Automated video surveillance: challenges and solutions. ACE Surveillance (Annotated Critical Evidence) case study

Dmitry O. Gorodnichy and Tony Mungham

Laboratory & Scientific Services Directorate,
Canada Border Services Agency
79 Bentley Ave, Ottawa, Ontario, K2E-6T7, Canada
Email: Dmitry.Gorodnichy@gmail.com / Tony.Mungham@cbsa-asfc.gc.ca

ABSTRACT

Currently deployed video surveillance systems and protocols are not fully efficient. In real-time monitoring mode, the problem is that an event may easily pass unnoticed due to false or simultaneous alarms and lack of time needed to rewind and analyse all potentially useful video streams. In archival mode, video data storage and manageability is the problem that makes post-incident investigation very difficult. - Due to the temporal nature of video data, it is very difficult for a human to analyse video data within a limited amount of time.

This paper presents an automated video surveillance technology named ACE Surveillance (Annotated Critical Evidence) that is developed by the National Research Council of Canada (NRC) for the purpose of enabling more efficient use of surveillance systems. This technology, which incorporates recent advances in objects detection and tracking, has been tested on several real-life long-term monitoring assignments, including an over a year testing with the existing CCTV surveillance cameras at the NRC campus. The results of these tests are described. – While quantitatively showing the advantage of using automated evidence extraction systems for enhanced security and providing a reference standard for measuring Intelligent Video systems available on the market, the presented study also exposes several problems related to development and deployment of such systems. Further steps for integrating automated evidence extraction systems for mainstream security applications are discussed.

1. INTRODUCTION

As a result of the increasingly growing demand for security, many countries have been deploying closed circuit television (CCTV) surveillance systems as in important tool for enhancing preventive measures and aiding post-incident investigations. Within Canadian government, specific interest in these systems is seen from Transport Canada who have announced multi-million funding towards surveillance technologies for rail and urban transit security in Canada and issued the first CCTV Reference Manual for Security Applications [1,2] and from Canada Border Services Agency (CBSA) who sees video surveillance as a key technological element in protecting the country borders. Many other federal departments have also expressed the need in video surveillance technology [3].

CCTV surveillance is used in three modes of operation: active, passive, and archival (through recordings). Active monitoring involves trained personnel who watch video streams at all times. Passive monitoring is most common and involves employees who watch video streams in conjunction with, or incidental to, other duties. In the third mode, CCTV systems record video data for the purpose of post-event analysis (Figure 1).

While evaluating the utility of CCTV surveillance systems, it has been realized and emphasized that currently deployed surveillance system and protocols are not fully efficient for either of described modes of operation [1-5]. In real-time monitoring mode, the problem is that an event may easily pass unnoticed due to false or simultaneous alarms and lack of time needed to rewind and analyse all potentially useful video streams. In archival mode, video data storage and manageability is the problem that complicates the most the efficiency of post-incident investigation. - Due to a temporal nature of video data, it may take very long for a human to analyse it. Besides, video files may be stored on different media and/or compressed using different proprietary algorithms. This was clearly seen from the London bombing video backtracking experience which reports that manual browsing of millions of hours of digitized video from thousands of cameras proved impossible within time-sensed period.

The solution to these problems is seen in developing Intelligent Video Surveillance systems capable of performing automated analysis of video data.. Such systems however, as reported in CCTV manuals [2], are still not considered mature by the majority of CCTV users and are therefore not incorporated with currently existing surveillance setups and protocols.

In order to improve the situation, the Video Recognition Systems team of the National Research Council of Canada’s Institute of Information Technology (NRC), which was working on developing video recognition solutions for Canadian Space Agency (Canadarm-2 project) and Ottawa health services (Nouse “Nose as mouse” project) , has geared its work into security and surveillance domain. As a result, an automated video surveillance technology based on the recent advances in object detection and tracking, named ACE Surveillance (Annotated Critical Evidence) has been developed [4-5]. Implemented as a software that runs on an ordinary desktop computer, ACE Surveillance performs real-time analysis of captured video streams for the purpose of automatically extracting and annotating Critical Evidence Snapshots, which are used to automatically alarm the system and which allow efficient summarization and browsing of captured video data.

In January of 2007, as a pilot project with the goal of testing the technology on real-life assignments, the NRC Administrative Services and Property Management Branch endorsed installing ACE Surveillance software with its existing CCTV systems inside the NRC campus. This paper presents the results obtained from this, almost a year long, testing. These results confirm that affordable and efficient automated video surveillance is possible. They also provide a reference standard against which can be measured other intelligent video solutions such as those coming from private industry.

