

S3-R1: The IBM Smart Surveillance System- Release 1

Contact: Arun Hampapur,
IBM T.J. Watson Research
Center
19 Skyline Drive, Hawthorne, NY
10532, (914)-784-7440,
arunh@us.ibm.com

ABSTRACT

The increasing need for sophisticated surveillance systems and the move to digital surveillance infrastructure has transformed surveillance into a large scale data analysis and management challenge. Smart surveillance systems use automatic image understanding techniques to extract information from the surveillance data. While the majority of the research and commercial systems have focused on the information extraction aspect of the challenge, very few systems have explored the use of extracted information in the search, retrieval, data management and investigation context. The IBM smart surveillance system is one of the few advanced surveillance systems which provides not only the capability to automatically monitor a scene but also the capability to manage the surveillance data, perform event based retrieval, receive real time event alerts thru standard web infrastructure and extract long term statistical patterns of activity.

Categories and Subject Descriptors

I.4.8 [IMAGE PROCESSING AND COMPUTER VISION]: Scene Analysis- *Tracking, Object Recognition*. H.3.1 [INFORMATION STORAGE AND RETRIEVAL]: Content Analysis and Indexing-- *Indexing methods*

Keywords

Video Surveillance, Object Tracking, Privacy Preserving Surveillance, Surveillance Event Retrieval, Smart Surveillance.

1. INTRODUCTION

Smart Surveillance is the use of automatic video analytics to enhance effectiveness of surveillance systems. In this paper we demonstrate a system which analyses activity in the monitored space in real time, and makes the events available for generating real time alerts and content based searching in real time. Section 2 of the paper describes the capabilities of the IBM Smart Surveillance System Release 1 (IBM S3-R1). Section 3 describes the architecture of the system. Section 4 describes the test deployment scenario in which the system is currently being tested and presents some results from that environment. Section 5 describes the demonstration that will be presented at the conference.

2. IBM S3-R1 CAPABILITIES

The IBM Smart Surveillance System (S3) is a middleware offering for use in surveillance systems and provides video based behavioral analysis capabilities. Release 1 of the Smart Surveillance System provides two components

- Smart Surveillance Engine (SSE) which provides the front end video analysis capabilities
- Middleware for Large Scale Surveillance (MILS) which provides data management capabilities.

These two components in conjunction with IBM DB2 and IBM WebSphere Application server support the following features.

- Local Real-Time Surveillance Event Notification: This set of functions provides real-time alerts to the local application which is running the SSE.
- Web Based Real-Time Surveillance Event Notification: This set of functions provide a web based real-time event notification within 3 seconds of the occurrence of a specified event within the monitored area, for example Speeding Vehicle
- Web Based Surveillance Event Retrieval: This set of functions provides the ability to retrieve surveillance events based on various attributes like object type / speed.
- Web Based Surveillance Event Statistics: This set of functions provides the ability to compute a variety of statistics on the event data. For example, distribution of people arriving and leaving a building over a day.

3. SYSTEM ARCHITECTURE

Figure 1 shows the high level architecture of S3-R1. The video from the surveillance camera is processed in parallel by the IBM Smart Surveillance Engine (SSE) and the video compression engine (VCE). The SSE produces a rich index which describes the events occurring in the scene. Middleware for Large Scale Surveillance (MILS) is a middleware layer which stores the compressed digital video from the VCE on a video management server. MILS also stores the meta data generated by the SSE is stored on a database server. The MILS middleware also provides the end user application with the services needed to receive real time alarms and perform content based queries on event data.

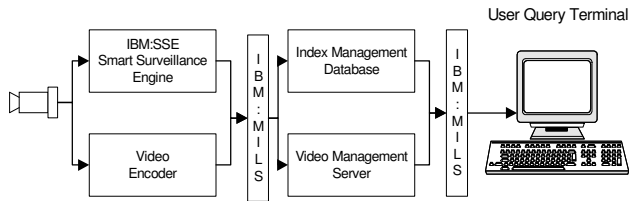


Figure 1: Block diagram of the IBM Smart Surveillance System. The video from the camera is simultaneously processed by the IBM:SSE (Smart Surveillance Engine) and encoded by the encoder. The index data generated by the IBM:SSE is loaded into a database server by IBM:MILS (Middleware for Large Scale Surveillance), which also provides the search functionality

3.1 Smart Surveillance Engine

The IBM Smart Surveillance Engine (SSE) is a software-only video based event detection technology which delivers “smart” video surveillance capabilities. The SSE is designed with the intent of making currently deployed surveillance systems “smart”. The IBM SSE is based on the following key video analysis technologies [1,2,3,4].

