Geppetto: An Environment for the Efficient Control and Transmission of Digital Puppetry

Daniel P. Mapes, Peter Tonner, and Charles E. Hughes

University of Central Florida, 4000 Central Florida Blvd.
Orlando, FL 32816-2362 USA
dmapes@ist.ucf.edu, ptonner@knights.ucf.edu, ceh@cs.ucf.edu

Abstract. An evolution of remote control puppetry systems is presented. These systems have been designed to provide high quality trainer to trainee communication in game scenarios containing multiple digital puppets with interaction occurring over long haul networks. The design requirements were to support dynamic switching of control between multiple puppets; suspension of disbelief when communicating through puppets; sensitivity to network bandwidth requirements; and as an affordable tool for professional interactive trainers (Interactors). The resulting system uses a novel pose blending solution guided by a scaled down desktop range motion capture controller as well as traditional button devices running on an standard game computer. This work incorporates aspects of motion capture, digital puppet design and rigging, game engines, networking, interactive performance, control devices and training.

Keywords: Digital puppetry, avatar, gesture, motion capture.

1 Introduction

Digital puppetry refers to the interactive control of virtual characters [1]. The artistry of using a human to control the actions of animated characters in real-time has been employed for several decades in a number of venues including television (Ratz, Muppets [2]) and location-based entertainment (Turtle Talk with Crush [3]). More recently, these techniques have found application in interactive networked simulations [4,5].

By affording responsive and adaptive behaviors while operating through the changeable context of an invented digital character, puppetry greatly increases the effectiveness of interactive experiences. Not unexpectedly, these benefits come with demanding requirements both from the puppeteers and from the technical aspects of projecting the simulation. From the puppeteer, there is the cognitive load associated with rapid switches between puppets; maintaining individual puppet behaviors during switches; following training scenario scripts; operating scenario specific controls; and operating puppet specific behavior controls. From a technical aspect, there is the need to reliably and efficiently capture and transmit behaviors over a long haul network; the necessity to achieve this transmission even in the presence of limited resources,

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 270–278, 2011.

including bandwidth and computational power; and the requirement to articulate these actions at the client site, providing smooth interactive performance.

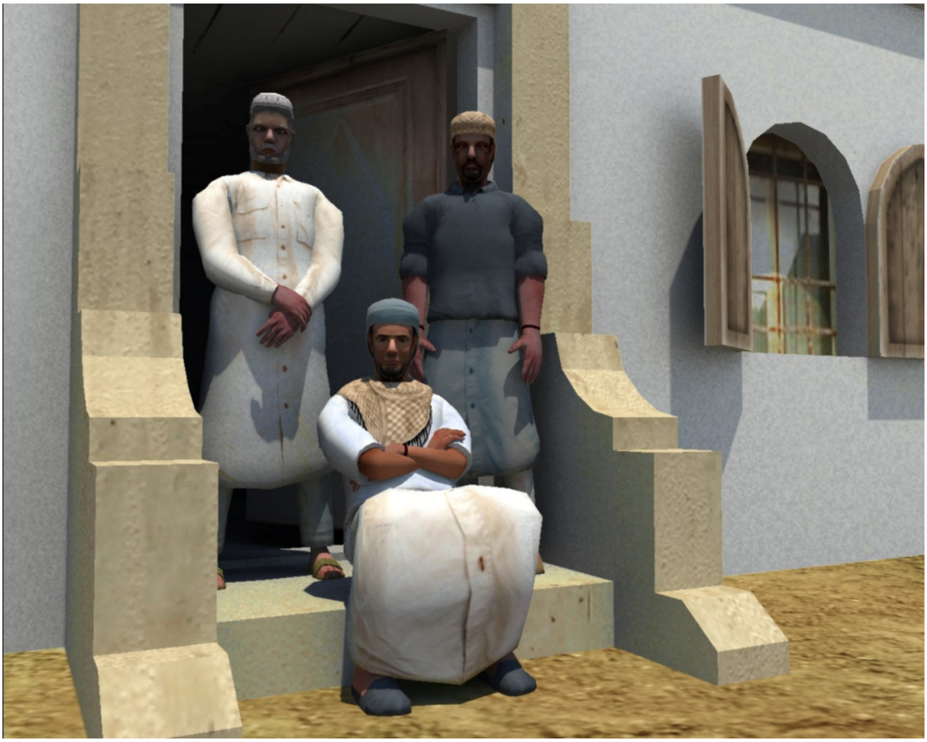


Fig. 1. Pashtun avatars used in cross-cultural training

In our paper we present an evolution of interaction paradigms that have made projected puppetry more intuitive and less cognitively draining while reducing network bandwidth requirements through a micro pose blending technique guided by body worn 3D controllers. The use of pose blending reduces visual artifacts of live motion captured puppetry; supports culturally appropriate gestures; and significantly reduces the streaming transmission bandwidth requirements over long-haul networks. Essentially, our contribution is interactive, affordable, reliable, believable, culturally attuned long-distance digital puppetry.

