Moving Towards Massively Scalable Video-Based Sensor Networks

Wu-chi Feng, Jon Walpole, Wu-chang Feng
Department of Computer Science and Engineering
Oregon Graduate Institute
{wuchi, walpole, wuchang}@cse.ogi.edu

Calton Pu
College of Computing
Georgia Institute of Technology
calton@cc.gatech.edu

1. Introduction

Networking and computing technologies are becoming advanced enough to enable a wealth of diverse applications that will drastically change our everyday lives. Some past examples of these developments include the World Wide Web and wireless data networking infrastructures. As is quite obvious, the World Wide Web has enabled a fundamental change in the way many people deal with day-to-day tasks. Through the web, one can now make on-line reservations for travel, pay bills through on-line banking services, and view personalized on-line newscasts. More recently, developments in wireless technologies have enabled anywhere, anytime access to information over wireless medium. As wireless technologies such as 3G continue to advance, the ability to support larger bandwidth applications will become possible.

Sensor networks are becoming a fairly hot area of research where the main focus is the development of networking technologies that support potentially thousands of sensors placed in a chosen environment [Estrin99]. Thus far, the sensors that have been described in the literature (and in some limited use) typically measure simple things such as humidity, temperature, or pressure. This results in a fairly limited amount of data generated, even over thousands of sensors. Now, if we look ten years into the future when video capture devices will most likely be small and inexpensive (with limited processing capabilities), the ability to create video-based sensor networks will be possible.

Video-based sensor networks can be used for a great number of applications that would undoubtedly revolutionize the way we go about our day-to-day lives. Two such examples include:

- **Automated video surveillance and notification systems** - automated video surveillance systems are useful for military and non-military applications alike. In the battlefield, rapidly deployable video surveillance systems could allow for the tracking and monitoring of troops and enemy activity. In buildings with classified information such as the Pentagon or industrial development labs, being able to monitor suspicious behavior and tracking of events could help track visitors and employees, signaling any alarms when necessary. In day care centers, such systems would allow anxious parents to follow their children around during the day.
- **Video and computer assisted living** - one can imagine a scenario where a home or office may be instrumented with a large number of these simple video sensors. Working together, this system can identify people within the home and actively track them as they move throughout the environment, providing services that make life easier such as automatic lighting. By having multiple camera angles available wherever the occupants are, techniques developed for gesture recognition become much easier, making the human computer interface more natural. For example, assume that the system can track a person continuously once authenticated. As the person reaches for a door knob (gesturing the intention to open the door), the system could verify the user has appropriate rights to open the door and can unlock it automatically without the user ever needing to take a key out.

One can easily imagine other applications where large-scale, distributed, video-based sensor networks (VBSN) could be useful. While many of the basic vision technologies exist for tracking and identifying human motion in video data, there remains a tremendous amount of research that needs to be solved beyond the current thinking in sensor networks in order support massively scalable, autonomous video sensing systems. We believe this research falls into three broad categories:

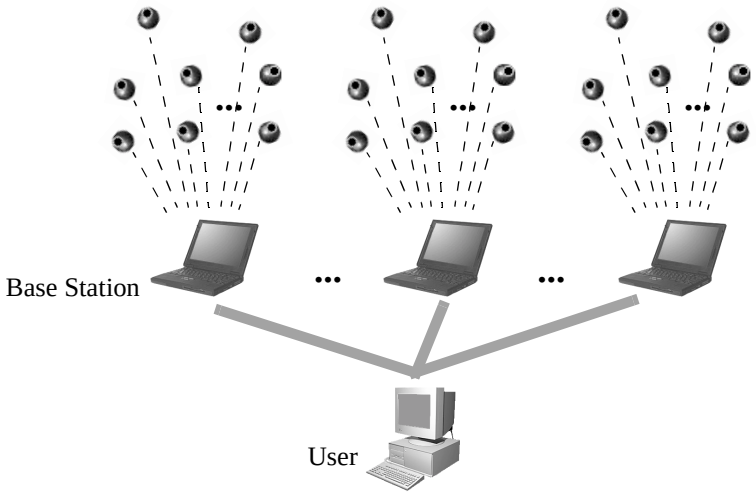

Figure 1: This figure shows an example sensor network with small video sensor devices that are inexpensive to build and have limited capacity to forward data to a base station which serves as an aggregator of data. Users access these base stations for information that they require.