- **Object Detection:** This set of technologies can detect moving objects in a video sequence generated by a static camera. The detection techniques are tolerant to changes in natural lighting, reasonable changes in the weather, distracting movements (like trees waving in the wind), and camera shake.
- **Object Tracking:** This set of technologies can track the shape and position of multiple objects as they move around a space that is monitored by a static camera. The techniques are designed to handle significant occlusions as objects interact with one another.
- **Object Classification:** These technologies use various properties of an object including shape, size and movement to assign a class label to the objects. Typical labels include, Person, Group and Vehicle.

The SSE provides the following functionalities.

- **Real Time Video Based Alerts:** These are alerts which depend solely on the movement properties of objects within the monitored space. Examples include 1) Motion Detection: 2) Directional Motion Detection 3) Abandoned Object Alert 4) Object Removal and 5) Intentional camera movement or blinding:
- **Viewable Video Index (VVI):** The SSE detects and tracks all moving objects within the cameras field of view. The SSE creates the VVI as a set of XML documents. The VVI encodes all the interesting activities in the video including, 1) Number of objects in a scene 2) Object Class, the current engine classifies objects in Single Person, Group of People and Vehicles, 3) Object appearance properties, including color, texture, shape, size and their changes over time, 4) Object movement properties, including position, velocity and trajectory, 5) Occlusion parameters when objects are in an occlusion, 6) Background changes due to changes in lighting and stopping of moving objects, 7) Event information: any events that may be flagged by the engine. The VVI index is also ideal for monitoring

activities over a mobile wireless device such as a PDA, because of its extreme low bandwidth requirements (~10MB/hour). The index encodes all the activity that occurs in the video in terms of evolving background models, evolving object models and motion trajectories of each of the moving objects. The information contained in the VVI can be used by the an application to render the activity in the video independent of the original video stream.

3.2 Middleware for Large Scale Surveillance

The MILS middleware provides the data management services needed to build a large scale smart surveillance application. While MILS builds on the extensive capabilities of IBM’s Content Manager and DB2 systems, it is essentially independent of these products and can be implemented on top of 3rd party relational databases. MILS provides the following functionalities

- **Converting the VVI to Relational Tables:** MILS is capable of taking the VVI generated by the IBM:SSE and inserting the data index into a relational database, thus allowing for SQL based querying of the index data by an application.
- **Query Services for Surveillance Data:** MILS provides applications with a set of services for querying the surveillance data. The queries supported by MILS cover a wide range of commonly used queries. The following are types of queries that will be supported by MILS. Each of these query types is illustrated by an example. The basic unit of storage in the database is a continuous movement event, for example, a car entering the camera field of view at time T1 and moving continuously to leave the field of view at T2, would be represented as a basic unit called an activity. All query responses are a set of activities.
 1. **Time Queries:** Show all activities between (7 PM Apr 17 2004) and (9 PM Apr 17 2004).
 2. **Object Size Queries:** Show all activities where the object size O is ($S1 < O < S2$)
 3. **Object Class Queries:** Show all activities with objects of class C1. Current set of classes are limited to Vehicles, Single Person and Group.
 4. **Object Motion Queries:** Show all activities where the object moved with speed V where, ($V1 < V < V2$) or where an object moved in a direction D (specified as an angle in the image plane).
 5. **Context-based object content similarity Queries:** Show all activities where there were “blue cars” or “cars similar to this car –here the user specifies an example car through an image”.
 6. **Queries based on region/location of interest:** Show all activities that occurred within a specified polygonal region.

7. Combination Query: A query which combines the one or more of the above specified query types.
8. Interactive object/event classification based on relevance feedback
9. Data Mining Services: MILS will support the following data mining applications.
 - Site activity summarization
 - Site activity pattern discovery
 - Automatic abnormal event/object detections

4. TEST DEPLOYMENT SCENARIO

The IBM S3-R1 system is being continually tested in the parking lot of the IBM Watson Center in Hawthorne, NY. Figure 2 shows the plan view of the Watson facility with a parking lot attached to the building. A security analyst would be interested in several types of activities including, finding cars speeding through the parking lot, finding cars that have been parked in loading zones, etc.

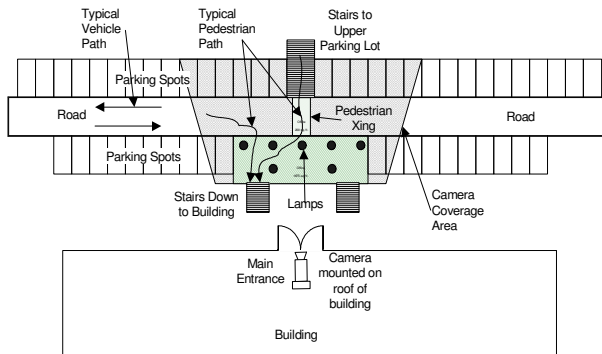


Figure 2 Figure 11: Plan view of the setup for monitoring the parking lot at an IBM facility. The camera, which is mounted on the building roof top is monitoring the entrance area to the building and a segment of the parking lot attached to the building. Typical activity in the monitored area includes cars driving by, cars parking in a spot, people exiting / entering cars and people walking in and out of the building.