2 Social Contexts and Puppetry

2.1 FPS Gaming

One of the most prevalent and affordable examples of projected puppetry applications is the networked first person shooter game where individual players log in online to control a puppet that can do a fixed set of behaviors such as walk, run, duck, jump, punch, kick, aim and shoot using devices such as game controllers, keyboards and mice. While the process of pressing a button to control a character may not be initially intuitive, devoted players to these games have become quite adept at using these controls under the pressure of competitive game play. The control data being transmitted over long-haul worldwide networks from the buttons and dual axis joysticks found in typical controllers is insignificant compared to the incoming game state stream or the streaming VOIP that players use to collaborate during play.

The ability for players to collaborate in teams led to the emergence of Machinima where remote participants could interact with each other to create short playful movies where each player acted out a predefined role, while the results were recorded through the game camera. Though the control model allowed for small efficient network streaming, with only the ability to run, walk, duck, jump, aim and shoot it was fairly clear that the game characters lacked the more complex facial expressions and body language to do live acting.

2.2 Second Life

In an online social networking game like Second Life over 120 user-playable puppet animations are provided by default with the option of players increasing this set with their own custom animations [6]. These are used much like emoticon symbols embedded in a text document. While this is an improvement over the FPS game control by showing both facial emotion and body language, even if a puppeteer were able to master such a sizable abstract mapping, the performance still lacks the subtle hand and body motions used to communicate non-verbally and to punctuate speech.

3 Evolution of Digital Puppetry in Geppetto

3.1 Pre-service Teacher Training

The TeachME™ experience provides pre-service and in-service teachers the opportunity to learn teaching skills in an urban middle school classroom composed of five puppets, each with its own prototypical behavior. Teachers use the virtual classroom to practice without placing “real” students at risk during the learning process [4]. Overall, the concepts developed and tested with the TeachME environment are based on the hypothesis that performance assessment and improvement are most effective in contextually meaningful settings.

In the design of TeachME the puppets (avatars) were always seated at desks which allowed us to focus only on upper body motions. We captured head, shoulder, elbow and waist orientations by placing 9 retro-reflective markers on the puppeteer and tracking them as 3D points using a 4 camera IR motion capture system. Four basic facial poses, smile, frown, wink and mouth open were simultaneously blended using the finger bend sensors of a CyberGlove [7]. Lip sync through audio evaluation was tried and rejected due to excessive evaluation period latency and feedback from the puppeteers on their need for more direct control. An Ergodex game pad was used to allow the puppeteer to rapidly switch control into one of the five middle school age puppets and provide a custom key interface to control the scenario. A Skype connection was opened up to provide bi-directional audio between the trainer and trainee as well as a trainer webcam video feed of the trainee.

3.2 Avatar

The Avatar project used digital puppets (Figure 1) to practice cross cultural communication skills. Two scenarios were created, one in the Pashtun region of Afghanistan and one in an urban setting of New York. Each scenario had three puppets: one elder leader, a young idealist, and an antagonist. In each scenario the participant was to practice successful negotiation with the trained Interactor operating through the puppets.

In order to minimize the limitations from live mocap in the TeachME project, a micro-pose system was designed where each puppet contained a set of full body end poses (Figure 2) that were characteristic of how each puppet might communicate through hand and body posture. A scaled down dual IR camera motion capture system was used to track the Interactor's head orientation and the location of the actor's wrists. A calibration process was created where each characteristic puppet pose was presented to the Interactor who was then asked to mimic the pose. A database of their real world wrist locations and the puppet's wrist locations was built and a routine was written to use the Interactor's wrist locations to find the two closest poses from the puppet with a confidence weight on each pose. These two poses were then blended in proportion to their weights along with previously used poses decaying in weight to zero over time. Facial pose was controlled using the same five-finger glove metaphor from TeachME. The Ergodex game controller was replaced with a WiiMote held in the non-gloved hand.

