ISSUES IN THE TRANSMISSION, ANALYSIS, STORAGE AND RETRIEVAL OF SURVEILLANCE VIDEO

D. Doermann, A. Karunanidhi, and N. Parkeh
Laboratory for Language and Media Processing, University of Maryland, College Park

M.A. Khan, S. Chen, H. T. Ozdemir, M. Miwa, and K.C. Lee
Panasonic System Solution Development Center of USA

ABSTRACT

Increased network capabilities and the prevalence of wireless networks have provided an environment where data and compute intensive surveillance applications can be realized as part of existing infrastructures. Nevertheless, careful consideration must be given to how such systems are designed and implemented. In our efforts to consider pervasive surveillance applications, this paper highlights some of the primarily issues related to the transmission, analysis, storage and retrieval of surveillance video and briefly describes our framework for video surveillance from mobile devices.

1. INTRODUCTION

One of the ultimate objectives of video surveillance systems is to provide an automated, highly reliable and complete end-to-end solution which can both monitor real-time activities and provide efficient archival for later analysis. These systems should provide effective transmission of video (possibly via low bandwidth connections from remote sites), visualization of video scenes to aid real-time monitoring, extraction of objects and events in a real-time to facilitate access, processing of the extracted objects and events, and efficient storage and rapid retrieval of data and associated metadata.

Traditionally surveillance systems are set up as closed circuit television (CCTV) systems with a large number of cameras, switches and recorders hooked up to a central bank of monitors and controllers over high bandwidth connections. With technology advances in Internet and IP technologies, surveillance systems are now becoming “networked” which opens up a tremendous number of opportunities for providing low cost systems through existing infrastructure, remote systems thru wired or wireless networks and the ability to use digital video analysis embedded in the network. In this paper, highlight some of the main issues required for moving to a digital IP environment including encoding standards, content delivery, streaming video and quality of data required for analysis, indexing and retrieval.

2. APPLICATIONS

There has been a great deal of literature on automated video surveillance that gives insight into real-time monitoring of security environments [1-4]. Automated processing seeks to supplement the traditional human monitoring and video archival scenarios with systems that can provide real-time metadata to supplement the video stream. Often, this type of information is the result of detection of activity where activity is otherwise not expected, recognition of activities, faces recognition, tracking and application specific processing. The technical challenges when using a multisensor approach, for example, include the ability to actively control sensors to cooperatively track multiple moving objects; fusing information from multiple sensors into scene-level object representation; and providing human users with a high-level interface for dynamic scene visualization and system tasking.

Although systems strive to provide these services for live real-time surveillance, there is an equally important task of providing an environment to retrieve archived video. Surveillance video is often used to identify the presence of particular individuals or to provide “evidence” of missing persons, theft or property damage. In many cases, such as a car thief who quickly picks a lock, a human observer may have trouble detecting the crime as it is happening. Systems properly equipped however, can rapidly search through archives to gather additional information. Unfortunately, time is the primary search parameter in current systems. Nevertheless, as more sophisticated analysis algorithms are incorporated, systems will be able to provide real-time deterrents.

Research in these application areas have focused on issues related to sensors and capture, automated detection and analysis of people and activities, networking, storage and retrieval. Before describing our work, we will highlight some of the key areas to be considered including content delivery, content analysis, storage and retrieval and issues related to IP based systems. For our discussion, we will assume that we are dealing with digital video assets.

3. CONTENT DELIVERY

In digital surveillance and security systems, video information may be either stored and viewed locally or captured and transmitted to a remote location. With the expansion of networked capabilities to the “last” mile, remote video surveillance is becoming a practical reality. In many cases, however, storage media and/or communication channels limit the amount of information that can be captured or transmitted. It is typically necessary to
compress video to reduce the amount of storage space that would be needed to save the video information. This leads to a tremendous number of questions as to what content is really necessary.

The general concept of video adaptation can be applied to encompass ways of changing the video to adapt to the network, the user, or the application. It is useful to consider the distinction between procedures that attempt to manipulate the underlying data directly (data adaptation), such as compression or reformatting, and those procedures that manipulate the underlying content (content-based adaptation). Previous work in our group has focused on providing dynamic adaptation of content, based on ranking of relevant video frames [5]. Assuming that standard compression techniques are exhausted, the idea is to provide a high quality temporal summary of the video by ranking frames and presenting only those which contain new information. The idea has been proposed for transmission over low bandwidth networks and can also be applied with more traditional content analysis techniques.

4. CONTENT ANALYSIS

In any indexing, retrieval or browsing application an important consideration is the need to have underlying metadata which supports access. For surveillance, the types of queries we are interested in typically require the recognition of certain activities or objects (people, vehicles, etc) in the scene. Although basic metadata about time can be useful for manually navigating the collection, intelligent systems will move toward identifying the content for efficient storage and as a pre-processing for activity analysis by subtracting the background [6,7,8], detecting and classifying motion [9,10] and recognizing key activities and objects such as people. Face detection research has evolved to a level where commercial systems are being applied with limited success in many environments. When general content analysis advances to the point where activity of interest can be reliably identified, the amount of data that needs to be stored will be further reduced.

5. STORAGE, INDEXING AND RETRIEVAL

Consider a scenario in which ten cameras are mounted in a remote location for an outdoor video surveillance. One minute of NTSC uncompressed video (2 bytes per pixel) from the camera requires approximately 1.2 GB while one minute of MPEG-2 compressed video with data rate of 4Mb/sec requires somewhat less, 30Mb. Even for a compressed video, however, one month of uninterrupted video surveillance from a single camera can require over 1.23 TB of disk space. If our goal is to provide efficient storage, indexing and retrieval, consideration should be given to the types of information we ultimately want to retrieve and how we want to retrieve it.

