AR Puppet: Animated Agents in Augmented Reality

István Barakonyi   *Dieter Schmalstieg*

Vienna University of Technology
{ bara | schmalstieg } @ ims.tuwien.ac.at

Abstract

AR Puppet is a hierarchical animation framework for Augmented Reality agents, which is a novel research field connecting Augmented Reality (AR), Sentient Computing and Animated Agents to a single coherent human-computer interface paradigm. This framework enables the control of the physical environment and virtual animated agents using the same interface, blurring the boundary between the real and virtual world. Any physical object that can be monitored or controlled by a computer can be incorporated into the framework and used in AR applications. AR Puppet explores the requirements for context-aware animated agents concerning visualization, appearance, behavior, as well as associated technology and application areas. The framework is built on a powerful collaborative AR middleware and it allows for experimentation with a wide range of applications and many different characters – physical and virtual alike.

1 Introduction

Body and facial gestures and speech are familiar and widely accepted means of human communication. Many people would prefer a knowledgeable colleague to a lengthy manual when learning to use the new fax machine in the office or a kind, helpful person to an automated guidance system to show the way through an unknown building. Animated characters, often with autonomous and affective behavior, have been recently researched as an interface to computerized systems to bridge the communication gap between man and computer, and the real and virtual world. Augmented Reality (AR) (Azuma, 1997) applications are aiming at the same goal by enhancing our real environment with useful virtual information, where virtual objects appear to coexist with the real world. To perceive this "mixed reality" world users may wear Head-Mounted Displays (HMD), carry camera-equipped mobile devices to provide an "augmented window" onto the real world, or watch monitors and projection screens connected to computers that are embedded into our surroundings. In Sentient Computing systems (Addlesee et al., 2001) computing devices maintain a model of the world obtained through sensors, which allows applications to react to changes in the environment in accordance with the user's perception of the world.

Our new research field, which we call Augmented Reality Puppet (AR Puppet), combines AR, Sentient Computing and Animated Agents to create a new, coherent human-computer interface paradigm. AR Puppet explores the implications of an AR scenario on animated agents concerning their appearance, behavior, application areas and associated technologies.

2 Related Work

Virtual Reality (VR) is a more mature field than AR and consequently already explored new interaction techniques involving animated agents, therefore it can provide many useful ideas. One of the outstanding examples in VR is the Jack animation system (Noma et al., 2000) that allows for using animated virtual human figures in a wide range of situations from military trainings to virtual presentations. The Improv system (Perlin and Goldberg, 1996) creates a novel interactive theater experience with real-time virtual actors on a virtual stage. The autonomous pedagogical agent called Steve (Rickel and Johnson, 1998) operates as a virtual trainer in an immersive VR environment presenting complex interactive, educational machine maintenance scenarios. The idea of Steve recurs in an AR setting in the EU project, STAR (Vacchetti et al., 2003), which aims at enhancing service and training in real factory environments using virtual humans.

MacIntyre et al. make the point that a new media, such as AR, starts to gain wider public acceptance once it entered the game, art and entertainment domain (MacIntyre et al., 2002). Their interactive theater experience places prerecorded video-based actors into an AR environment. As they demonstrate, AR has started to step beyond the usual instructional and presentational domain and explore new fields of application. Cavazza et al. put a live video actor of a real person into a Mixed Reality setting (Cavazza et al., 2003), and interact with a digital storytelling system with body gestures and language commands. Cheok et al. also experiment with Mixed Reality entertainment with live captured 3D characters (Cheok et al., 2002). They integrate Ubiquitous Computing, tangible interaction and Social Computing within a Mixed Reality space. ARQuake (Piekarski and Thomas, 2002) recreates the famous first-person shooting game in a real campus setting using a mobile AR setup, where the user has to shoot virtual monsters lying in ambush behind real buildings, and uses a tangible interface to fire virtual weapons. Balcisoy et al. experiment with interaction techniques with virtual humans in Mixed Reality environments (Balcisoy et al. 2001).

3 Motivation and research goal

With the exception of the work of MacIntyre et al., who are developing a general animation framework (MacIntyre and Gandy, 2003) for AR applications, all of the related researches are bound to a single domain or technology and lack a general approach to create a reusable framework.

Our goal is to turn physical objects, including everyday items like a printer or a digital piano, into context-aware animated agents that perform various tasks together with digital actors, virtual presenters and other synthetic visual elements. The visualization of the agents is however not limited to anthropomorphic forms. A lesson learned from Microsoft's "intelligent" staple is that human-like behavior may sometimes be more annoying than useful. A simple arrow may communicate more information than a fully articulated human character, therefore we experiment with various character forms.

