Video Agents

Kazuhiro Asai* Takashi Nishimoto† Yoshinori Hirano‡ Emiko Hama§ Yoshifumi Kitamura¶
Fumio Kishino‖
Osaka University

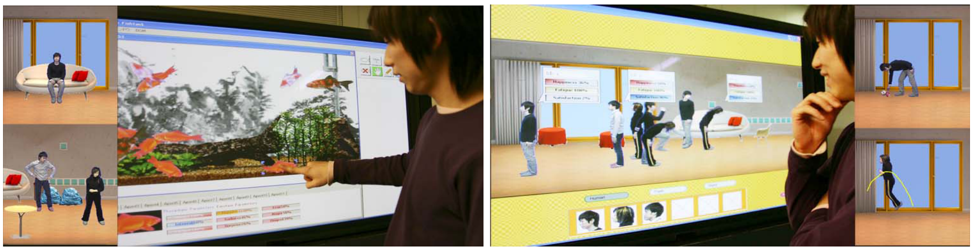

Figure 1: *Video Agents*

Abstract

This project establishes an interactive environment in cyberspace in which users interact with autonomous agents generated from video images of real-world creatures. Each agent has autonomy, personality traits, and behaviors that reflect the results of various interactions, which are determined by an emotional model.

CR Categories: I.3.4 [Computer Graphics]: Graphics Utilities—Virtual device interfaces; H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities

Keywords: video agent, autonomous agent, character animation, authoring tool

1 Introduction

1.1 Project Description

Our project establishes an interactive environment in cyberspace, based on video images taken from the real world. Each agent in cyberspace is an autonomous agent generated from video images of real-world creatures. Therefore, the agent is called a ”video agent,” who embodies such real-world creature characteristics as motion, shape, and texture that have been extracted from video using image processing techniques. The behavior of each autonomous agent is determined by an emotional model with fuzzy logic. Here, the emotions of each agent are generated based on its individual personality and physiological parameters, which are varied based on user interactions and events encountered in the environment.

After an agent's behavior is determined, a sequence of video images that best match the determined behavior is retrieved from the database in which a variety of video image sequences of the real creature's behaviors are stored. The retrieved images are applied to the agent to make it responsive; the image sequences are sometimes automatically edited to satisfy the requirements made by the emotional model.

Users can freely interact and play with the video agents in cyberspace. As a result, users participate in a form of entertainment in which autonomous video agents act with them. Through these experiences, we believe that users will learn significant yet common events that they failed to perceive earlier from the world that surrounds them. These events may come from their daily experiences in inter-personal communities or from other abstract experiences.

1.2 Vision

Our primary goal is to provide users with an entertaining cyberspace environment that is also educational. The system's entertainment value is derived from allowing users to easily construct interactive environments that include many of their favorite creatures as autonomous video agents. Users can create video agents based on their real-world pets and allow the video agents to live in cyberspace forever, even after the real-world pet is gone. The system's educational value stems from the knowledge users gain through interactive experiences.

Another goal of this project is producing a system that facilitates movie or cinema creation. Scenes of animals or humans moving or walking generated by our system will be a powerful tool to produce background scenes including extras in movies.

Finally, another goal of this project is creating an ecosystem in cyberspace for science, education, and relaxation. A cyberspace ecosystem can be based on a real environment; however, our ecosystem is interactive, and each creature that autonomously exists in it has an individual personality. Such a novel approach to interactive ecosystem simulation carefully addresses the fragile balance and tradeoff between the autonomy of the simulated ecosystem and

*e-mail: asai@ist.osaka-u.ac.jp
†e-mail: nishimoto.takashi@ist.osaka-u.ac.jp
‡e-mail: hirano.yoshinori@ist.osaka-u.ac.jp
§e-mail: hama.emiko@ise.eng.osaka-u.ac.jp
¶e-mail: kitamura@ist.osaka-u.ac.jp
‖e-mail: kishino@ist.osaka-u.ac.jp

freedom of user interaction. This project will be useful for future computer simulations of natural ecological systems (not only zoological but botanical environments) for science, education, and relaxation.

