Multiple Perspective Interactive Video

Arun Katkere Don Kuramura Saied Moezzi Patrick Kelly Deborah Swanberg
Koji Wakimoto Edward Hunter Li-Cheng Tai Shankar Chatterjee Ramesh Jain
email: katkere@vision.ucsd.edu
Department of Electrical and Computer Engineering
9500 Gilman Drive, Mail Code 0407
University of California, San Diego
La Jolla, CA 92093-0407.

Abstract

Crucial to the transformation of computers into intelligent multimedia systems is the ability to access, manipulate and manage visual information. In particular, video data is becoming a prevalent source of such information. Our *Multiple Perspective Interactive Video (MPI-Video),* project integrates a variety of visual computing operations with modeling and interaction techniques, to extract, synthesize and manage dynamic representations of scenes observed from multiple perspectives. Automatic and semi-automatic analysis of video data from multiple cameras is performed. This analysis is used to build a three-dimensional model of the environment monitored by the cameras. *MPI- Video* has applications in a variety of areas including the development of *immersive video* for telepresence systems, traffic monitoring and control and the analysis of physical performances, e.g. sports and dance. This video presents an overview of *MPI- Video* along with demonstrations of two *MPI-Video* prototypes developed in the *Visual Computing Lab.* Further details on *MPI- Video* are described by Jain and Wakimoto in [Jain and Wakimoto, 1995] and in our technical report [Kelly *et* a/., 1995]

First, we present an *MPI-Video* system which automatically identifies and tracks objects moving in an environment recorded by four cameras. Each camera "sees" the environment from a different perspective. Moving object data is extracted from each camera image using image differencing techniques. This information is integrated by a centralized modeling and assimilation module to construct a three-dimensional model of object behavior and events in the environment. An interface is provided allowing users to interact with the three-dimensional model and the video data. The interface presents both the raw video data, annotated to indicate moving objects in the scene, along with the three dimensional model. Users can view the environment from a variety of perspectives which are not limited to the real camera views.

Using a cursor, viewers can select objects and areas in the scene. A marker indicating the selection as seen from each camera perspective, is placed in each camera window. This demonstration also introduces the notion of a "best view." That is, a particular camera view for which some criteria, either specified by the system or a user, is optimal. For instance, the "best view" might be the vista in which a particular object of interest is largest. Here, the "best view" is that for which the distance along the line of sight from camera to selected position is minimum. When a location in the model is selected the system determines the line of sight distance and chooses that view, from the four available, for which this distance is shortest.

To perform the video analysis and modeling, our *MPI-Video* system uses information about the "static" world. Camera calibration maps related locations in the two-dimensional video to a fully three-dimensional representation of the world recorded by the cameras. The video demonstrates software developed to assist in this calibration. Details on this *MPI-Video* prototype can be found in our technical report [Kelly *et al.,* 1995].

The use of *MPI-Video* to support an *Immersive Video* application is also shown. In *Immersive Video* visual information from multiple live video images of a real-world event is obtained and integrated to provide a photo-realistic rendition of the dynamic environment, thus supporting a sense of total immersion in the environment. Using the actual camera data, a view from a location anywhere in the environment can be constructed. These views are created by mosaicing pixels from the video data. Several such "virtual" views are shown and an animated walk through[1] of the courtyard environment illustrates a sequence of such views. *Immersive Video* is described in greater detail in our technical report [Moezzi *et ai,* 1995].

The second portion of the video highlights an early *MPI- Video* prototype, named *Fumble.* In this prototype, we analyze football data from three cameras covering a Superbowl. The system tracks different players on the field. While our *MPI-Video* system described above, provides fully automatic detection and tracking, this second system relies on manual analysis of the video data. Here, objects in the video data were identified and this information stored in a database. This knowledge is used to track players in the video sequences and to support user queries of the data. An interface allows users to query about players in the system and a three-dimensional cursor is used to select and locate players in the video data. For instance, a user can place the cursor on a particular player in the video frame and ask "Who is this?" The system will then identify the player. Alternately, a specific player can be chosen from a list and the system will indicate, using the three-dimensional cursor, where the player is in the frame. As in the previous system, *Fumble* also provides a "best view". In this case, the frame which keeps a selected player of interest most central to the frame. A "best view" sequence illustrates this capability of the *Fumble* prototype. Jain and Wakimoto [Jain and Wakimoto, 1995] present more information on the *Fumble* prototype.

References

[Jain and Wakimoto, 1995] Ramesh Jain and Koji Wakimoto. Multiple perspective interactive video. In *IEEE Multimedia Conference,* Washington, May 1995. IEEE, to appear.

[Kelly *et al,* 1995] Patrick H. Kelly, Arun Katkere, Don Y. Kuramura, Saied Moezzi, Shankar Chatterjee, and Ramesh Jain. An architecture for multiple perspective interactive video. Technical Report VCL-95-103, Visual Computing Laboratory, UCSD, 1995. A version submitted to ACM Multimedia 1995.

[Moezzi *et ai,* 1995] Saied Moezzi, Arun Katkere, Shankar Chatterjee, and Ramesh Jain. Visual reality: Rendition of live events from multi-perspective videos. Technical Report VCL-95-102, Visual Computing Laboratory, UCSD, 1995.

[1]An MPEG-1 version is located at http://vision.ucsd.edu:80/papers.