Architectures for Distributed Smart Cameras

Wayne Wolf, Burak Ozer, and Tiehan Lv
Department of Electrical Engineering
Princeton University
{wolf,iozer,lv}@ee.princeton.edu

Abstract
This paper describes our new multiple-camera architecture for real-time video analysis. This architecture uses an array of relatively inexpensive cameras to gather images in order to provide high resolution at low cost. The system also uses a hierarchy of cameras, including both wide-angle and telephoto views. Wide-angle cameras are responsible for camera coordination while telephoto cameras are primarily responsible for detailed processing of parts of the scene.

1 Introduction
Smart cameras perform real-time video processing in situ. A smart camera may perform a variety of tasks: face recognition, tracking, etc. A smart camera both captures the images and processes them in a single package. High levels of integration make smart cameras both less expensive and physically smaller.

This paper describes a new architecture for distributed real-time video processing. We consider aspects of both the imaging architecture—how the cameras cover the scene—as well as the distributed computing architecture used for real-time video analysis. Our architecture takes advantage of advances in VLSI technology in two ways. First, we assume that cameras are cheap and that we can use many cameras in the system; in the past, cameras and their associated processors have been relatively expensive. Second, we assume that powerful processors can be distributed throughout the system in order to handle computational tasks close to the location at which the video is gathered.

The distributed smart camera architecture proposed here has two novel features. First, it uses an array of relatively inexpensive cameras to gather images. Viewing a subject with several lower resolution cameras provides better results than using fewer high-resolution cameras. Second, the array uses a hierarchy of cameras, including both wide-angle and telephoto views. Wide-angle cameras are responsible for camera coordination while telephoto cameras are primarily responsible for detailed processing of parts of the scene. This architecture helps reduce processing time and improves real-time responsiveness.

The next section summarizes related work. Section 3 shows how our previous work in real-time video analysis on embedded processors evolved toward our approach to distributed smart cameras. Section 4 discusses trade-offs between the number of cameras and the resolution of each camera. Section 5 discusses our hierarchical camera organization. Section 6 describes the challenges in real-time and low-energy operation of a distributed camera array.

2 Previous Work
Several groups have developed algorithms and architectures for single-node real-time video systems. Pentland’s survey describes several real-time video analysis efforts. Chai et al and Watlington and Bove have developed real-time architectures for video analysis.

Rourke and Badler conducted early work on human motion analysis. Hogg studied periodic walking activity. Rehg and Kanade developed a 3D hand model based on truncated cones. Kakadiaris and Metaxas developed a method for fitting an articulated 3-D model to an object. Davis et al at the University of Maryland have developed a multi-camera video analysis system.

Collins, Lipton, and Kanade developed a multi-camera video surveillance system. Matsuyama and Ukita developed a multi-camera real-time vision system in which the cameras were mechanically panned and zoomed.

3 Distributed Smart Cameras
Our work in distributed smart cameras is based upon our previous work in real-time video analysis using embedded processors. Figure 1 shows some of the steps in our real-time human gesture recognition system. The system starts with background elimination and skin tone detection. The system then extracts contours that bound the skin-tone and non-skin-tone regions. The contours are fit to ellipses...
that model various regions. A graph is then built to describe relationships between the ellipses. That graph is matched to a library of standard gestures stored as graphs. Not shown is the classification of these gestures over time using hidden Markov models (HMMs). This system runs on a Trimedia TM-1300 VLIW video signal processor running at 100 MHz. The system can classify gestures at 25 frames/sec, performing all steps from low-level video operations through HMM classification on every frame.

Using multiple cameras for real-time vision has many advantages. The most significant advantage is improved occlusion handling. A single camera will not be able to see all the elements of a scene. A person may, for example, self-occlude parts of his/her body, such as when a person turned to one side self-occludes the arm on the other side of the body. Objects in the scene may occlude other objects as well. Single-camera systems must take steps to infer the existence of temporarily occluded objects. Multiple-camera systems, in contrast, can see different objects from different cameras and synthesize an un-occluded view of the scene.

But just as real-time image capture implies highly-integrated architectures for single-node smart cameras, distributed multiple cameras requires inherently distributed processing. Sending raw video data to a central node for processing is simply infeasible in most real-world environments. Transmitting large volumes of video data over a network inherently imposes delays impede real-time processing. These high-speed networks are not only expensive, they also burn excessive power. Centralized processing also requires large memory buffers that add to the cost and power consumption of both the cameras and the central processing node.