Today protecting the earth's environment is vital. Therefore, computer simulations of natural ecological systems have an increasingly important role in various fields. Moreover, sophisticated interactive simulation systems need to be established to assess the environment and to enlighten, educate, and expand the range of this research field. In the future, our computer simulation techniques for natural ecological systems will be used in both zoological and botanical environments for science, education, and relaxation purposes.

2 Previous Related Work

Creating a long continuous video stream by utilizing a set of relatively short video fragments is one of the hot topics in multimedia processing. For example, Schödl et al. [2000] proposed a technique to provide a continuous, infinitely varying stream of images by repeating video frames. They applied their technique to video sprite animations by using a number of video frames and defining transition costs [Schödl and Essa 2002]. For example, this technique was extended to systems creating cartoon animations [de Juan and Bodenheimer 2004]. Recently, much literature is being devoted to segmentation-based image cutout systems. For example, Wang et al. [2005] presented an interactive video cutout system using selected regions that form foreground objects in 3D video volumes. Starck et al. [2005] introduced video-based representation for free viewpoint visualization and motion control of 3D character models created from multiple view video sequences of real people. Motion graphs presented by Kovar et al. [2002] smoothly connected motions based on the analysis of motion capturing data.

Kombis et al. [1998] demonstrated a "Virtual FishTank" in which fish characters generated by computer graphics swim. Users choose the appearances and personalities of the fish by controlling such parameters as timidity or curiosity; however, the emotional or physiological models seemed undeveloped. A similar scheme using human characters can be found in Cube World [RADICA 2005]. Autonomous and interactive agents can be used in various applications, such as simulation systems, video games, and so on. AlphaWolves [Tomlinson et al. 2002] presented an interactive semi-autonomous system in which the behavior of wolves reflected their emotional states. Su et al. [2005] provided story characters whose postures were controlled by their personalities and emotions. Users were requested to make a rule-based program to create agent behaviors [Loyall et al. 2004].

In our work, we use a video processing technique similar to [Schödl et al. 2000]; however, we developed a unique feature-based extraction and classification technique for real creatures in actual video image sequences taken by a single camera to create video agents. In addition, the motions of video agents are generated based on their inherent personalities and emotions that reflected experiences in the created cyberspace. For this purpose, we implemented a unique emotional model. In addition, we extended an idea of [Kovar et al. 2002] to establish links among video frames having higher similarities in terms of motions and other features of video agents to obtain smooth motion transition. In our work agents are generated from real video images taken in real space. Since each agent's behavior is determined by emotional models with fuzzy logic, our work shows a variety of behaviors. Images representing agent behaviors are retrieved from a video database and displayed after appropriate special effects are added. Finally, our work provides a graphical user interface with which users can intuitively create content with autonomous video agents.

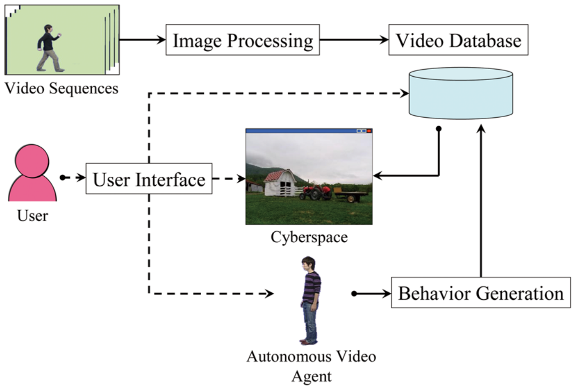

Figure 2: *System overview*

3 Technical Innovations

3.1 System Overview

The most important feature of our work is that agents are generated from live video taken from the real world. Our work successfully integrates such diverse technologies as image processing, computer vision, database, artificial intelligence, and post-production. Based on these technologies, a graphical user interface allows users to create content with video agents without making program code. Figure 2 shows an overview of our system. The following are the core technical innovations:

- Image Processing: extraction of a target creature in each frame of video sequences, segmentation of the video sequence based on feature values of the extracted target creatures, and establishment of links among video segments.
- Behavior Generation: behavior of each autonomous agent is determined by an emotional model with fuzzy logic, and personality is implemented based on well-known five big factors.
- Video Database: automatic generation/maintenance of a video database and automatic editing of retrieved video segments based on generated behavior.
- User Interface: a graphical tool to create video agents, and an authoring tool to create an interactive cyberspace in which autonomous video agents move.