Figure 1. Physical, augmented and virtual representations of
a LEGO Mindstorms® robot as an AR Puppet

AR Agents – or AR Puppets – can be represented by physical, virtual or augmented objects or combinations thereof (see Figure 1 for an example). The physical representation means that any object can be turned into an AR Puppet that has a communication channel to a computer and can consequentially be monitored and controlled by it. Elements of the physical environment can be used as "stage" and "actors" as well as they can serve as input and output devices. They accept commands and are able to react to them and send feedback. The current pose of the physical objects is tracked. The virtual representations are synthetic visual and audio elements independent of the real world: 2D graphical information (labels, pictures, videos) or animated 3D models. Augmented representations are combinations of the physical and the virtual. They assume the presence of a physical representation and only superimpose necessary virtual information. Virtual agent representations may and augmented agent representations need to have an associated tracked physical object, which serves as a tangible control interface already familiar for us from our everyday life. The virtual and augmented objects can be rendered on several platforms from PCs to mobile devices and on a wide range of display types from monitors through HMDs to mobile phones. Consider a machine maintenance scenario, where a printer accepts a print command from the user and responds with an "empty tray" error message. The physical representation handles and executes printer messages. The virtual counterpart on the computer screen shows an animation using a 3D model how to remove the tray from the printer. The augmented representation highlights the tray on the real printer, places a label with the error message above it, while a virtual technician imitates the removal of the tray on the real machine with gestures.

AR Puppets and users are always associated with a context determined by their current and relative pose, behavior, decisions and the application(s) in which they are embedded. For AR applications the most important context element is the physical context. Through this physical context (just like in Sentient Computing) the animated agents use the real world as an interface to applications. Changes in the environment generate events to which the agents can react. AR Puppets are constantly aware of the current context. This information can be exploited when making decisions based on past contexts stored in a database such that characters adapt to users, environments and applications.

4 AR Puppet framework

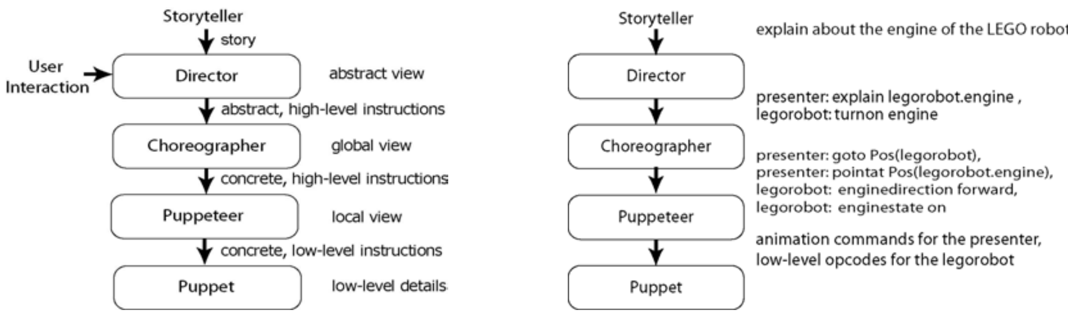

Figure 2. (a) AR Puppet Framework (b) Command flow in the framework

Our hierarchical animation framework uses a puppet theater metaphor, that is the reason why we call our Augmented Reality characters “AR Puppets”. We chose the multilevel structure shown in Figure 2 to accommodate the needs of controlling physical and virtual objects with a standard interface. At the lowest level of the framework the puppets handle the physical representation of an AR Puppet (low-level commands and feedback), and various virtual and augmented representations appearing on different devices and display types (visual appearance). In the puppet theater skilled puppeteers animate their puppets and know which strings to pull to express their thoughts. Every puppet has its own puppeteer that masters its control. The Puppeteer component handles a group of puppets, the different representations of an AR Puppet. It is responsible for communication, animation, etc. Many functions like precise, low-level instructions for the puppets, tracking support, standard command interface and command synchronization are also implemented here. The next level is the choreographer, which, unlike the puppeteer, has an overview of the whole stage and all the characters and scene elements. The Choreographer component is aware of what the puppeteers are doing and gives high-level commands to them (e.g. motion planning, group behavior, etc.) and assures synchronization between character commands. It can choreograph the higher-level commands from the director and choose the appropriate animation, speed, path, and emotion. The director is the story engine, a state machine that implements the story and executes high-level abstract commands. All the components can be configured dynamically, even during a running story. User interaction events influence the story flow and thus the behavior of the characters.

5 Monitoring applications and users

It has been always a challenge for interface agents to monitor the current state of the application they are embedded into, and the behavior of the users they are interacting with, without altering the application itself. One of the most powerful aspects of the AR Puppet framework is the easy way to monitor the current applications and user state. The framework is built-on the Studierstube collaborative AR platform (Schmalstieg et al., 2002), which is a powerful middleware allowing for a wide of range of multi-user AR applications. On this platform applications and users are represented as “black-box” objects just like the AR Puppet components, and interface one another via input and output attributes.