- Designing "many-to-one" or "many-to-many" information flow topologies between sensors and users as well as sensor to sensor.
- Designing massively scalable infrastructures that support real-time and stored video movement and coordination.
- Designing autonomous, coordinated sensors that are sensitive to both the environment and available power.

In the rest of this white paper, we will describe the computing and networking technologies that need to be developed in order to deploy massively scalable VBSN's. We will first describe our vision of VBSN architectures and will then present several broad areas of research that will need to be pursued.

2. A Vision for Video-Based Sensor Network Architectures

Our vision of future VBSN architectures is shown in Figure 1. The VBSN is comprised of three primary elements: video sensors, base stations, and user end-systems.

Each video sensor will be responsible for capturing a single view of the physical world. We expect that the number of video sensors could number in the tens of thousands with smaller groups under the control of a single base station. Further, we expect that each video sensor will be approximately an inch or less in every dimension and will be equipped with a video capture device and a wireless transmitter to deliver the data to the base station. The video sensor may potentially contain mechanisms to orient the camera as well as zoom capabilities at a slightly higher cost. For transmitting the actual video data, a number of technologies may be used including a Bluetooth-like transmission protocol, radio, or cellular telephony. We say Bluetooth-like because we believe that the limited transmission range (10 meters) of the current Bluetooth specification may not provide sufficient reach for each base station and its set of sensors. One indication that such devices will be possible was the recent announcement by Toshiba that described their TC35273XB chip which contains an integrated memory, encoder, and decoder for MPEG-4 audio and video [Shim01]. Because the memory is integrated into the processing chip, it is touted to be more power efficient for use in small appliances like cell-phones. We believe that derivatives of such technologies will also enable video-based sensor networks. While we focus primarily on the video aspect of the sensor, we believe that these sensors may also potentially be augmented with additional sensor capability that measure physical parameters such as temperature, humidity, or pressure.

The base stations are higher-powered, mobile devices that aggregate the video information from the various sensors. The purpose of the base stations is three-fold:

(i) They provide a central store for the cameras to store their information. This store alleviates the video sensors of having any backing mechanism and significant computation power, making them as inexpensive as possible. We expect that the base stations will contain a fairly substantial backing store, enabled by technologies such as the IBM's microdrive. The microdrives themselves can currently hold a gigabyte of data on a drive that measures less than 0.19 in. by 1.7 in. by 1.5 in. and at a cost of less than $300. As drive technologies continue to advance, we expect that extremely large stores will be feasible on a mobile platform.

(ii) They provide more computing power for aggregation of video data. The base station may contain filters for aggregating and distilling the large amounts of video data. Thus, the base station is the first level at which higher-level knowledge will be applied to the video data. Computation at the base station may involve simple artificial intelligence techniques that provide context-based coordination of multiple video feeds.

(iii) They provide easier mechanisms to provide for long-term monitoring. The deployment of VBSN's may be in environments that don't have power readily available such as the monitoring of river basins and battlefield situations. Because the number of base stations is expected to be small relative to the number of video sensors, we expect that the base stations may be equipped with a solar panel that allows the battery in the mobile unit to be recharged as sunlight is available. These solar-based devices can currently be purchased to run a laptop-like machine for less than $300. The advantage of having the solar-power driven device is that the video sensor networks can be deployed without regard to the power-line ownership. Obviously, in single ownership facilities such as a building this is not a concern.

In summary, the base stations are expected to have more processing power, have a large backing store, and will have more power available to it than the video sensors.

At the top level are the processes that the users require such as automated tracking (for use in surveillance), image understanding (for use with traffic routing and accident detection), and gesture recognition (for computer assisted living spaces). These systems are expected to be wired for both power and networking. Thus, we expect that they may have tremendous amount of capabilities in relation to the base stations.