3.2 Image Processing

We suppose sequences of real video images taken in real space where a creature's moves are given to the system. Our system's image processing flow, which is shown in Figure 3, consists of three steps. First, the creature is extracted in each frame of the video sequence. Here we use simple background subtraction by assuming an ideal environment with a blue background to simplify the problem. For environments with more complicated backgrounds, a sophisticated image processing technique is used. In the second step, segmentation of the video sequence is conducted based on the

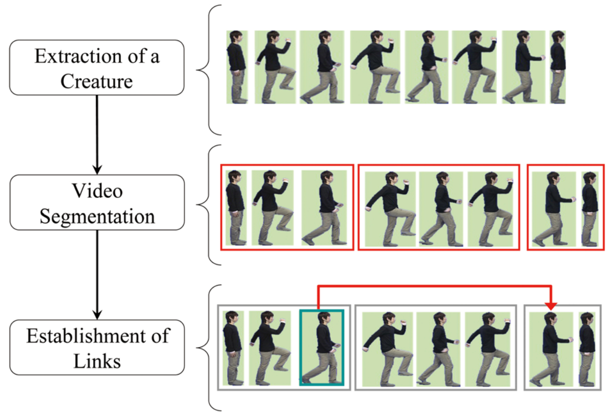

Figure 3: *Flow of image processing*

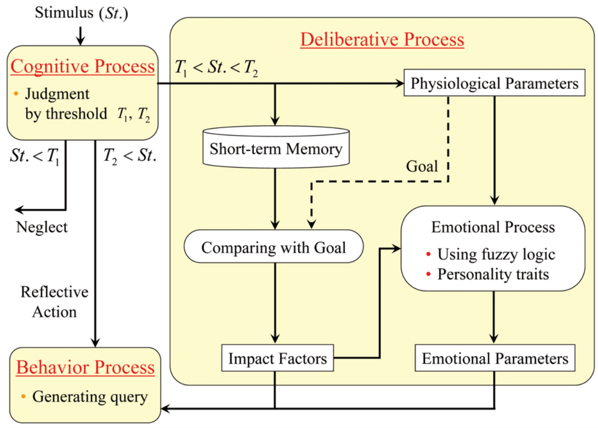

Figure 4: *Architecture of emotional model*

feature values of the motions of the extracted target creatures. Finally, in the third step, links among video segments are established by calculating similarities of the motions and other features of the extracted target creatures in the segments. This idea of establishing links is expanded from [Schödl et al. 2000]. Similarity is calculated by:

$$S_{total} = \frac{S_m + S_g + S_c}{3}$$

where S_m is the similarity of motion vectors, S_g is the similarity of intensity, and S_c is the similarity of the target creature's contours. Here, S_g and S_c are calculated by algorithms of Intel OpenCV. When two images, I_a and I_b, are compared, similarity S_m between I_a and I_b is calculated as follows:

$$S_m = \left(\frac{D_a - D_b}{D_{MAX}} - \frac{A_a - A_b}{180} \right) \times 0.5$$

where $(D_a, D_b) and (A_a, A_b)$ are motion vectors; the former is the length and the latter is the angle (degree) to the axis.

Finally, the video segments are stored in the video database with calculated feature values and established links. In addition, a user can give adequate annotation to each video segment stored in the database for quick and easy access.

3.3 Behavior Generation

As shown in the demo movie, each video agent has its own individual autonomy and is driven by an emotional model with fuzzy logic. The model effectively uses several internal sensors to detect changes in the environment and the physiological parameters of the agents themselves. The architecture of the emotional model is shown in Figure 4. It consists of cognitive, deliberative, and behavior processes.

