A Visual Computing Environment for Very Large Scale Biomolecular Modeling

1 Speech and Gesture Interface

To fully exploit the potential that visual computing environment offers, there is a need for "natural" interface that allow the manipulation of such displays without cumbersome attachments. In this section we describe the use of visual hand gesture analysis and speech recognition for developing a speech/gesture interface to VMD. The free hand gestures are used for manipulating the 3-D graphical display together with a set of speech commands. We describe the visual gesture analysis and the speech analysis techniques used in developing this interface. The dual modality of speech/gesture is found to greatly aid the interaction capability.

The communication mode that seems most relevant to the manipulation of physical objects is hand motion, also called *hand gestures*. We use it to act on the world, to grasp and explore objects, and to express our ideas. Now virtual objects, unlike physical objects, are under computer control. To manipulate them naturally, humans would prefer to employ hand gestures as well as speech. Psychological experiments, for example, indicate that people prefer to use speech in combination with gestures in a virtual environment, since it allows the user to interact without special training or special apparatus and allows the user to concentrate more on the virtual objects and the tasks at hand [**?**]. We explore this multimodal nature of HCI involved in manipulating virtual objects using speech and gesture.

To keep the interaction natural, it is desirable to have as few devices attached to the user as possible. Motivated by this, we have been developing techniques that will enable spoken words and simple free-hand gestures to be used while interacting with 3D graphical objects in a virtual environment. The voice commands are monitored through a microphone and recognized using automatic speech recognition (ASR) techniques. The hand gestures are detected through a pair of strategically positioned cameras and interpreted using a set of computer vision techniques that we term automatic gesture recognition (AGR). These computer vision algorithms are able to extract the user hand from the background, extract positions of the fingers, and distinguish a meaningful gesture from unintentional hand movements using the context. We use the context of the VMD environment to place the necessary constraints to make the analysis robust and to develop a command language that attempts to optimally combine speech and gesture inputs.

The older version of the VMD system uses a keyboard and a magnetically tracked pointer as the interface. This is particularly inconvenient since the system is typically used by multiple (6-8) users, and the interface hinders the interactive nature of the visualization system. Thus incorporating voice command control in MDScope would enable the users to be free of keyboards and to interact with the environment in a natural manner. The hand gestures would permit the users to easily manipulate the displayed model and "play" with different spatial combinations of the molecular structures. The integration of speech and hand gestures as a multi-modal interaction mechanism would be more powerful than using either mode alone, motivating the development of the speech/gesture interface. Further, the goal was to minimize the modifications needed to the existing VMD program for incorporating the new interface. The experimental prototypes that we built for both the speech (ASR) and hand gesture analysis (AGR) required the following addition to the VMD environment.

Software. In order to reduce the complexity and increase the flexibility of the program design, a communications layer was added so external programs can be written and maintained independently from the VMD code. These use the VMD text language to query VMD for information or to send new commands. The VMD text

This work was supported in part by the U. S. Army Research Laboratory under Cooperative Agreement No. DAAL01-96-2-0003 and in part by the National Science Foundation Grant IRI-96-34618.