3. Research Directions Needed to Support Massively Scalable Video Sensors

There are a large number of technological advances that are required to enable video-based sensor networks, ranging from low-level circuit design and packaging to high-level image understanding and processing techniques. Some of the technologies are already under development including wireless networking technologies and small video capture and compression devices. In the rest of this section, we describe three broad areas of systems and networking-related research that we see as requirements for building massively scalable VBSNs.

3.1. Vision-Based Information Fusion

For video-sensor technologies, one of the key primary developments will be in managing a tremendous number of sensors, both human controlled as well as under autonomous control using context-based information (i.e. the objects in the video data). The underlying theme here is providing infrastructures and mechanisms that allow aggregation and filtering of potentially thousands upon thousands of video sensors. We break the discussion into two main areas: downstream and upstream video fusion and control.

Downstream video fusion: With potentially tens of thousands of video sensors capturing real-time video data, mechanisms are required that reduce the video streams into more manageable numbers. To make this *information fusion* tractable, we envision using basic vision technologies that exist for tracking objects and providing automated mechanisms that scale to the large numbers of sensors. Among the issues to be solved are:

- *Distributed vision management* - The first step in making use of potentially thousands of video streams is to filter and aggregate the video flows to more useable groups of sensors. The sensors may be statically placed, in which case the inter-sensor relationship is relatively well-known. For this static deployment, mechanisms

are needed to allow a “tracking thread” to move from camera to camera while tracking an object and potentially combine multiple streams to allow more views of a particular object. In addition, sensors may also be placed in a more ad hoc fashion under rapid deployment conditions or terrain monitoring. Under these conditions, the system would need to be augmented with ad hoc group management policies.

- *Compression management* - In order to have some hope of managing large amounts of video data, compression of the video data is a must. The design of video compression algorithms needs to be optimized along several dimensions. First, they need to result in highly compressed video streams with high fidelity (possibly lossless). Second, they cannot consume large amounts of power as the video sensors are expected to last for a long time after being deployed. Thus, algorithms that are DCT-based as in the MPEG compression standards may not be suitable for deployment in these devices. Alternative compression algorithms such as wavelet-based compression algorithms need to be examined. Third, while high levels of compression are required, they also need to provide for some levels of adaptation and network availability (e.g. the wireless bandwidth is expected to vary quite wildly over time). Finally, there need to be mechanisms that allow for some aggregation of data to occur in the compressed domain to help speed and minimize power drain on the infrastructure itself.
- *Low power image processing algorithms:* In order to minimize the amount of data transmitted over scarce wireless links and to maximize the scalability of the VBSN, some of the power in the video sensor needs to be used to filter the incoming video data. This entails re-examining many of the traditional vision algorithms that exist and framing them in a power-sensitive way. Thus, the ability to do tracking on a limited power budget will only help the long-term survivability of the video sensors themselves. To the extent that generic filters can be built (and extended to support specific-domains) will undoubtedly help minimize the deployment time required.
- *Massively scalable networking quality of service*: It is expected that the number of sensors will be on the order of tens of thousands, while sharing potentially limited wireless resources. Thus, we envision that massively scalable network management will be a key to allowing the video sensors to transmit their data to the base stations. This includes both quality of service issues as well as protocol design issues.

Upstream video control: In creating the VBSN, mechanisms also need to be put into place that allow potentially few people to automatically control thousands of video sensors. In particular, one might envision a user saying “track that person” and the infrastructure converts this automatically into a “tracking thread” that manages multiple cameras (potentially controlling individual cameras if equipped with panning, zooming, and orientation control). Thus, the distributed vision management, as described above, would have to export interfaces that allow a command to be issued to sensors such as “track that object” or “inform me when something comes into view in these cameras”. Thus, software mechanisms need to be developed that allow multiple “threads of monitoring” to coordinate themselves and that allow triggers to be constructed and fired based upon the movement of an object within the network. Again, a large portion of this work will have to be focused on the systems infrastructures (including massively scalable group membership protocols and monitoring thread handoff mechanisms) necessary to group and maintain the sensors in a sensible way.

