ABSTRACT
We describe a system of small wireless autonomous mobile cameras that operate as a group to solve machine vision tasks. Given sufficient on-board computing to do real-time scene analysis without external processing resources, these cameras enable us to investigate how scene understanding can be improved when each camera is independently capable of analyzing its own sensor data, and sharing information on what it sees with its fellow robots.

INTRODUCTION
The availability of inexpensive, low-power embedded processors that match the processing capacity of the workstations that many of us used for machine vision in the recent past suggests that we combine these chips with the increasingly widespread wireless networking technologies to create the hardware substrate for a range of smart wireless devices, in particular, intelligent mobile cameras. An intelligent mobile camera is often thought of as a robot, but interest is growing in applications where such devices can perform visual problem-solving as a cooperating group.

Applications for coordinating multiple cameras are faced with a number of logistical and technological challenges. Logistically the cameras must be correctly positioned at known locations, and this calibration information associated with the data from the camera. Processing of the visual data is usually done in a centralized location, in which case limitations on communications bandwidth and processor speed can restrict system scalability, while the existence of a single point of failure (the processing system) limits robustness.

Despite these challenges, the rapidly dropping cost and size of cameras and the increased ease of acquiring images from many of them simultaneously have enabled a growing number of multi-camera applications. The film industry has recently made extensive use of multiple-camera-array shots, typically for specialized effects as exemplified by the film *The Matrix*. Similarly machine vision researchers have shown increased interest in their application to a number of practical scene and object analysis and virtual reality problems.

BACKGROUND
The problem of multi-camera synthesis has received increasing attention from several different fields of research, and in several different forms. Virtual reality researchers such as Moezzi [1] and Kanade [2] have examined the field from the perspective of integrating real and virtual environments, with statically placed sets of cameras for three dimensional data capture. Hilton [3] discusses the use of multiview color images to tackle the problem of building animated, realistic, 3D models of people. Mittal [4] presents a system that uses 16 CCD cameras to track people in a cluttered scene, and has
also examined the related problem of
tracking people through the views from
widely dispersed cameras.\cite{5} An overview
of the requirements for a multi-agent
approach specifically for surveillance
camera control is described by Orwell.\cite{6}

The use of multiple cameras along with
other sensors is also being explored for
Human Computer Interaction, in particular
in the creation of “intelligent rooms.” \cite{7}
Hoover \cite{8} demonstrates the use of a
network of fixed cameras to track and steer a
mobile robot.

The related problem of coordinating
multiple independent robots to perform tasks
collaboratively is also receiving more
attention as the computational power
available increases. There is still a
noticeable reluctance to move to complete
local autonomy; most systems such as that
described by Hajjawi \cite{9} prefer an approach
that includes partial local autonomy but still
involves some kind of remote supervisory
control system.

There has been some exploration of
specifically camera-based problems using
robots; Jennings \cite{10} in particular presents
an approach to robot navigation where two
stereo camera equipped robots co-
operatively build a mutual map of their
landscape.

**EYE SOCIETY**

In this project we are creating a “society” of
small, cheap, autonomous wireless mobile
cameras each controlled by an embedded
processor board running Linux. Each camera
can independently pan and tilt and move
along an overhead lighting track, and is in
constant communication with its fellow
cameras as it does so. With this as a basis
we are investigating how scene analysis can
be improved when each camera is
independently capable of analyzing its own
data, and sharing information on what it sees
with its fellow robots.

![Figure 1: Eye Society robots on an overhead track.](image)

Networks of independent sensors with
limited processing (e.g. \cite{11}) are becoming
common, but the long-term goal of Eye
Society is to put as much programmable
processing as possible in a small and mobile
package, while providing higher-level self-
organizing abilities so that the robots can
form arbitrary task-oriented groups.

The fundamental principles behind the Eye
Society project are:

- It is possible to put sufficient
  processing into an information-
  capture device such that a central
  processing server is not needed for
  many sophisticated tasks.
- A user or programmer does not need
to think in terms of an identifiable
  individual device but of the overall
  system.
- The absence of centralized external
  processing or control – or even a master-
  slave hierarchy among the devices –
  eliminates the possibility of a single point
  of failure, while moving tasks to the
“ecosystem” of cameras improves system scalability, as each sensor brings along enough processing power to handle the data it acquires.

Each camera in the project operates as a fully autonomous entity, with wireless (multicast) communication to other robots and access to “offshore” resources such as databases and file space as required. The current version of the robot controller centers on a commercial processor board containing a StrongARM SA-1110 processor running at 206Mhz, 32MB of flash memory, an SA-1111 companion chip (which allows interfacing to USB devices), and additional analog and digital inputs and outputs used to control locomotion and camera servos and to accept input from sensors such as microphones. An add-on board of our design contains motor drivers. Communication is through IEEE 802.11b.

Because of the limited memory on-board, each robot can mount an external NFS file system if needed to provide access to additional storage space or external data.

Although the processor is fairly modest, we are able to acquire color video at 5 frames/second at a resolution of 320 by 240 pixels; running a simple color-histogram object tracking algorithm on the incoming video slows down the throughput rate by less than one percent, suggesting that the acquisition speed is limited by bus bandwidth rather than processor performance. Complex algorithms that perform more operations per pixel do of course execute more slowly.

The cameras operate in a relatively normal laboratory area, with varying lighting and movable furniture. Since none of the cameras has a fixed position, (although each can run only in one dimension along a linear railing), visual servoing is being developed that allows them to determine their relative positions using fixed points in the environment. We are currently exploring environmentally adaptive ways to do this as a collaborative problem among the robot group, rather than using pre-positioned calibration points.

Allowing the cameras to run autonomously permits us to develop control and visual algorithms with elementary feedback loops. This applies to a single camera, which can adjust its target using both the pan and tilt controls and its lateral movement on the track. In effect this allows the camera to capture and operate on a set of multiple views as it moves under its own control along the track. While an obvious use of the movement is for traditional stereopsis and egomotion algorithms, an additional possibility is that of self-adjusting position and heading to find optimal locations for acquisition and calibration (e.g. locating good three-point perspective), or to compare predicted views generated by rendering a model with actual views and then updating the model to account for the differences.

When this principle is applied to a group of cameras working together things get even more interesting. The ability of the cameras to move and collectively self-adjust allows us to take a more dynamic approach to problems of visual analysis such as occlusion. Similarly when the cameras are operating as an organized group, they can use the pre-computed knowledge of their relative positions and views to provide each other with additional perspective information of the scene from the individual points of view. Besides rather elementary advantages such as compensating for individual failure, or lighting effects (such as specular reflections) that disproportionately affect one camera...
position, this also allows hypothesis-testing results to be computed for objects in the scene from other cameras’ perspectives.

Current work with the robots is aimed at resolving the group calibration problem, and exploring mechanisms that allow the robots to maximize the information provided by their shared view of the space in which they are positioned.

Longer-term work includes increasing the processor power of the robots to allow more sophisticated real-time processing, incorporating additional sensors such as microphones, and enabling other sorts of devices (both stationary sensing devices and unconstrained-motion floor robots) to become part of the acquisition-and-understanding “ecosystem.”

REFERENCES


