Energy, In Proc. Fifth Intl. Conf. Computer Vision, IEEE Computer Society, pp. 76-83.

[15] Deering. M., (1992) High Resolution Virtual Reality. Computer Graphics, Vol. 26, 2, July 1992, pp. 195-201

[16] Featherstone R., (1987) Robot Dynamics Algorithms, Kluwer Academic Publishers, Boston.

[17] Fisher S.S., Girard M. and Amkraut S., Menagerie, (1993) Tomorrow’s Realities, SIGGRAPH-93 Visual Proceedings, ACM SIGGRAPH 1993, pp. 212-213.

[18] Girard, Michael and A. A. Maciejewski. Computational Modeling for the Computer Animation of Legged Figures. Pro ceedings of SIGGRAPH 85 (San Francisco, CA, July 22-26, 1985). In Computer Graphics 19, 263-270.

[19] Horn, B.K.P., (1991) Robot Vision, M.I.T. Press, Cambridge, MA.

[20] Hutchins E. L., Hollan J. D. and Norman D. A., Direct Manipulation Interfaces, (1988) in: User-centered system design: new perspectives on human-computer interaction, D. A. Norman and S.W. Draper (eds), Lawrence Erlbaum Associates, pp 87-124.

[21] Krueger M.W., (1990) Artificial Reality II, Addison Wesley.

[22] Maes P. (1991), Designing Autonomous Agents: Theory and Practice from Biology to Engineering and Back, Bradford Books/MIT Press.

[23] Maes P., (1994) Agents that Reduce Work and Information Overload, Communications of the ACM, July.

[24] McKenna, Michael and David Zeltzer. Dynamic Simulation of Autonomous Legged Locomotion. Proceedings of SIGGRAP H 90 (Dallas, TX, August 6-10, 1990). In Computer Graphics 24, 4 (August 1990), pp.29-38.

[25] Mountford S.J. and Gaver W.W., Talking and listening to computers, (1990) In: The art of human-computer interface design, Brenda Laurel (ed), Addison Wesley.

[26] Raibert M. and Hodgins J., (1991) Animation of Dynamic Legged Locomotion, Computer Graphics: Proceedings of SIGGRAPH ’91, 25(4), ACM Press, July.

[27] Renault, O., Thalmann, N., Thalmann, D., A vision-based approach to behavioral animation. The Journal o f Visualization and Computer Animation 1(1),1990, pp.18-21.

[28] Reynolds C.W., (1987) Flocks, Herds and Schools: A Distributed Behavioral Model, Computer Graphics: Proceedings of SIGGRAPH ’87, 21(4), ACM Press, July 1987.

[29] Terzopoulus, D., and Waters, K., (1993) Analysis and Synthesis of Facial Image Sequences Using Physical and Anatomical Models, IEEE Trans. Pattern Analysis and Machine Intelligence, 15(6), pp. 569-579, June

[30] Thalmann, N., and Thalmann D.,, (1994) Artificial Life and Virtual Reality. John Wiley & Sons, Chichester U.K.

[31] Tu, X., and Terzopoulos, D., Artificial Fishes: Physics, Locomotion, Perception, Behavior Computer Graphics: Proceedings of SIGGRAPH 1994, ACM Press, July 1994.

[32] Vincent V.J., (1993) Mandala: Virtual Village, SIGGRAPH-93 Visual Proceedings, Tomorrow's Realities, ACM SIGGRAPH 1993, pp. 207.

[33] Wilhelms J., R. Skinner. A 'Notion' for Interactive Behavioral Animation Control. IEEE Computer Graphics and Applications 10(3) May 1990, pp. 14-22.

[34] Williams, L., Performance-driven facial animation, Computer Graphics: Proceedings of SIGGRAPH 1990, 24(4), ACM Press, pp. 235-242.

[35] Wren, C., Darrell, T., Starner, T., Johnston, M., Russell, K., Azarbayejani, A., and Pentland, A. (1995) pfinder: A Real-Time System for Tracking People, SPIE Conference on Real-Time Vision, M. Bove, Ed., Philadelpia, PA, July 1995.

[36] Zeltzer D., (1991) Task-level graphical simulation: abstraction, representation and control, in: N.I. Badler, B.A. Barsky and D. Zeltser (editors), Making them move: mechanics, control and animation of articulated figures, Morgan Kauffman, pp. 3-33.

Figure 1: The ALIVE "Magic-Mirror": a user sees himself in a virtual world.

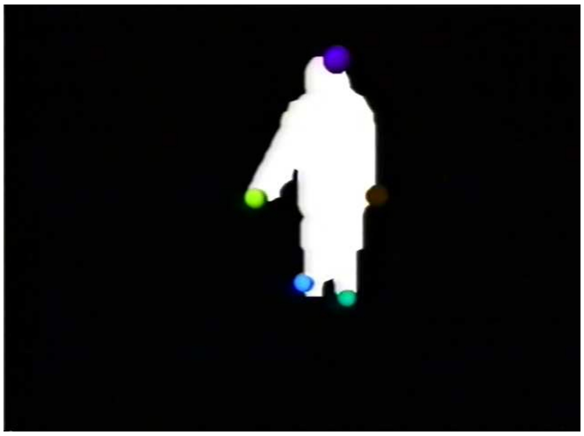

Figure 2: Simple vision routines extract figure from background and label salient points of contour.

Figure 3: Image of user is composited with computer graphics. Here the Dog responds to pointing gesture by sitting.

Figure 4: Another example of a recognized gesture. Dog walks in direction indicated by user.

Figure 5: Dog shakes hands with user. Dog respods to hand gestures differently depending on stance of user.

Figure 6: Image of Dog standing on hind legs to mimic user's gesture.