Variable Viewpoint Reality
9807-28

Progress Report: January 1, 2001 – June 30, 2001

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 that 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 Through June 2001

To create this new viewing paradigm, there are a number of important computer vision and graphics problems that must be solved. These include issues of real-time 3D reconstruction, coordination of large numbers of cameras, rendering of arbitrary viewpoints, learning to recognize common activities, finding similar visual events in archival video, and many other associated problems. We have made progress on many of the problems related to the goals of the Variable Viewpoint Reality project:

- We have developed a number of basic algorithms for 3D reconstruction. One approach is designed to work in real time on many cameras. Another is a bit slower, but is designed to yield higher quality results. A third attempts to find the arm, leg and body positions of a human being from one or multiple camera views.

- We have designed and set up a multiple camera system for acquiring data in real-time. The first system was designed to be flexible and to work indoors. The system uses 16 cameras working in synchrony, and has been used to acquire a great deal of multi-camera data.

- As an alternative to direct 3D reconstruction, we have investigated the use of visual hull methods, in collaboration with Prof. Leonard McMillan. We have used a 4 camera version of the visual hull approach to capture real time renderings of arbitrary views of a person and his/her activities.

- In collaboration with students working on another project we have been observing outdoor activities. This system provides coarse tracking information of multiple people and cars. The system can also recognize simple activities.

- We have demonstrated the system performing real-time 3D reconstruction using 16 cameras. This system combines many of the results mentioned above.

- We have developed several new algorithms for the reconstruction of 3D shapes. In previous reconstruction algorithms, each camera attempts to segment the object from the background. These segments are then intersected to form a 3D shape. In some cases noise in the cameras leads to incorrect segmentation. This in turn leads to poor reconstruction. Our new algorithm explicitly models this noise and introduces a prior over shapes. The result is the Bayes optimal reconstruction that is very insensitive to noise. These results are described in a paper at CVPR 2000, which is available from the MIT/NTT web page: http://www.ai.mit.edu/projects/ntt.

- We have developed new algorithms for the automatic calibration of the camera array. Typically, the calibration of 16 cameras is a very difficult and time-consuming task. Our approach requires little human intervention and can be used to dynamically update the calibration over time. The algorithm proceeds by a random search over camera poses in an effort to maximize the reconstruction volume.

- Foreground-Background segmentation is a key problem for all of our reconstruction algorithms. To be successful, the methods must separate out image regions that correspond to objects of interest in each camera, so that the information can be combined to create a single unified representation of the object. We have developed a number of new algorithms for segmentation that use prior information in order to improve results. Both of our new results in this area are based on the concepts of Graph-cuts, a technique which can rapidly find the optimal segmentation give a prior model for shapes.

- One difficult problem that can be solved is to reduce the effects of shadow in the segmentation process. Our basic image segmentation approach is to use image differences between the current image and a “background” image in order to highlight regions where a person appears. There are two problems that graph cuts solve: 1) in some cases the person is wearing clothes which are the same color as the background; 2) in some cases the person casts a shadow that appears different from the background. Using graph-cuts and a 3 label MRF the foreground, background and shadow regions can be identified.

- Capturing arbitrary renderings of a person is the central step for some applications of this approach. For other applications, however, it is equally important to recognize the person or persons involved in the activity and to categorize that actual activity. In conjunction with students working on a related project, we have been applying machine learning and classification methods to these problems. Using gait information extracted from people walking through the visual hull, we are able to recognize the individual with high reliability, even when the video is taken over a period of several weeks. We are also able to combine this gait recognition data with face recognition methods to further improve performance.

- We are currently examining the use of these classification methods to categorize particular actions of individuals within the Visual Hull space. Examples include a person dribbling a basketball, throwing a baseball, running and skipping through the scene, and so on. An example of different activities is shown below. Our goals are to categorize these activities into generic classes, and to analyze the actual motions at a fine detail in order to find similar actions in a stored database and to measure variances in these actions over time.

Research Plan for the Next Six Months

- Our primary goal is to extend the use of the Visual Hull method for real time acquisition of actions. This includes developing and testing methods for automatically selecting the optimal viewpoint for specific tasks: for example, if we want to recognize a person, automatically determining the best viewpoint for acquiring gait information and for acquiring face information; if we want to analyze a particular motion, automatically determining the best viewpoint for isolating that motion from other actions. It will also include developing and testing methods for extracting actions of multiple people in a setting, and separating those actions into distinct individual ones. And it will include developing and testing methods for categorizing those actions: against known, learned classes of actions; and against previous versions of those actions to determine changes with respect to previous examples.