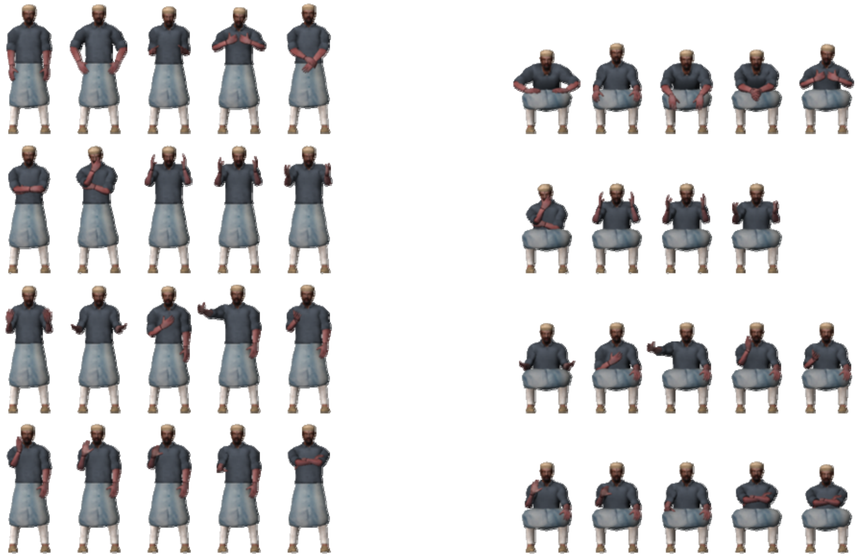


Fig. 2. Key poses for avatars

The end result was a puppet designed to exhibit only culturally appropriate gestures where these gestures could be controlled seamlessly using 3D locators held in each of the Interactor's hands while the Interactor moved the puppet's head directly by their own head movement. Since the puppet poses and the controlling hand locations were mapped nearly one-to-one, this system was almost as intuitive as raw mocap but with fewer motion artifacts.

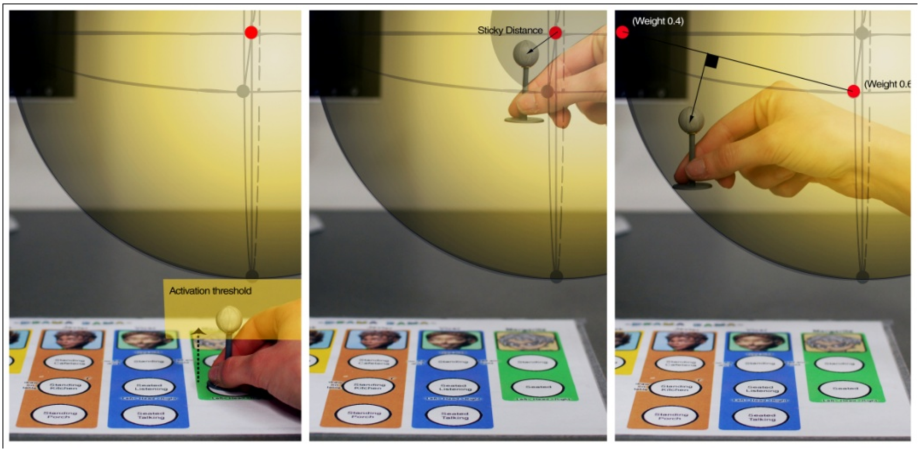
The Avatar effort improved the speed of puppeteering by feeding a much smaller vector of data points (two wrist locations) into the animation database. This addressed issues in previous studies where motion capture was stored in a database of animations and full motion capture was used to map into this database in real-time. Incorporating more animations would require scaling the database appropriately and optimizations to mapping of the mocap data to these animations [8]. In addition, all aspects of the pose, including hand and finger postures, could be included where the raw mocap system had only flat unexpressive hands. Another clear benefit of this system over raw motion capture was the realization that only the poses and their relative weight need be communicated over the Internet to describe the puppet state. Instead of streaming 17.28 mbps of skeletal joint angle data, this pose blended system required on the order of 1.44 mbps. While this is not significant for projecting a single puppets performance, when social networking games and machinima servers emerge supporting 100's of puppets the need for performance compression will become more apparent.

After the initial trials with this system, with only 3D points at their wrists, the puppeteers complained that the selection criteria was too ambiguous to find some of the key gestures under the pressure of performance. An example might be a hand held palm up or palm down. Another minor problem was the need to maintain individual calibration settings for each puppeteer as they changed shifts.

3.3 Evolving Avatar

In an effort to further minimize the need for full body motion capture and give puppeteers the control they requested, the representative poses from the initial study were sorted into bins based on modes within which the puppeteers were operating. Many of the poses were designed to be used together and, once separated, the blending became trivial to control with two points; but now a method for choosing which bin to operate within needed to be found. Button events from either an Ergodex or WiiMote were considered but this added to the buttons already being used to control the scenario and to switch between puppets.

