S3: THE IBM SMART SURVEILLANCE SYSTEM: FROM TRANSACTIONAL SYSTEMS TO OBSERVATIONAL SYSTEMS

A. Hampapur, S.Borger, L.Brown, C.Carlson, J.Connell, M.Lu, A.Senior, V.Reddy, C.Shu, Y.Tian
Exploratory Computer Vision Group, IBM T. J. Watson Research Center

ABSTRACT

Pervasive sensor based systems are transforming Information Technology systems from being ***transactional*** *in nature to being* ***observational*** *in nature. Observational systems are inherently distributed and capture information at a much finer grain of space and time. Enabling and building such systems also poses many technology challenges, extracting information from sensor signals, indexing and searching sensor meta-data, data mining and scalability. In this paper we use S3: The IBM Smart Surveillance System as an example of an observational system to explore several of these issues through real world deployment examples.*

Index Terms- Observational systems, video surveillance, smart surveillance, intelligent video, security systems.

1. INTRODUCTION

The increasing need for better security, the dropping costs of cameras, RFID and other sensors, and the drop in cost per MIPS of processing power have provided the ideal conditions for developing large scale distributed observational systems. The term **observational systems (O-Sys)** was first coined and used in a presentation by Jain [6] and is defined as a sensor based event capture and storage system. As such systems become pervasive; in addition to enhancing security they also provide significant operation benefits. In this paper we use the learning from the development and deployment of the IBM Smart Surveillance System (S3) in various customer environments as the basis for the vision of Observational Systems (O-Sys). The goal of this paper is to explore the benefits of observational systems both from a security and operational perspective and to explore the technical challenges in developing and deploying such systems.

2. SCENARIOS FOR OBSERVATIONAL SYSTEMS

Consider a retail department store, which has deployed a digital video surveillance system in the store, covering a significant fraction of the floor area with cameras that are being recorded and monitored. Let us consider a typical customer, customer #2342, who enters the store @10.23 AM, pauses before a special promotion item on his way to the shoe section, where he tries on shoes (2 pairs) and is assisted by an associate after a 10 minute wait. He later spends 25 minutes in the sporting goods section before heading to the check out register where he buys ~$154 worth of merchandize after waiting for 40 seconds in line, leaving the store at 11.17AM. The total visit to the store has lasted 54 minutes.

TABLE 1: Transactional View of Customer #2342 captured by Point of Sale System.

Cash Register #, associate ID	22, 2363
Frequent Customer Number	23A5389
Check out start time	11.07 AM
Shoes – Nike # XXXX	$85.99
……….	………..
Total	153.78
Tender Credit Card	VISA # XXXX
Checkout Complete	11.16AM

TABLE 2: Observational View of Customer #2342 captured by Digital Video Surveillance System.

Cam#	Location	Event	Time	
22	Door 1	Customer Entry	10:23	
22	Door 1	Customer at map	10:24	35 secs
24	Jeans	Customer views disp	10:26	1 min
….	………	………	…	……
54	Checkout	Customer Arrives	11.06	
54	Checkout	Customer Waits	11.07	40 secs
55	Checkout	Associate interrupted	11.10	5 mins
55	Checkout	Customer leaves	11.16	
28	Door 2	Customer exits store	11.17	

Table 1 captures the information logged into the Point of Sale System when customer #2342 checks out merchandize in the store. Table 2 captures the activities of customer #2342 that can be observed by a person monitoring the DVS system. Clearly the information captured is largely complementary. While the point of sale (transactional system) logs the "sales as reported by the checkout clerk for customer #2342" during the 9 minutes that the customer spends at the checkout. The DVS system (observational system) provides insight into the customer activity over the complete 57 minutes the customer spent in the store. While the information captured on the POS system can readily be used for a number of data mining activities, the information is limited to the explicitly executed transactions. For example, the sales logged for customer #2342 could be different from the actual merchandize that customer #2342 takes out of the store – this is a common form of fraud called sweet-hearting where the customer acts in collusion with the clerk. Both such "sweet-hearting" activity and the fact that the customer waited 10 minutes for help in the shoe aisle are captured by the DVS system. The key distinction between a transactional system and observational system is the fact that a transactional system requires explicit human input to record information, while the observational system automatically captures all the activity that occurs within the specified limits of the sensor. While the DVS system captures all the details of the customers visit, it requires a person to watch the video to extract this information. Using people to monitor video cameras is known to be ineffective [7] and off-course cost prohibitive. In the past decade automated video analysis and pattern recognition technology has been applied to surveillance cameras (called video analytics) [3,4] making it possible to extract much of the information in table #2 automatically, *thus enabling the first "observational systems –based on digital video surveillance and other sensor infrastructure"*

