Interactive Character Animation with Vision

T. Edmunds   D. Kaufman   P. G. Kry   D. K. Pai   S. Sadhukhan   S. Sueda   D. Wang   Q. Wei

Rutgers, The State University of New Jersey, USA

Abstract

We demonstrate the Scarlet Knight Interactive Persona, a life-sized animated human avatar which interacts with users using stereo vision. Interaction includes mimicry (where the Knight moves around his domain to reflect the user's position in the world), and game play (in which the user's movement is construed as mark placement in a game of tic-tac-toe).

Categories and Subject Descriptors (according to ACM CCS): I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism

1. Introduction

Our system is an implementation of a human-like agent that dynamically responds to the actions of its audience. The avatar is rendered on a large projection display to yield an approximately life-sized character. The behaviour of the Knight is constructed on the fly, using a repertoire of high-quality motion capture data, sequenced by a finite state machine. To emphasize the perception of physical presence in front of the audience, the Knight's shining armour reflects the actual environment, including the audience, as captured by a separate camera-and-mirror system.

In addition to the environment-reflection camera, the system utilizes a Bumblebee binocular stereo vision system (from Point Grey Research Inc.) to track the location of audience members in real time. Our initial experiments at a day-long public event demonstrated the system to be both robust and engaging.

2. Overview

2.1. Audience Observation

The audience-observation that provides the Knight with awareness of his environment is accomplished by stereo depth-reconstruction of the area in front of the display. Audience members are localized in 3D, and their positions categorized according to an imaginary grid laid over the audience area. By discretizing user positions into grid squares (of approximately 1 m^2), minor fluctuations in the detected position are smoothed out without relying excessively on temporal smoothing. This manner of trading spatial resolution against latency keeps the system responsive (<20 ms latency for the vision component).

2.2. Avatar Behaviour

The behaviour of the Knight is controlled by a finite state machine, in which state transitions are influenced by the input from the stereo vision system described above. This structure allows for rapid development of new behaviour modes; to program a new interaction context for the Knight, the designer need only create a new set of state instances whose transition rules correspond to the desired behaviour within the context. The audience state (as detected by the observation system above), along with any auxiliary state information that is required for the context (such as game-board state when playing tic-tac-toe) is recorded in a common world-state object that is accessible to any state instance that becomes the active state. Each state instance has an associated motion sequence that is played out by the avatar whenever that state becomes the active state.

2.3. Motion Sequences

The state transitions underlying the Knight's behaviour are manifested by animating the avatar with pre-recorded motion-capture sequences. Fixed motion sequences are used (as opposed to synthesized motions, as in [AF02], [LCR*02], or [KGP02]) to give a 1-to-1 correspondence with the behaviour states; this guarantees that quantities such as net global translation will be the same every time the state is activated.

To allow different motion sequences to be played consecutively without obvious position or velocity discontinuities of the Knight's limbs, two approaches are taken. In the motion recording stage, the motions are scripted to begin and end in a reproducible rest pose (a measured board on the floor helps achieve consistent relative placement of the feet). To deal with the impossibility of accurately reassuming an identical pose, and with the noise introduced by motion capture and post-processing, the animation procedure incorporates blending when transitioning from one motion sequence to another.

Figure 1: *A demonstration of the Scarlet Knight during an all day public event.*
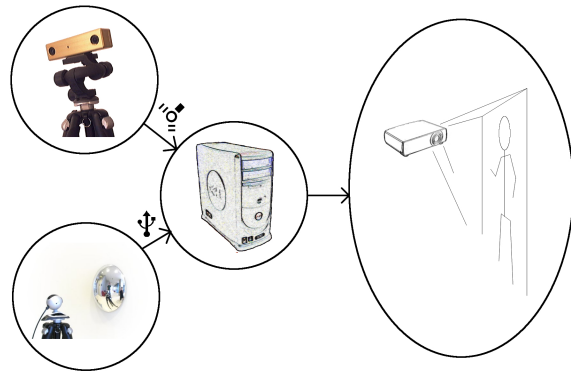

Figure 2: *A schematic overview of the system components. The input from the stereo vision system and from the environment mapper are used to generate the Knight's behaviour and appearance, which is projected onto the display screen.*

The two motion sequences are overlapped, with the skeleton configurations being mixed at each overlapping frame - the terminating sequence fades out as the new sequence fades in.

To play the motion sequences, the skeleton configurations recorded by the motion capture system must be mapped to vertex locations for the Knight model. Since the proportions of our motion capture subject differ from that of Knight, we define a skeleton for the model that has the same structure as the motion capture skeleton, but with different bone lengths. The captured joint angles are mapped onto the Knight's skeleton at load time to avoid redundant computation during the animation phase.

The vertex positions of the flexible portions of the Knight model (the undergarments) are determined by the common Skeletal-Subspace Deformation (SSD) technique [MTT91]; the Knight's armour plates, however, must be transformed rigidly to avoid deformation. To position the rigid portions of the model, an auxiliary set of bones is defined as a function of the skeleton's primary bones. An individual armour plate is attached to an auxiliary bone, which moves in response to the configuration of the primary bones.

2.4. Environment Mapping

In addition to the avatar's response to the audience actions, the impression of physical presence is reinforced by emulating real reflections from the Knight's shining armour. A standard webcam is used to acquire an image of the scene in front of the display; to give the camera a wide field of view approximating a spherical environment map, the camera is aimed at a convex mirror. The images, acquired at ∼15 Hz, are used as the the environment map for texturing the armour plates.

2.5. Physical Architecture

The entire system runs on one dual 2.0 GHz Xeon machine running Windows 2000 with an NVIDIA GeForceFX 5950 graphics card. The stereo vision camera communicates over FireWire, while the webcam is connected via USB. The display consists of Christie LX20 projector front-projecting onto a ∼2 m x ∼2.5 m screen mounted ∼1.4 m above the floor (to keep the projector beam over the audience's head). See Figure 2 for a schematic layout of the system.

3. Observations

We exposed our system to the public (attendees of a Computer Science Department open house) for a day, over the course of which we observed the response of the audience. Users found the knight sufficiently engaging to relish playing tic-tac-toe against him, despite the simplicity of the game. We also noted the tendency of the audience to personify the agent, to the extent of ascribing interactive behaviour that was not actually present (such as full-body mimicry rather than just positional imitation); although this tendency is often present in any human-computer interaction, its prevalence is encouraging for the use of of this type of system for autonomous interactive agents.

References

[AF02] ARIKAN O., FORSYTH D.: Interactive motion generation from examples. In *Proceedings of SIGGRAPH 2002* (July 2002).

[KGP02] KOVAR L., GLEICHER M., PIGHIN F.: Motion graphs. In *Proceedings of SIGGRAPH 2002* (July 2002).

[LCR∗02] LEE J., CHAI J., REITSMA P. S. A., HODGINS J. K., POLLARD N. S.: Interactive control of avatars with human motion data. In *Proceedings of SIGGRAPH 2002* (July 2002).

[MTT91] MAGNENAT-THALMANN N., THALMANN D.: Human body deformations using joint-dependent local operators and finite-element theory. 243–262.