The solution that evolved was a hybrid between motion capture and a game controller solution, where the need to switch between puppets and different puppet pose spaces were combined. A single 3D point from the motion capture system was used as a controller within a virtual 3D volume above a desk surface. A printed map of each of the pose bins of all of the puppets was attached to the desk surface. Placing the marker below a fixed height above the desk surface put all puppets into an automated non-puppeteered state (Figure 3a) and lifting the marker above a pad designated which puppet was to be controlled as well as the set of poses to be blended. When lifted above a pad the controller values were converted to a vector relative to a sphere of fixed radius and centered above the pad. The vector was



(a) (b) (c)

Fig. 3. Using marker to select character and behavior genre and then effect behaviors

normalized to have a length of 1.0 at the sphere surface. The poses within each bin were then assigned a representative vector within the sphere coordinates that best expressed the nature of the pose relative to other poses in the bin. A vector pointing up (0,1,0) might be mapped to a pose with arms above the head or a character making the thumbs up gesture. Sometimes poses progressed along an axis. For instance a left moving gesture might have a critical pose representation at (1,0,0) followed by another at (2, 0, 0). A design workflow quickly developed between the puppeteers and the artists making poses where they expressed their requests and results as up, down, left, right, forward and back with extended poses as up-up, down-down etc. While there was no real restriction on the use of vectors like (0.5, 1.0, 0.0) , the human element in the design process did not work well with abstract numbers either in design or control. Another nice aspect of this design workflow was that it allowed artists to work from representative pictures provided by cultural content experts involved in the project. Once a puppet was rigged for animation it was an easy matter to create and edit poses based on expert feedback.

Finding the two closest poses within a bin was done by picking the two closest pose vectors to the controller, a process which no longer required individual user calibration. The active two poses were weighted proportional to each other when transitioning from one to the other by projecting the 3d marker to the line between each pose's control point. When the projected controller fell halfway between each pose, the weights were 0.5 and 0.5 respectively (Figure 3c). When the projection fell past either of the endpoints the closest endpoint was weighted at 1.0 and the other at 0.0. When a previously active pose became deselected its current weight was decayed to 0.0 at a fixed rate configured by the puppeteers. In some cases when a pose was spatially organized in between other poses a radius was added to the endpoint which gave a "sticky" zone where the pose was held (Figure 3b).

This new system gave a great deal of control over the appropriate cultural feel of the puppet, gave the puppeteer much better control of both the puppets they would control and also the set of poses they wished to use. The pose blending now was operated by a single hand with the elbow comfortably resting either on the desk or chair armrest, leaving the other hand to control the scenario and facial poses. The head orientation could still be directly controlled by the same IR camera system measuring the control marker. Compared to previous studies where puppet actions would be visibly displayed, and then require costly comprehension by the user [Lee et al., 2002], an Interactor in this system can store possible actions mentally and draw from these possibilities in real time.

3.4 DramaRama

This study is developing and testing an interactive virtual environment that is designed to build the skills in middle school girls needed to resist peer pressure. Figure 3 depicts a prototype of the control pad used by puppeteers in this study. Along the top are the five characters that appear in the set of scenarios. Below each character is a set of behavior types (Figure 4).

This project was the first to use the enhancements to the Geppetto system derived from the Avatar project trials. Anywhere from 1 to 3 bins were created for each of the 5 puppets. The bins for each puppet were arranged left to right according to where each puppet typically appeared on screen relative to the other puppets. Multiple bins for a puppet were stacked vertically top to bottom based with standing pose sets on the top followed by seated variations. Printed images of each puppet titled each column and descriptive annotations for each bin provided additional feedback during initial training.

Head pose from the full body blend was further modulated by the orientation of the head measured using an IR marker pattern mounted on a cap worn by the puppeteer. A Logitech G13 advanced gameboard was used to control facial pose blending of three expression-pair opposites and the mouth open pose. As the key for any pose was held the weight of the gesture would grow until it reached full weight at a rate specified by the puppeteers. When released the pose would lose influence at a weight decaying slower than initially used for rising to allow rich blending. The mouth open pose had a faster rise and decay time to allow for better lip sync. The quality of the key approach over the glove was in part to simplify the setup. In comparison, the glove required calibration and was generally less reliable. A novel lip tracker using 4 IR markers attached around the puppeteers mouth was explored but the puppeteers preferred the simplicity and control afforded by the gameboard. What was important in all approaches was the ability to create weighted blends involving all facial poses simultaneously. Having multiple poses mixing over time allows Geppetto to transcend the plastic look often associated with digital puppets, thereby bringing the character to life. While this was not done, a certain amount of random background noise is suggested from this result. There was an automated blink pose on all puppets that also added a great deal to the overall effect.