3. OBSERVATIONAL SYSTEM STRUCTURE

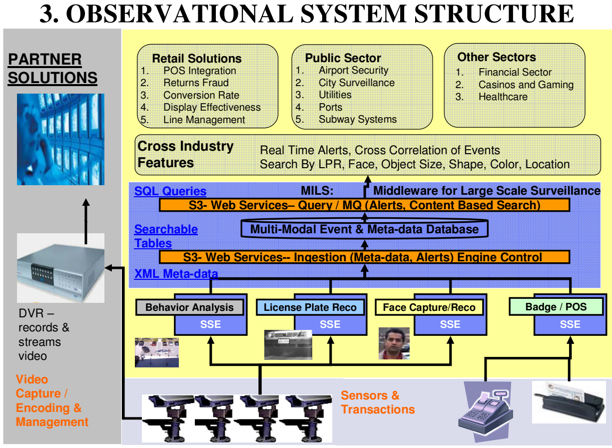

Figure 1 An example observational system: IBM S3: An open architecture for extensible event based surveillance.

The IBM Smart Surveillance System (called S3 –shown in figure 1) is an architectural framework for event based surveillance. S3 is a real world observational system. Using standard video cameras in conjunction with existing Digital Video Surveillance infrastructure, S3 provides the following functionality [10]. **Real-time Alerts** including, motion detection, trip wire, abandoned object, object removal, camera move/blind, e.g. sound alert when person approaches the fence. **Search Functionality** including search by object presence, time, location, color, speed, event duration and object size, e.g. find all red cars that entered the parking lot. **Specialized Analytics** including face capture and license plate recognition. The **Cross-Correlation** function of S3 allows the user to correlate events across analytics, e.g. search for license plate #, relate it to parking lot activity and badge's information. S3 supports statistics of events and extensibility.

A generic observational system has five distinct layers of technology on top of the sensors.

Layer 1: Data Capture Layer: This layer includes technologies which take the raw sensor data and convert it to digital and store it on a disk and provide live streams for monitoring. S3 relies on partner technologies which already exist for layer 1 (cameras from company X, DVR from company Y).

Layer 2: Data Analytics Layer: This layer consists of signal processing, pattern recognition and learning technologies, which process the sensor data to extract meta-data. S3 implements this layer thru Smart Surveillance Engines (SSEs) which is a plug in framework of analytics algorithms like object detection, tracking, license plate recognition, face recognition, object attribute extraction (color, shape, size, etc). Each SSE can generate real-time alerts and generic event meta-data. The meta-data generated by the SSE is represented using XML. Table below shows snippets for two types of meta-data generated by 1) SSEs with behavior analysis plug-in and 2) SSE with license plate recognition plug-in.

Behavior Meta-data	**License Plate Meta-data**
Engine ID: 23	Engine ID: 35
Unique Event ID: 2379406	Unique Event ID: 4926402
Start: 9/10/06:02:22:15:100	Start: 9/10/06:02:12:15:100
End: 9/10/06:02:22:55:300	End: 9/10/06:02:12:25:453
Keyframe: 23567.jpg	Keyframe: 563783.jpg
Video : //mils/xx/file1.wmv	Video : //mils/xx/file3.wmv
Object Type: Person	License Plate #: 525sds
Additional Fields:	Additional Fields:

Layer 3: Meta-data Management Layer: The meta-data management layer aggregates meta-data from multiple sensors-analysis modules across the system. S3 implements this layer thru MILS (Middleware for Large Scale Surveillance) a Java framework. The XML meta-data is received by MILS and indexed into predefined tables in the IBM DB2 database. This is accomplished using the DB2 XML extender. This allows for fast searching using the primary keys. MILS provides a number of query, retrieval and real-time alert notification services based on the types of meta-data available in the database.

Layer 4: Data Mining Layer: The meta-data accumulated across multiple sensors & transactions, over extended periods of time provide a very fertile ground for extracting insights into trends that are occurring in the observed environment. S3 does not implement any data mining capability at this point.

Layer 5: The Solution Layer presents all of the data gathered by an observational system to the user in the context of the business application. An number of industry specific solutions have been implemented using S3.

4. O-SYS TECHNICAL CHALLENGES

Figure 2 below shows the architecture for implementing S3 across a chain retail store with both local and centralized monitoring of events. The video is stored locally at the store while the event meta-data is centralized on a regional basis. The monitoring of alarms and investigation functions are centralized. Given this architecture, there are several dimensions of technical challenges.