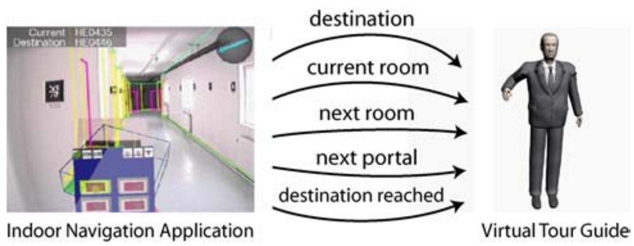

Figure 3. Indoor AR navigation system interacting with a virtual tour guide

Users have fix, standard attributes like current tracked pose and display type. Their actions can be monitored through tracked interaction devices. Animated agents in the AR Puppet framework are thus always aware of the application users’ pose and behavior that generates a large number of new events that may control agent behavior.

The interface to applications is defined by the developer who packs application functionalities into one component and defines attributes relevant to the outside world and help track the current application state. Figure 3 shows an example, where an indoor AR navigation system exposes its attributes to a virtual tour guide agent so that it can deliver location-based information. The application outputs the selected destination of the user, the current and next room in the route suggested by the system, the next portal to go through and a flag indicating whether the user has reached its target location. By watching these attributes the virtual tour guide can deliver location-based descriptions about the current room and useful navigation information. Using hand and head gestures it is able to show the right direction, point out locations of interest in the building, is able to warn users when a door is approaching, and send a notification when the destination has been reached.

6 Applications

The Studierstube collaborative AR middleware allows for experimentation with different tracking technologies and a wide range of distributed, multi-user applications. Three applications have been developed to demonstrate the capabilities of our framework (see Figure 4 for screenshots).

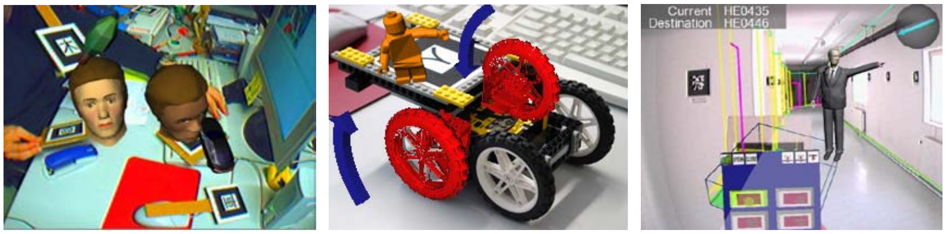

Figure 4. Screenshot of the applications: kanji teaching, AR Lego® and virtual indoor tour guide

6.1 Kanji teaching

The first demo application teaches Japanese kanji characters to language students. The language teachers are represented by two animated affective facial agents featuring various facial gestures and synthesized speech with lip synchronization. The text-to-speech engines and the tutorial text can be configured to support multiple languages, for example English and Japanese. The task of the user is to guess the meaning of kanji characters appearing inside ARToolKit optical markers, which are used as tangible interfaces to position the talking heads in the physical world. When the student is instructed to find a kanji, he/she has to choose an optical marker with a symbol. Initially a virtual question mark is floating above the marker. Placing the chosen kanji marker near a teacher's marker indicates that the selected solution is submitted to that teacher, who will answer in its own language. This is an intuitive way to obtain the English meaning from the English, and the Japanese pronunciation from the Japanese teacher. In case the correct answer has been given, a virtual object conforming to the meaning of the character appears above the marker and gets associated with it (e.g. a virtual car model is displayed above the kanji of the car) so that students are able to remember their answer. The progress of the user is monitored and commented by the facial agents with speech and facial expressions.

6.2 Augmented LEGO® robot

The second application implements a machine maintenance scenario using a LEGO Mindstorms® robot and a presentational agent. The demo explains how the robot works: which motors are moving, which cogwheels are turning in which direction, how the light and the touch sensor are used, etc. The LEGO® robot is a full-featured AR Puppet since all the three representations are used in the application. The physical representation of the robot communicates with the computer via an infrared link. If the real robot is present and functional, the computer senses its "alive" messages and displays the augmented representation with virtual information over the real robot. This virtual information include highlighting engines that are currently turned on, indicating the cogwheel's actual direction with virtual arrows, revealing hidden parts allowing for an "X-Ray view", and labeling buttons and other important functional parts with explanatory text. If the physical robot is unavailable or malfunctioning, a virtual mock-up model is displayed instead, which copies the behavior of its real counterpart. An anthropomorphic animated LEGO® minifig character explains about the different robot parts and functions, its gestures (like walking and pointing) are synchronized with the actions of the robot.