The result is a lifelike, almost infinitely varied set of gestures that, together with interactive dialogue, that provides deep immersion on the part of the user.

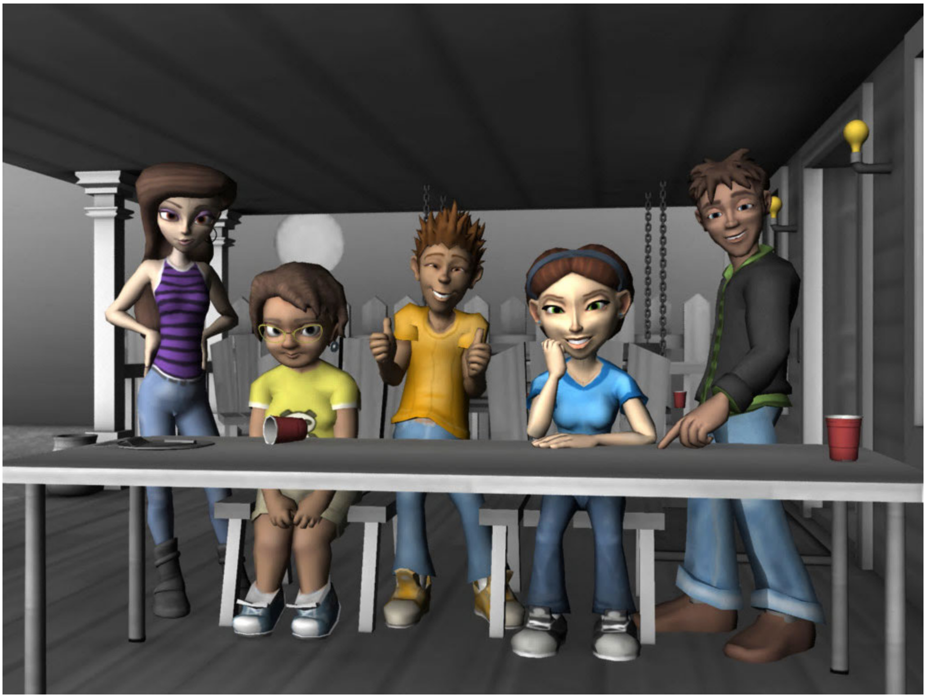


Fig. 4. Characters in DramaRama

4 Conclusions

The primary gains of this approach to puppetry over full-body capture have been the reduction in noise that arises from tracking a large number of points; a reduced hardware footprint; reduction in overall hardware cost and a significant reduction in data that must be transmitted; Additionally, the approach greatly reduces the cognitive and physical load on our puppeteers, especially when they must be concerned about varying cultures to which this must be delivered – non-verbal cultural awareness exists at the receiving end not at the puppetry end.

Acknowledgements. The material presented in publication is based on work supported by the National Institute for Nursing Research (R15NR012189-01) and the National Science Foundation (CNS1051067). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Institute for Nursing Research or the National Science Foundation.

References

1. Sturman, D.J.: Computer Puppetry. Computer Graphics and Applications 18(1), 38–45 (1998)
2. Walters, G.: The Story of Waldo C. Graphic. Course Notes: 3D Character Animation by Computer. In: Proceedings of the 16th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 1989), Boston, pp. 65–79. ACM Press, New York (1988)
3. Trowbridge, S., Stapleton, C.: Melting the Boundaries between Fantasy and Reality. Computer 42(7), 57–62 (2009)
4. Dieker, L., Hynes, M., Hughes, C.E., Smith, E.: Implications of Mixed Reality and Simulation Technologies on Special Education and Teacher Preparation. Focus on Exceptional Children 40, 1–20 (2008)
5. Mazalek, A., Chandrasekharan, S., Nitsche, M., Welsh, T., Thomas, G., Sanka, T., Clifton, P.: In: Proceedings of the 2009 ACM SIGGRAPH Symposium on Video Games, New Orleans, Louisiana, pp. 161–168. ACM Press, New York (2009)
6. Rymaszewski, M.: Second Life: The Official Guide. Wiley-Interscience, New York (2007)
7. Kessler, G., Hodges, L., Walker, N.: Evaluation of the CyberGlove as a Whole-Hand Input Device. ACM Transactions on Computer-Human Interactaction 2(4), 263–283 (1995)
8. Lee, J., Chai, J., Reitsma, P.S.A., Hodgins, J.K., Pollard, N.S.: Interactive Control of Avatars Animated with Human Motion Data. In: Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 2002), San Antonio, pp. 491–500. ACM Press, New York (2002)