3.2. Avoiding Information Implosion

The base stations in the VBSN are the first level at which, higher-level understanding of the aggregated video data will be assigned. Each base station is expected to handle 100’s to 1000’s of video feeds simultaneously, all while managing potentially limited availability of power. Among some of the major threads of research to be conducted in order to avoid information implosion are:

- *Object filtering and network management:* At a higher layer, the base station should be able to filter and prioritize information for the users that will be accessing the information. This filtering may include prioritizing the objects within the system with the highest priority objects being assigned the most resources. This also implies that the video algorithms may potentially have to operate with limited access the video data itself. Along with filtering, the processes running on the base stations will need to actively manage the availability of networking resources between the sensors and the users. Obviously, for applications where data is very important, over-provisioning to some degree will be expected. Nevertheless, with the amount of

data to be managed at each base station network adaptation will probably be necessary in addition to the object-tracking work already described.

- *Stream storage and retrieval mechanisms:* We envision that the VBSN will be used in two primary modes: on-line monitoring and forensic reconstruction. In on-line monitoring, a single user may be controlling 1,000s of cameras through simple commands such as “track that vehicle”. Thus, the number of actual video feeds that might be viewed simultaneously might be less than ten in number. For this event and other events that are currently not in “focus”, mechanisms need to ensure efficient storage and retrieval of the video streams when needed. This is not simply a video-on-demand problem. Rather, *active thinning* technologies need to be developed that start to automatically thin the video data stream when system resources are calculated to be running low. As an example, the video streams that are coming in from the sensors to the base stations may already have some of the data filtered out (due to lack of a significant event in view). The base station may further reduce the stream (e.g. reducing redundant views, etc.). As the captured data ages, the base station may actively thin the streams by transcoding a 30 frame per second video into 15 frame per second video when it is a day old. As the video continues to age, they may be thinned further by the system. For applications that are mission critical where the users want *all* the data stored (and have the resources to do so), we envision *active thinning* as a mechanism that essentially helps cache video data from the tertiary storage, allowing users to more quickly manage a tremendous number of video streams.

3.3. Power Computing

For deployment in power-limited environments, the base stations are expected to be the first down-stream hosts with solar-backed battery power. In these situations, *power computing* will be necessary. Much of the traditional wireless and mobile research tends to focus on only low-power or full-power scenarios. With solar-backed devices, the power is expected to vary depending on the whether it is sunny or cloudy, or day or night.

- *Base station management:* Above all else, the base station need to remain available for long-periods of time (particularly when wired power is unavailable). Under these conditions, it is imperative that all the image processing, tracking, filtering, and network management activities work under potentially variable amounts of processing power availability. Under conditions where the power is plentiful (i.e. sunny for solar-backed base stations), the algorithms should be allowed to consume as much power as needed, keeping in mind the processing requirements of the other algorithms that may be running.

3.4. Other Complementary Technologies

We expect that the development of other related technologies may also be useful to VBSNs. Two such examples include visualization techniques and vision techniques. Visualization techniques could efficiently map the events that are being viewed by the video sensors to be represented in a coherent way to the users, allowing them to quickly change focus to events that may have a continually rising priority. Vision-based techniques would allow automatic recognition of objects.

4. Conclusion

In this white paper we have described a video-based sensor network architecture for video surveillance and environment monitoring. There are a tremendous number of systems and networking related problems to solve in order to make the video-based sensor networks feasible including: low-power video computing, aggregation of video flows, massively-scalable low-power networking, and distributed computing.

5. References

Estrin99 Deborah Estrin, Ramesh Govindan, John Heidemann, Satish Kumar, “Next Century Challenges: Scalable Coordination in Sensor Networks”, in *Proceedings of Mobicom 1999*.

Shim01 Richard Shim, “Toshiba Announces Handheld Video Chip”, *ZDNet News*, January 15, 2001.