Sensor Analytics Algorithmic Challenges: Clearly, this is a significant challenge, especially for video cameras. The current state of art can deal with scenes involving sparse activity and can extract limited types of information from them. As observation systems become pervasive, the demand for finer grained information extraction increases. For example, retail operators as of today can count customers entering the store using cameras. However, the value of understanding demographics of customer, age, socio-economic status, group shopping behavior, expressions of person looking at a new display, fraud behavior of employees and customers, etc is significant. The utility of observational systems depends largely on the development of algorithms to extract such complex information from video and other sensors.

Processing Power Challenge: The video analytics algorithms are very processor intensive. The current model for video analytics is centralized processing. As O-Sys's become pervasive, the processing is being moved into dedicated processors (like the TI-DSP) and new specialized media, processors like the IBM CELL [5]. The trend is also to move these processors to the edge (into the cameras). However, as the algorithms get more and more sophisticated, processing will need to be distributed between the edge and central site, bringing in a number of new challenges in distributed processing.

Data Management: Back end scalability: Unlike transactional data, observational meta-data can be voluminous, especially if "retro-active searching for unanticipated events" is supported by the observational system. A typical camera in a retail store (which see's 1000 objects per day for an average of 1 minute) would generate 1.60GBytes of meta-data. These numbers are significantly larger for a city surveillance application. Clearly, the back end system for handling this type of meta-data will need to deal with a number of issues including, meta-data management, effective search and indexing structures, etc.

Information Integration: As seen in the retail example, while the observational system can capture very fine grained information, the value of this information greatly increases, once it is related to the corresponding transactional activity For example, relating the people count within the TV department in a store to the total number of TV's sold provides a metric very useful in retail. The technical challenge here involves designing the right architecture to integrate information and the algorithms that can tie the appropriate piece of observational data to transactional data.

Data Mining: One of the biggest values of accumulating observational (and transactional information) is the ability to extract trends from the data and the ability to automatically detect anomalies, while this technology has been extensively researched for transactional systems [1], the addition of observational data brings in a whole new dimension of probabilistic trend extraction to the picture.

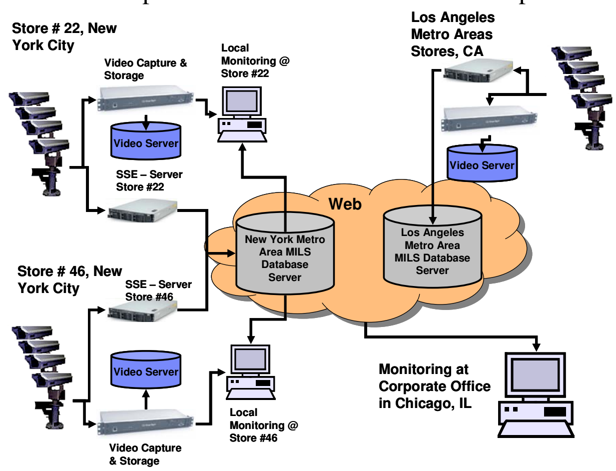

Figure 2 S3 based distributed surveillance system architecture

5. O-SYS: IMPLICATIONS

As observational systems become more pervasive they have significant implications on the following.

Security Implications: Clearly, observational systems can be used to enhance the security of critical facilities in various ways, including providing real-time alerts on know security violations (like perimeter breaches), enhancing investigative capabilities (like searching for green get-away vehicle in the vicinity of a crime scene) and potentially moving towards proactive measures based on intelligence gathering (a van has been consistently seen around the embassy premises). Additionally, the activity trend information that gets aggregated in an O-sys over a period of time can be used for planning purposes to enhance security.

Operational Implications: A retail store may deploy camera infrastructure to prevent "loss or theft", however, the camera covering the checkout (which is one of known locations for loss in retail—thru cashier fraud) can also be used to measure the wait-time at the checkout, thus providing operational benefits to the store. Similarly airports can use cameras at the security check, to capture facial images to match against a database and use the same cameras to manage the wait time at security checkpoints

Privacy Implications: As observational systems (cameras & other sensors backed by event detecting intelligence) proliferate, these systems begin to accumulate large amounts of information about activities of people in a wide range of places, work, stores, airports, shopping malls, city streets, etc. Today the information gathered by O-sys is "anonymous", i.e. "the camera used to count people at the entrance to the retail store – is not capable of uniquely identifying people", this situation is likely to change in the future due to the following two reasons

- Improvements in non-intrusive biometrics technology, like face recognition, gait recognition etc.
- Cross linking of Observational data to Transactional data. For example, linking the anonymous "trajectory of person X thru the store" to the credit card tendered at checkout..

