Variable Viewpoint Reality
9807-28

Progress Report: July 1, 2000 – December 31, 2000

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).

Summary of Progress Through July 2000

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 rapid 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. Each of these algorithms continues to be tested and refined.

- 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. At present, we have 16 cameras working in synchrony. We are investigating alternative systems as well.

- We have acquired a great deal of multi-camera data. This is allowing us to test our algorithms and to develop new ideas.

- 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.

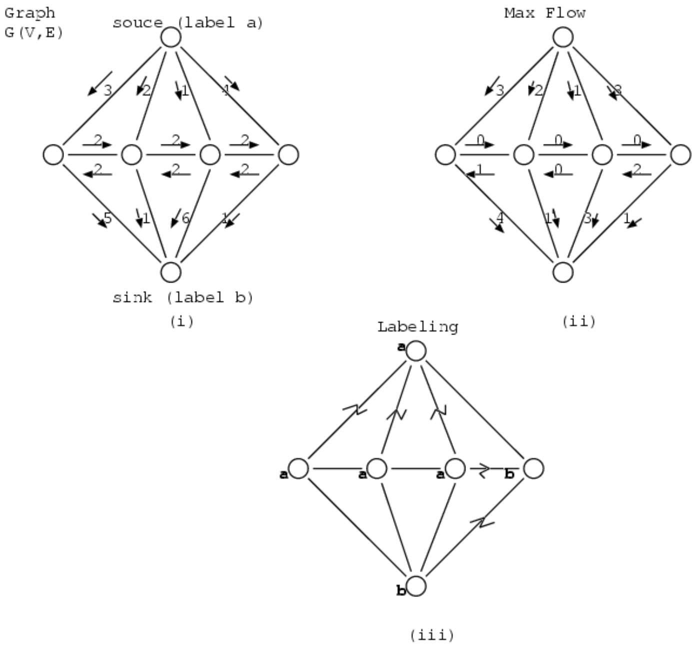
- We have demonstrated the system performing real-time 3D reconstruction using 16 cameras. This system combines many of the results mentioned above.

- We have developed a new algorithm for the reconstruction of 3D shapes. This algorithm addresses one of the key problems we have encountered to date -- noise in the camera observations. 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.

Progress Through December 2000

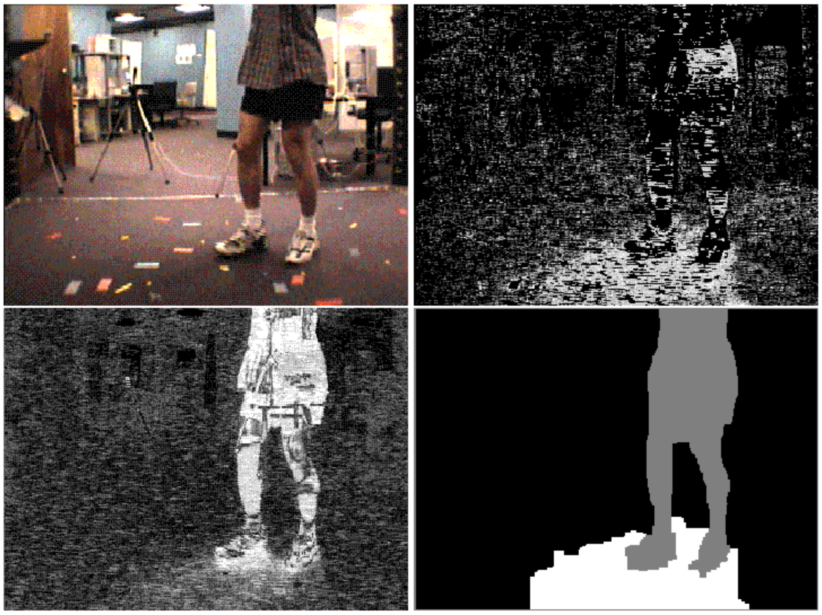
- Foreground-Background segmentation is a key problem for all of our reconstruction algorithms. To be successful, the methods must separate out image regions that correspond to objects of interest in each camera, so that the information can be combined to create a single unified representation of the 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.