The cognitive process, where all stimuli given to the agent serve as input, determines whether the stimuli are perceived or neglected. Here, determination is performed by comparing stimulus strength with preliminary determined thresholds (described as T_1 and T_2 in Figure 4). Stimuli smaller than T_1 are ignored. Stimuli larger than T_1 and simultaneously smaller than T_2 are passed to the deliberative process. However, when larger than T_2, stimuli are directly passed to the behavior process to generate reflective action.

In the deliberative process, stimuli that passed from the cognitive process are stored in the short-term memory. Physiological parameters, which are those having strong relationships with instincts (i.e., hunger, fatigue, and interest), are computed based on stimulus affects. Also, the stimuli stored in short-term memory are compared with a goal, and impact factors of the stimuli are calculated. Here the goal is a short-term purpose to increase the value of physiological parameters. Similarly, emotional parameters are computed using fuzzy logic based on the calculated impact factors of the stimuli and physiological parameters. Resembling [Ekman and Davidson 1994], we adopt agents with six emotional parameters. Moreover, to achieve personality traits for the agents, the so-called big five factors [Costa and McCrae 1985] are adopted. Each personality trait takes a value of three levels. By combining these personality traits, the membership functions used by fuzzy logic are varied. The impact factors and the emotional parameters computed for each stimulus are passed to the behavior process.

In the behavior process, a video agent's behavior is determined based on the information passed from the deliberative process. The behavior is basically determined by the stimuli, emotional parameters, and impact factors. Finally, based on the determined behavior, a query is generated to retrieve an adequate video segment from the video database.

Consequently, video agents portray a variety of behaviors based on environment variations and interactions among agents and with users.

3.4 Video Database

A sequence of images of the video segments that best match the behavior determined in the behavior generation step is retrieved from the database. Here, annotations, motion vectors, and feature values of all images in the video segments in the database are compared. The retrieved sequence of the video segment is applied to the action of the video agent. Extracted image sequences may sometimes be automatically edited to satisfy the requirements made by the internal model.

3.5 User Interface

As shown in Figure 5, two tools are implemented, i.e., a graphical tool to create video agents and an authoring tool to create an interactive cyberspace in which autonomous video agents move. In addition we have an execution program with which users play interactively.

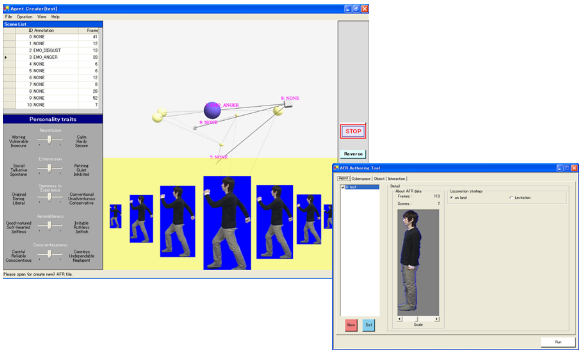

Figure 5: *User Interface*

Agent Creator Contents in the video database, i.e., sequences of images of video segments, are visualized as a node-and-link 3D diagram based on the TKS method [Kruskal and Seery 1980] and can be interactively manipulated. Each node shows a piece of a video segment, and a link between nodes shows a link among the video segments established by calculating the similarities of the motions and other features of the extracted target creatures in the segments. We have two different display modes, i.e., whole view and detailed view. In the former mode, a user can observe the entire structure of the diagram. In the latter mode, only the selected node and nodes that have links to the selected node are displayed. When a user selects the other node, the diagram layout is updated with the displayed transition animation. Finally, the output of the agent creator is a diagram stored in a format called agent data.

Cyberspace Authoring Tool This authoring tool creates an interactive cyberspace in which autonomous video agents move. The agents and objects that appear in cyberspace are registered. Varieties of interaction among agents, users, and objects in the environment are also defined.

Interactive Cyberspace Execution The created cyberspace is executed with adequate I/O devices in which autonomous video agents move. Users interactively play with the cyberspace.