The camera which is mounted on the roof of the building is wired to a central server. The video from the camera is analyzed by the smart surveillance engine to produce the viewable video index which is stored in the index database which is a commercial database system like IBM DB2. The video from the camera is also independently encoded by the video encoder and stored on a video server like the IBM Video Charger. An application can launch SQL queries against the index database to locate events of interest (like locating all speeding cars between 9AM and 10AM on Jan 13th 2004). The events located in the index database are always associated with a pointer to the actual video data on the video server which can be used for browsing purposes.

Following is a presentation of results of performing queries against our experimental activity database. This activity database was generated based by monitoring activity from our experimental setup at an IBM facility as described in figure 2. The database captures activity from 12.00 AM Jan 13th 2004 to 12 AM Jan 14th

2004. Figures 3 and 4 show the output of queries and the distribution of arrival of people into IBM Watson Research on Jan 14th.

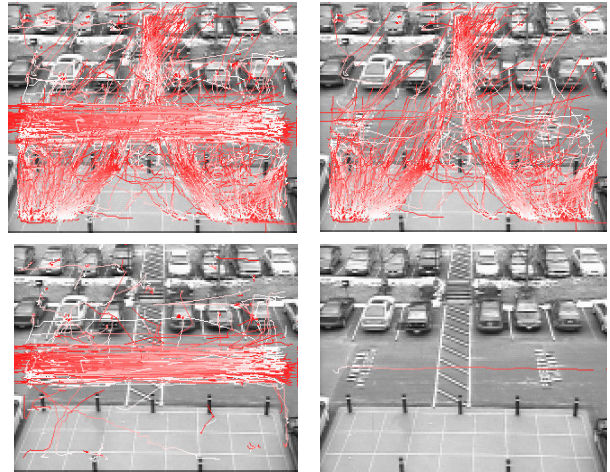


Figure 3: Top Left: Tracks of all moving objects during the 24 hour monitoring period. Top Right: Tracks of objects with (50 pixels < Area < 1000 pixels) yielding only the people walking through the parking lot. Bottom Left: Tracks of objects with (1000 pixels < Area), yielding only vehicles. Bottom Right: Track of a speeding vehicle (Velocity > 10 pixels / frame ~ covering 10 m in 2 secs ~ 11.25 MPH). All tracks in the diagram are color coded, tracks are colored white at the beginning and gradually turn red as time progresses.

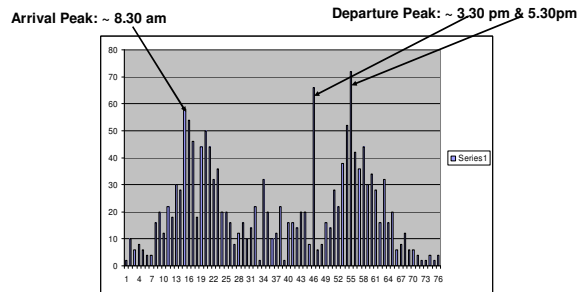


Figure 4 Shows the temporal distribution of tracks corresponding to people (arriving and leaving the building). The tracks were selected using an area based query (results presented in Figure 14 Top Right). The peak of the arrivals is around 8.30 am in the morning and the two significant departure peaks are around 3.30 PM and 5.30 PM approximately.

5. DEMO DESCRIPTION

The demo at ACM Multi-Media 2004 will demonstrate remote surveillance of the Watson Center parking lot. Users at the conference will be able to do the following.

- Real Time Alerts: Users will be notified in real time upon the occurrence of certain selected events, for example when a large vehicle drives into the lot.
- Event Retrieval: Users will be able to retrieve a variety of events from the parking lot, for example all cars that arrived between 10 am and 11 am.

- Event Statistics: Users will be able to get the statistics of certain events that are occurring in the parking lot, for example the arrival and departure distribution of people at the Watson center on a given day.

6. REFERENCES

- [1] A. Hampapur, L. Brown, J. Connell, S. Pankanti, A.W. Senior, and Y-L Tian , Smart Surveillance: Applications, Technologies and Implications, IEEE Pacific-Rim Conference On Multimedia, Singapore, Dec
- [2] A. Hampapur et al, The IBM Smart Surveillance System, Demonstration at IEEE CVPR 2004.
- [3] PeopleVision: Demo videos of on going work at IBM research: www.research.ibm.com/peoplevision
- [4] A. Senior, A. Hampapur, Y-L. Tian, L. Brown, S. Pankanti, and R. Bolle. Appearance Models for Occlusion Handling. Proceedings of Workshop on Performance Evaluation of Tracking and Surveillance (PETS2001), December 2001.