Graph-cuts can be used to find the segmentations of high likelihood under an MRF prior for shape. Classically MRF's were solved using a time intensive technique called Gibbs

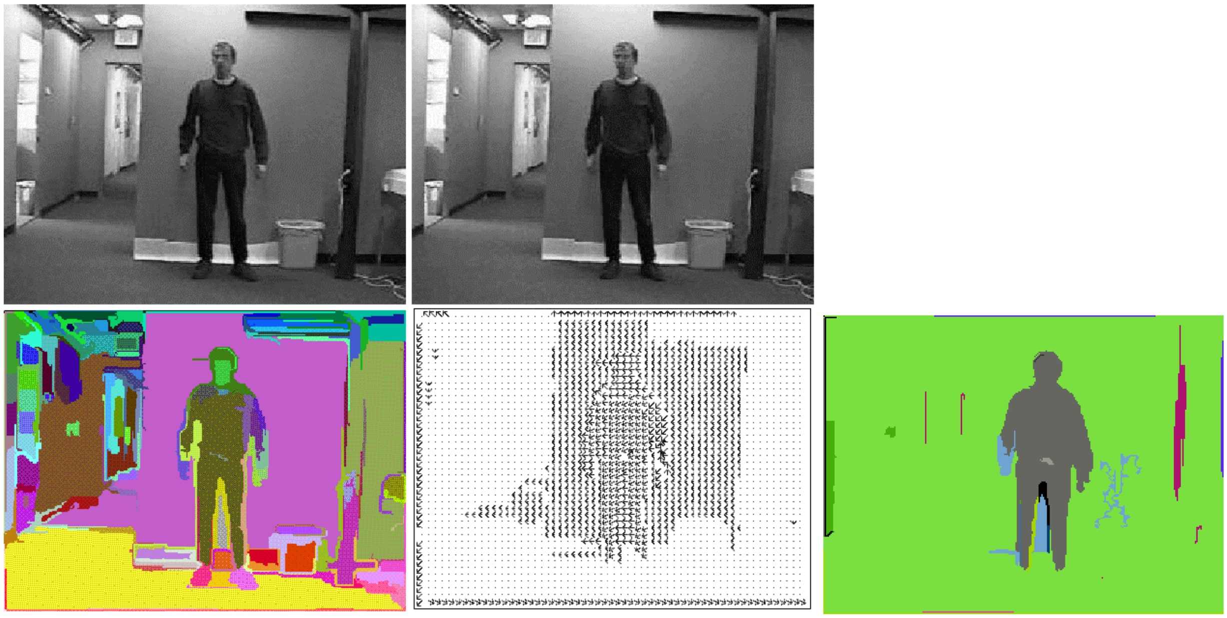
Sampling. This new approach transforms the MRF into a conventional graph and uses polynomial time graph cutting algorithms to find the lowest cost solution.

- One difficult problem that can be solved is to reduce the effects of shadow in the segmentation process (see Figure Below). 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.



Upper left image shows the input (background omitted). Upper right and lower left show evidence used for segmentation. Lower left is image differences (notice the holes within the body). Upper right show shadow evidence. Locations where the color remains the same but intensity is reduced. Lower right shows the final segmentation.

- In many other cases there may not be a background image for use in segmentation (e.g. in the analysis of existing video). Conventional approaches use image segmentation approaches that attempt to group pixels based on similar colors. The problem is that the person may where clothes of various colors. These would appear as difference regions. Another source of information is motion. Often the person is the only moving object. We have built a system that fuses these two sources of information in order to get much better results.



Upper images show two images from a video sequence. Person is moving to the left. Lower left shows image segmentation based on color information. Lower middle shows motion information. Note this motion field is extracted using the boundaries estimated using color. Lower right show the segmentation using motion and color information.