Thus as observational systems evolve, the need to incorporate privacy safe guards into these systems becomes essential. Several approaches to enhancing privacy have been presented in [9].

6. RESULTS

In this section, we provide a high level description of deployment of S3 in a retail store environment and discuss the various functions being provided by S3 (security and operational). There are several cameras at a chain retail store that are currently being monitored by the S3 system. The S3 system is providing the following functionality.

Returns Fraud Investigation Tool: In this application, S3 is being used to provide a tool for matching people in different parts of the store. Specifically, S3 captures customer presence events at the returns desk, correlates it to returns transactions, and allows the investigator to easily find the same person entering the store. The tool is being used to investigate the occurrence of return fraud [2] – specifically, to check weather the merchandize being returned was actually carried into the store by the customer.

Customer Count at Entrance: Here S3 is being used to count the number of customer / customer groups that are entering the store. These counts are then related to the number of sales in the store. Table 4 shows the customer counts over a week in April 2006. The number of customers almost double over the week due to a promotional sale by the retailer.

Display Effectiveness: Here S3 is being used to monitor selected display cases, to understand display effectiveness which is measured as the ratio of (*Number of customer spending time at the display)* **to** *(total number of customers passing by the display.)*

Enhanced Customer Service: S3 is being used to provide real-time alerts when there are customers in selected parts of the store, and there is no sales associate present in the area.

The key point of the S3 retail pilot is the fact that the camera and observational system are being used not only to address "security –or theft issues" in the store, but also operational issues. This theme has been consistent in other customer deployments as well.

TABLE 4: CUSTOMER ENTRANCE COUNT AT STORE

DATE	ENTRANCE COUNT
4/23 (Sunday)	**1055**
4/24 (Monday)	**882**
4/25 (Tuesday)	**999**
4/26 (Wednesday)	**748**
4/27 (Thursday)	**1331**
4/28 (Friday)	**1812**
4/29 (Saturday)	**2148**
4/30	**1495**

7. CONCLUSION

The pervasive deployment of camera, digital video and sensor infrastructure brings to fore the challenge of "effectively using" the video/sensor data. This has lead to the creation of a number of "video/sensor analytics technologies " which address point problems like behavior analysis, license plate recognition, etc mainly aimed at enhancing security. As discussed in this paper, camera/sensor and analytics technologies with the right architectures lead to the creation of "observational systems", which create a whole new class of information which is complementary in nature to that generated by existing transactional systems. The key to successful observational systems is an open and scalable architecture, which inherently supports the large scale distributed nature of sensor based observation. This paper has presented the IBM S3 system as an example of an observational system. The technical challenges of scalability at various levels, the sensor data analysis, back end meta-data indexing and management, data mining and vertical solution development need to be addressed for Observational Systems to become as pervasive as Transactional Systems.

8. REFERENCES

1. C. Apte. Data Mining – An industrial research perspective. IEEE Computational Science and Engineering, April-June 1997
2. CNN, Ex-White House aide arrested in alleged refund scam, Saturday, March 11, 2006, http://www.cnn.com/2006/US/03/11/claude.allen.arrest/
3. R. Collins, et al. `A system for video surveillance and monitoring', VSAM Final Report, Technical Report, CMU-RI-TR-00-12, May 2000
4. Arun Hampapur, et al, Smart Video Surveillance, Exploring the concept of multiscale spatiotemporal tracking, IEEE Signal Processing Magazine, March 2005.
5. IBM CELL Processor: http://domino.research.ibm.com/comm/research.nsf/pages/r.arch.innovation.html
6. Ramesh Jain, Observational Systems for GRV, Presentation at DARPA, EC ASSIST Kickoff Meeting, May 2005, Washington DC.
7. Miller et al, Crew fatigue and performance on US coast guard cutters, Oct 1998, US Dept of Transportation.
8. S3: Demo videos of on going work at IBM research: www.research.ibm.com/s3
9. A. Senior, S. Pankanti, A. Hampapur, L. Brown, Y.-L. Tian, A. Ekin, Blinkering surveillance: Enabling Video Privacy Through Computer Vision, IBM Technical Report, Vol: RC22886, 2003
10. Shu et al, IBM Smart Surveillance System (S3): A Open and Extensible Framework for Event Based Surveillance, IEEE AVSS 2005, Como, Italy.