Video compression schemes can be classified into two approaches: scalable and non-scalable coding [11]. A non-scaleable video encoder generates one compressed bit stream while scaleable video encoder compresses a raw video sequence into multiple sub-streams. MPEG-1 uses non-scaleable video encoding while MPEG-2 allows both scaleable and non-scaleable coding. Due to object-based representation of MPEG-4, scalable coding is more suitable. Scaleable coding is also suitable for video over IP used for surveillance. QoS can be addressed using scaleable coding where only the base sub-stream is transmitted when there is congestion in the network or when there are no alarm situations. Pons et al [12] have developed a fast algorithm to detect motion in the compressed domain for video surveillance. Shaun and Gary [13] provide a framework for real-time on-site video image storage that enables increased functionality in the retrieval of video images. Users can retrieve live or archived video images through a client workstation that communicates with the off-site server over the public Internet.

A video storage system called PRISM [14] has been built on Linux that provides better response times and throughput. The major components are QoS context management module, a scheduler to schedule I/O requests to the storage system and application layer for providing integrated services. PRISM provides sufficient separation of service classes and it is possible to simultaneously achieve multiple performance (in particular, bandwidth and delay) objectives in a single storage system.

Video visualization describes the joint process of analyzing video and the subsequent derivation of representative visual presentation of the essence of the content. Visualization consists of video retrieval methods like text based query, content based query, metadata based query and video browsing. Jung et al. [15] proposes an efficient method for semantic video retrieval based on the object motion trajectory for visual surveillance applications. This search scheme also supports various queries based on trajectory information of the video object. Marir et al. [16] discuss an event based indexing and retrieval for surveillance video.

Video browsing often provides a high-level representation of video in the form of visual summaries. One common approach is to have a video browser in which thumbnail images are tiled together, usually arranged in a time order or based on the key frames to give an overview of the visual content in the video sequence. Fong et al. [17] propose a content-based video sequence representation which helps in scenario understanding and can be employed in real-time for security monitoring systems. The algorithm comprises spatio-temporal video object segmentation, feature extraction and scene analysis.

6. IP BASED SYSTEMS

We are currently developing a media processing framework that addresses many of the issues described above, that will allow mobile devices to make use of network based storage and processing to overcome many of the challenges faced by the limited resources of remote surveillance applications.
The goal is to provide a set of services that will facilitate various content adaptation, network video analysis and ultimately, pervasive applications. This will allow us to address scenarios where either the user has access to a security system through limited resources (such as a guard with a PDA over a wireless network) or where the surveillance system is monitoring a remote site, and the data needs to be analyzed, transmitted and stored at the far end of a limited bandwidth link.

The major components of the framework are the display engine, component engine, storage management engine, video analysis engine and a transcoder (shown in Figure 1). The component engine and display engine are associated with the client application in the mobile device, while the storage management engine, video analysis engine and transcoder are associated with the media-processing server. The client application in the mobile device interacts with the camera by means of a MMAPI (Mobile Media Application programming interface) [18]. The MMAPI is a simple, lightweight package and it provides multimedia functionalities for the mobile devices with which the developers can gain access to native multimedia applications on the device.

![Figure 1: Media Component Architecture](image)

The media-processing server has a transcoder which transcodes the video from one format to another, so that the mobile device at the receiver end can interpret and display it. The storage management engine manages the storage and retrieval of the media content in media processing server. Video content is then retrieved for video summarization and indexing by the video analysis engine. The video analysis engine processes the video and transmits the summarized video to the mobile device. The client application receives the video and requests services from the component engine for video playback and the video player is generated dynamically based on the playback application.

We have implemented a Java version of the Media Engine along with a simple database application where a user can send a media stream from mobile device on to the media processing server or retrieve a media object from the server and view it. A diagram is shown in Figure 2. We have used an IPAQ version on Windows CE platform with pocket camera attached to it and wireless connection so as to store/retrieve the files to/from the server. The interface also allows the user to display images, and store them in one of two folders, either a local folder or a remote folder. The remote folder will be used to test our server side communications. The images are stored in the MySQL server in blob format. Initially, the device sends a device identifier to the multithreaded server and the server sends back an XML file to the client that has the list of capabilities for the mobile device. Since the mobile devices are resource constrained devices, we have only the minimal setup on the client and the other functionalities can be downloaded as components when required. Once the client gets these capabilities, it is able to capture, store and retrieve the images.

To test the media analysis capabilities in a server setup, we are integrating our existing text detection and recognition process so that the user can request that a captured scene image be processed and the results returned to the client. The application has been implemented in Java to run across various platforms.

7. SUMMARY AND DISCUSSION

Future video surveillance systems will focus on using intelligent cameras where the local processing layer uses digital compression methods to save bandwidth resources. These cameras will support background subtraction as well as on demand adaptation of image resolution so that the high resolution image can be provided when needed. In general, the goal will be to provide an organized representation that can be accessed by content to perform both real-time analysis (alerts) as well as post access (querying for specific people, situations or activities, for example). Our work will focus on providing a low level toolkit for processing of surveillance video with the plan to integrate these lower level analysis components into a full intelligent indexing and retrieval system.
8. REFERENCES


