Interactive Control of Avatars Animated with Human Motion Data

Jehee Lee
Carnegie Mellon University

Jinxiang Chai
Carnegie Mellon University

Paul S. A. Reitsma
Brown University

Jessica K. Hodgins
Carnegie Mellon University

Nancy S. Pollard
Brown University

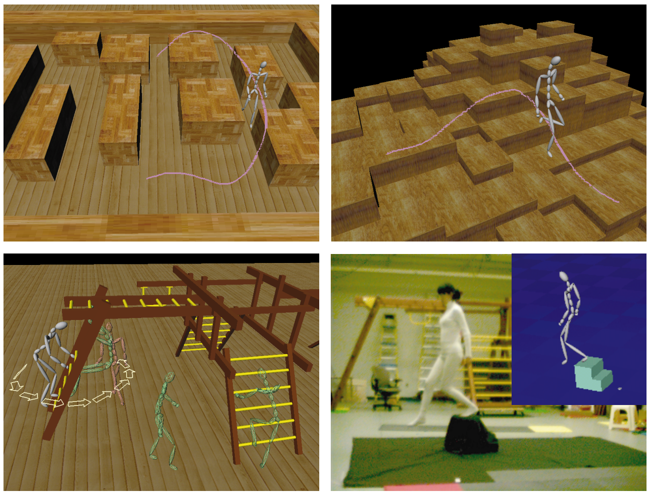

Figure 1: Real-time avatar control in our system. (Top) The user controls the avatar's motion using sketched paths in maze and rough terrain environments. (Bottom left) The user selects from a number of choices in a playground environment. (Bottom right) The user is controlling the avatar by performing a motion in front of a camera. In this case only, the avatar's motion lags the user's input by several seconds.

The popularity of three-dimensional computer games with human characters has demonstrated that the real-time control of avatars is an important problem. Two difficulties arise in animating and controlling avatars, however: designing a rich set of behaviors for the avatar, and giving the user control over those behaviors. Designing a set of behaviors for an avatar is difficult primarily due to the real-time constraint, especially if we wish to make use of relatively unstructured motion data for behavior generation. The raw material for smooth, appealing, and realistic avatar motion can be provided through a large motion database, and this approach is frequently used in video games today. Preparing such a database, however, requires substantial manual processing and careful design so that the character's behavior matches the user's expectations. Such databases currently tend to consist of many short, carefully planned, labeled motion clips. A more flexible and more broadly useful approach would allow extended, unlabeled sequences of motion capture data to be exploited for avatar control. If such unstructured data is used, however, searching for an appropriate motion in an on-line fashion becomes a significant challenge.

Providing the user with an intuitive interface to control the avatar's motion is difficult because the character's motion is high dimensional and most of the available input devices are not. Input from devices such as mice and joysticks typically indicates a position (go to this location), velocity (travel in this direction at this speed) or behavior (perform this kick or pick up this object). This input must then be supplemented with autonomous behaviors and transitions to compute the full motion of the avatar. Control of individual degrees of freedom is not possible for interactive environments unless the user can use his or her own body to act out or pantomime the motion.

In this research, we show that a rich, connected set of avatar behaviors can be created from extended, freeform sequences of motion, automatically organized for efficient search, and exploited for real-time avatar control using a variety of interface techniques. The motion is preprocessed to add variety and flexibility by creating connecting transitions where good matches in poses, velocities, and contact state of the character exist. The motion is then clustered into groups for efficient searching and for presentation in the interfaces. A unique aspect of our approach is that the original motion data and the generalization of that data are closely linked; *each frame* of the original motion data is associated with a tree of clusters that captures the set of actions that can be performed by the avatar from that specific frame. The resulting *cluster forest* allows us to take advantage of the power of clusters to generalize the motion data without losing the actual connectivity and detail that can be derived from that data. This two-layer data structure can be efficiently searched at run time to find appropriate paths to behaviors and locations specified by the user.

We explore three different interfaces to provide the user with intuitive control of the avatar's motion: choice, sketch, and performance (figure 1). In choice interfaces, the user selects among a number of options (directions, locations, or behaviors) every few seconds. The options that are presented to the user are selected from among the clusters created during the preprocessing of the motion data. In the sketching interface, the user specifies a path through the environment by sketching on the terrain, and the data structure is searched to find motion sequences that follow that path. In performance interfaces, the user acts out a behavior in front of a video camera. The best fit for his or her motion is then used for the avatar, perhaps with an intervening segment of motion to provide a smooth transition. For all three interface techniques, our motion data structure makes it possible to transform possibly low-dimensional user input into realistic motion of the avatar.

We demonstrate the power of this approach through examples in four environments (figure 1) and through comparison with directly recorded human motion in similar environments. We note that the vision-based interface, due to the higher dimensional nature of the input, gives the most control over the details of the avatar's motion, but that the choice and sketch interfaces provide the user with simple techniques for directing the avatar to achieve specific goals.