As a result, we are developing a distributed smart camera system that analyzes video in real-time by dividing the work across camera nodes. Individual cameras perform low-level video operations and provide only higher-level descriptors of their imagery to other nodes. Nodes in the system cooperate in order to create a world view of the scene from the multiple views of the scene.

The distributed architecture must encompass both the image capture configuration as well as the processing architecture. We argue in this paper that a system composed of a mixture of wide-angle and telephoto cameras provides both higher resolution and more responsive real-time operation.

Figure 2 shows some sample frames from the initial 3-D gesture recognition system that we derived from our 2-D system. A 3-D model of the person is synthesized from two cameras placed approximately 90 degrees apart. As with the single-camera system, the analysis is performed in real
time. Each camera analyzes each frame through 2-D ellipse fitting; one of the cameras, working as the master, then fits 3-D ellipses based on the 2-D models. This early implementation uses shared memory communication rather than a network, allowing faster node-to-node communication. This system can recognize gestures at a rate of 10-15 frames/sec.

4 Resolution vs. Proliferation

One question that must be decided about the imaging architecture of a distributed camera system is the typical number of cameras that will be able to see each point. We argue in this section that the system should have multiple lower-resolution sensors looking at each point in the scene, rather than using one large sensor to view each point in the scene.

The cost of the image sensors in the system is roughly proportional to the total number of pixels in all the sensors. For a given pixel budget, we can choose to put those pixels in any number of cameras, ranging from a single camera to one pixel per camera. While a sensor with more pixels can resolve more from that angle, we can also estimate the position of an object from the views of several lower resolution cameras. Using several cameras to observe each point in the scene can provide the high resolution as well as provide some fault tolerance—in case one camera fails, the points in its field of view will be covered by other cameras.

5 Hierarchical Image Understanding

The many cameras in the array must be coordinated in order to work effectively. We propose a hierarchical imaging architecture in order to coordinate the cameras in real time.

Figure 3 illustrates the organization of a section of our hierarchical camera array. The hierarchy consists of telephoto and wide-angle cameras. Telephoto cameras are positioned for detailed views of the scene. As described in the last section, each point of interest in the scene is covered by more than one telephoto cameras. Wide-angle cameras are used to track movement between telephoto cameras and to help patch images at the boundaries between telephoto cameras. Of course, 3-D analysis requires cameras at different angles to the subject, so a realistic setup would have telephoto and wide angle cameras at different orientations; for example, sets of cameras may be placed along perpendicular walls in a room.

Hierarchical camera arrays help reduce processing delays and provide improved real-time performance. Wide angle cameras perform tracking algorithms to determine when a subject will move into the view of a particular camera. Early warning allows the camera to switch into the proper mode to process the subject, reducing real-time latency. It also gives the camera information about the position of the subject that can be used to generate an initial estimate for analysis algorithms.

6 Real Time and Low Energy

Real-time operation and low-energy operation are generally compatible goals in distributed camera systems. Real-time
operation allows both higher frame rates and lower latency. Producing analysis results in low latency is generally important in closed-loop systems in which the results of video analysis are used to help decide whether to take action.

Low-energy operation is important whenever cameras are deployed in everyday environments because power consumption is directly related to installation costs. If the cameras consume more power, they require higher-cost power distribution networks. High-energy cameras also generate more heat, which requires more expensive, bulky and potentially noisy heat dissipation mechanisms. While we do not expect smart cameras to be powered by batteries in the foreseeable future, low-energy operation is still very important for realistic distributed smart camera systems.

The hierarchical camera organization should improve both real-time performance and energy consumption. Information from the wide angle cameras can be used to manage the workload on the computational nodes in the system. The wide angle cameras help to predict the workloads on the telephoto cameras. Latency can be reduced using simple predictions and computational loads can be better managed. In addition, simple predictions allow us to avoid unnecessary computation and save energy.

7 Summary

Advances in VLSI technology are now making it possible to perform real-time video processing in embedded systems. Distributed arrays of smart cameras allow us to capture a unified world view and process the video information using distributed computation nodes. We are developing a hierarchical smart camera system to take advantages of the characteristics of relatively inexpensive smart cameras. A hierarchical imaging architecture allows us to synthesize a high-resolution view of the scene using inexpensive cameras. Distributed processing allows us to analyze the scene with low latency and at high frame rates.

Acknowledgments

This work was funded by the New Jersey Center for Pervasive Information Systems.

References