4 User Experience

SIGGRAPH 2007 attendees can have three different experiences based on their interests. First, they can enjoy an interactive experience in cyberspace generated from real-world video images using interactive displays with touch overlays. SIGGRAPH attendees will also be able to experience two example works of autonomous agents: fish and human agents. We also plan to develop other examples including dogs, cats, and elephants in a zoo that may be available for SIGGRAPH attendees. Attendees' interactions will include copying, deleting, dragging agents, and so on. In addition, they can draw a trajectory along which an agent can move and manipulate such spatial obstacles as rocks or chairs. Attendees can also customize their cyberspace by changing its parameters including brightness and temperature. For human agents, interactions resemble the case of fish; however, the agents will be more emotional. Therefore, SIGGRAPH 2007 attendees will enjoy more profound interactions. Second, attendees can create contents by using the graphical tool to create video agents and/or the authoring tool to create an interactive cyberspace in which autonomous video agents move. Third, attendees will also have the opportunity to act as video agents if they desire. In front of a blue background, their actions are recorded by video camera and stored in the video database through the image processing described mentioned above.

References

COSTA, P. T., AND MCCRAE, R. R. 1985. *The NEO Personality Inventory*. Psychological Assessment Resources.

DE JUAN, C., AND BODENHEIMER, B. 2004. Cartoon textures. In *Proceedings of the 2004 ACM SIGGRAPH/Eurographics symposium on Computer animation*, ACM Press, New York, NY, USA, 267–276.

EKMAN, P., AND DAVIDSON, R. J. 1994. The nature of emotion: fundamental questions. Oxford University Press.

KOMBIS, S. 1998. Virtual fishtank. In *ACM SIGGRAPH 98 Conference abstracts and applications*, ACM Press, New York, NY, USA, 116.

KOVAR, L., GLEICHER, M., AND PIGHIN, F. 2002. Motion graphs. In *Proceedings of ACM SIGGRAPH 2002*, ACM Press, New York, NY, USA, 473–482.

KRUSKAL, J. B., AND SEERY, J. B. 1980. Designing network diagrams. In *Proceedings 1st General Conference on Social Graphics*, 22–50.

LOYALL, A. B., REILLY, W. S. N., BATES, J., AND WEYHRAUCH, P. 2004. System for authoring highly interactive, personality-rich interactive characters. In *Proceedings of the 2004 ACM SIGGRAPH/Eurographics symposium on Computer animation*, ACM Press, New York, NY, USA, 59–68.

RADICA, 2005. Cube world. http://www.radicauk.com/.

SCHÖDL, A., AND ESSA, I. A. 2002. Controlled animation of video sprites. In *Proceedings of the 2002 ACM SIGGRAPH/Eurographics symposium on Computer animation*, ACM Press, New York, NY, USA, 121–127.

SCHÖDL, A., SZELISKI, R., SALESIN, D. H., AND ESSA, I. 2000. Video textures. In *Proceedings of ACM SIGGRAPH 2000*, ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 489–498.

STARCK, J., MILLER, G., AND HILTON, A. 2005. Video-based character animation. In *Proceedings of the 2005 ACM SIGGRAPH/Eurographics symposium on Computer animation*, ACM Press, New York, NY, USA, 49–58.

SU, W.-P., PHAM, B., AND WARDHANI, A. 2005. High-level control posture of story characters based on personality and emotion. In *IE2005: Proceedings of the second Australasian conference on Interactive entertainment*, Creativity & Cognition Studios Press, Sydney, Australia, Australia, 179–186.

TOMLINSON, B., DOWNIE, M., BERLIN, M., GRAY, J., LYONS, D., COCHRAN, J., AND BLUMBERG, B. 2002. Leashing the alphawolves: mixing user direction with autonomous emotion in a pack of semi-autonomous virtual characters. In *Proceedings of the 2002 ACM SIGGRAPH/Eurographics symposium on Computer animation*, ACM Press, New York, NY, USA, 7–14.

WANG, J., BHAT, P., COLBURN, R. A., AGRAWALA, M., AND COHEN, M. F. 2005. Interactive video cutout. In *SIGGRAPH '05: ACM SIGGRAPH 2005 Papers*, ACM Press, New York, NY, USA, 585–594.