Virtual Viewpoint Reality
9807-28

Progress Report: July 1, 1998—December 31, 1998

Paul Viola and Eric Grimson

Project Overview

In the foreseeable future, sporting events will be recorded in super high fidelity from hundreds or even thousands of cameras. Currently the nature of television broadcasting demands that only a single viewpoint be shown, at any particular time. This viewpoint is necessarily a compromise and is typically designed to displease the fewest number of viewers. In this project we are creating a new viewing paradigm which will take advantage of recent and emerging methods in computer vision, virtual reality and computer graphics technology, together with the computational capabilities likely to be available on next generation machines and networks. This new paradigm will allow each viewer the ability to view the field from any arbitrary viewpoint—from the point of view of the ball headed toward the soccer goal; or from that of the goalie defending the goal; as the quarterback dropping back to pass; or as a hitter waiting for a pitch. In this way, the viewer can observe exactly those portions of the game which most interest him, and from the viewpoint that most interests him (e.g. some fans may want to have the best view of Michael Jordan as he sails toward the basket; others may want to see the world from his point of view).

Progress

Early in January we hosted a visit from several members of NTT: Dr. Ishii, Dr. Kurakake, Mr. Yamato, and Mr. Kuano. During that visit we discussed both the Virtual Viewpoint Reality (VVR) project as well as the Image Database project. We received much useful feedback during this visit, including an overview of related projects at NTT. The minutes of this visit were compiled and sent to us by Mr. Yamato. We were very much in agreement with the suggestions described in these minutes. In order to further support collaboration we have setup a web page containing our presentations, demonstrations, and other related information http://www.ai.mit.edu/projects/NTTCollaboration

During our discussions it became clear that both NTT and MIT are very interested in basic research questions. Some of the basic questions that underlie VVR include: 3D model construction, real-time tracking, human movement analysis, and the fusion of information from cameras and other sensors. While we are very committed to basic research goals, we have also begun to narrow down our practical goals.

Multi-camera video acquisition and 3D modeling: We intend to construct a real-time (or near real-time) multi-camera data acquisition system. Each camera will be connected to a dedicated computer. The video output of each camera will be processed in parallel to extract the locations of players and other features. The information will be integrated on a single multi-processor machine. This machine will compute a 3D representation of the scene and send this to yet another machine which will include 3D visualization hardware.

Movement Analysis: In parallel to the virtual viewpoint computations we will track the movements of players both individually and in concert. This will allow us to analyze typical plays, and potentially provide commentary.

Ancillary Processing: We will also explore a number of ancillary information sources in an attempt to integrate them with VVR environment. For example, we will explore the use of sound acquired from a large array of microphones. Hopefully we will be able to "listen in" on players and conversations with referees. We will also explore the use of other types sensors such a large scale tactile array developed by NTT. When placed on the floor, the tactile array can be used to localize players as they participate in the VVR environment. NTT has also developed an eye tracker which can give us information about the viewer. Perhaps the VVR user interface can be extended to incorporate this information (e.g. the user may be able to direct the system to follow a particular player simply by staring at him.)

Demonstration Domain: In discussion with Dr. Ishii we have settled on soccer as a good domain to demonstrate our algorithms. At first we will build a system that will work indoors which will accommodate one or perhaps a pair of people. We will also build a rough scale model of sports field in order to explore that integration issues at a larger scale. It is our hope that in later years of the project that we can build larger systems.

Research Milestones

In addition to defining and refining our goals, we have achieved a number of research milestones:

New Lab Space: We set up a lab space for use in VVR. The new lab includes a 14 computers equipped with cameras and frame-grabbers. We have setup a high speed network and parallel processing software (MPI) so that the inherently parallel nature of VVR computation is easily supported. The lab space is split into two parts: an enclosed space in which lighting can be controlled, and an open space through which people can pass.

Constructed a 3D scanner: We have built a computer controlled "scanner" that can move a camera in a controlled fashion around a small static object. The scanner can be

used to acquire up to 256 images of a single object. This data has allowed us to debug and evaluate our reconstruction algorithms.

Development of a Tomographic Reconstruction Algorithm: We have developed an entirely novel scheme for 3D reconstruction that works on the principles of tomography. The technique uses the reconstruction technique known as "filtered back projection" to compute a 3D model from many images. A full description of the approach is being prepared now (and will be included on our web site).

Development of a Probabilistic Reconstruction Algorithm: The tomographic reconstruction algorithm, while potentially very efficient, is limited to the reconstruction of mostly convex objects. We have constructed a new probabilistic algorithm which can be used to compute the shape of non-convex objects. A full description of the approach is being prepared now (and will be included on our web site).

Goals

Our goals for the next six months include: i) setting up a near real-time coarse 3D reconstruction system; ii) further developing our 3D reconstruction algorithms; iii) developing initial algorithms for rendering our 3D representations; iv) developing better tracking algorithms for humans in 3D; and v) exploring the fusion of sound and video information.