6.3 Virtual Tour Guide

The third demo embeds a virtual human acting as a tour guide into a mobile indoor navigation application (already mentioned in Section 5) called SignPost. The user wears a mobile AR setup and perceives the augmented world through an HMD. While she is walking around with the setup, the system guides her in the selected destination. The virtual tour guide agent is placed in the reference frame of the real building and explains about the featured demos of the different labs in our institute with body gestures (e.g. looking, pointing), 2D and 3D visual elements and sound. In this application the tour guide has only a virtual representation, and is controlled by the AR Puppet of the physical environment, which has a real and an augmented representation.

7 Conclusion and future work

We have created a hierarchical animation framework for Augmented Reality agents called AR Puppet, which combines several research areas into a single novel field. We have made the following contributions to the animated agent research field:

- Physical objects can be easily turned to animated agents and used as input and output devices in AR environments via a unified command interface. Consequently, the agents of AR Puppet can not only talk back but hit back as well.
- AR Puppet is the first general framework for animated agents that have been developed specifically for AR applications. The underlying middleware allows for experimenting with a wide range of applications, tracking technology and display devices.
- AR Puppets are interchangeable because of the unified command interface and default command implementations, therefore it is simple to try the same story with different types of characters.
- It is easy to monitor the current status of the application and users. Events generated by the framework components drive the application forward.
- AR Puppets are dynamically configurable, their attributes and command fields can be intuitively controlled even during a running application.

We believe that Augmented Reality applications bring new challenges and fresh perspectives for the animated agent research community. In the near future the storing and recalling contexts will be implemented focusing especially on character adaptation observing the user's obtained knowledge, cultural aspects, etc. The interaction with the characters will be enhanced with multimodal communication including speech recognition and gestures. Many application ideas such as augmenting everyday items like a fax machine, educational scenarios and games will be realized together with an extensive user evaluation.

8 References

Addlesee, M., Curwen, R., Hodges, S., Newman, J., Steggles, P., Ward, A., Hopper, A. (2001). Implementing a Sentient Computing System, *IEEE Computer* 34, 8, 50-56.

Azuma, R. (1997). A Survey of Augmented Reality. *Presence: Teleoperators and Virtual Environments* 6, 4, 355-385.

Balcisoy, S., Kallmann, M., Torre, R., Fua, P., Thalmann, D. (2001). Interaction Techniques with Virtual Humans in Mixed Environments, *Proc. of International Symposium on Mixed Reality*, Tokyo, Japan

Cavazza, M., Martin, O., Charles, F., Mead, S.J., Marichal, X. (2003). Interacting with Virtual Agents in Mixed Reality Interactive Storytelling. *Proc. of Intelligent Virtual Agents*, Kloster Irsee, Germany

Cheok, A. D., Weihua, W., Yang, X., Prince, S., Wan, F. S., Billinghurst, M., Kato, H. (2002). Interactive Theatre Experience in Embodied and Wearable Mixed Reality Space. *Proc. of International Symposium on Mixed and Augmented Reality* (ISMAR'02), Darmstadt, Germany

MacIntyre, B. and Gandy, M. (2003). Prototyping Applications with DART, The Designer's Augmented Reality Toolkit, *Software Technology for Augmented Reality Systems Workshop* (STARS 2003), Tokyo, Japan

MacIntyre, B., Bolter, J., Vaughan, J., Hannigan, B., Moreno, E., Haas, M., and Gandy, M. (2002). Three Angry Men: Dramatizing Point-of-View using Augmented Reality. *SIGGRAPH 2002 Technical Sketches*, San Antonio, TX

Noma, T., Zhao, L., Badler, N. (2000). Design of a Virtual Human Presenter, *IEEE Computer Graphics and Applications*, 20, 4

Perlin, K., and Goldberg, A. (1996) Improv. A System for Scripting Interactive Actors in Virtual Worlds. *Proc. of SIGGRAPH '96*, 205-216.

Piekarski, W., and Thomas, B. (2002) ARQuake: The Outdoor Augmented Reality Gaming System. *ACM Communications* 45, 1, 36-38.

Rickel, J., Johnson, W. (1998). Steve: A pedagogical agent for Virtual Reality. *Proc. of the 2nd Int'l Conf. on Autonomous Agents*, 332-333.

Schmalstieg, D., Fuhrmann, A., Hesina, G., Szalavári, Zs., Encarnação, M., Gervautz, M. and Purgathofer, W. (2002). The Studierstube Augmented Reality Project. *PRESENCE -Teleoperators and Virtual Environments*, MIT Press

Vacchetti, L., Lepetit, V., Papagiannakis, G., Ponder, M., Fu, P. (2003). Stable real-time interaction between virtual humans and real scenes. *3DIM 2003*, Banff, AL,Canada