The paper is organized as follows. Section 2 describes in detail ACE Surveillance technology. Section 3 presents the results of technology testing on real-life long-term assignments. The last section summarizes the important lessons learnt from the current study and presents the outlook for further integration of automated evidence extraction systems such as ACE Surveillance for mainstream security applications.

2. ACE SURVEILLANCE

ACE surveillance technology is based on the concept of *Critical Evidence Snapshot* (CES) defined below.

Definition: *Critical Evidence Snapshot (CES)* is defined as a video snapshot that provides to a viewer a piece of information that is both useful and new.

Definition: A surveillance system that deals with extraction and manipulation of Annotated Critical Evidence snapshots from a surveillance video is defined as *ACE Surveillance*.

The architecture of the *ACE Surveillance* system is shown in Figure 1. It consists of the *ACE Capture* module, of which there can be several -- each connected to one or more video-cameras and performing real time video analytics, and the *ACE Browser* module, which operates either remotely or locally and which manages the data captured by the ACE Capture module.

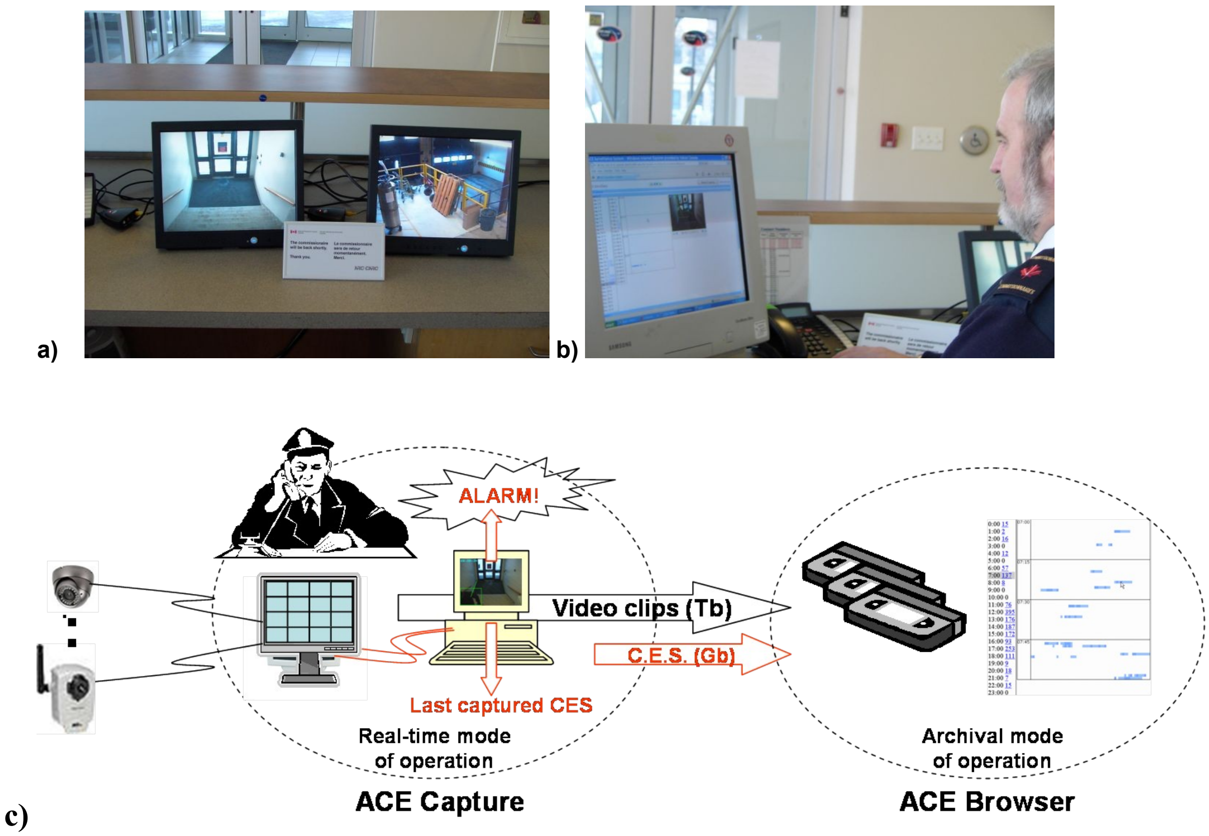


Figure 1: Monitoring with a) conventional CCTV surveillance system - A dedicated officer has to look at the monitors at all times, and b) with ACE-equipped system - In real-time mode: alarm sounds and the last captured CES is shown; In archival mode: "zoom on the evidence" – zoom on day, on hour, then on event – to quickly localize an abnormal event. c) ACE Surveillance architecture.

2.1 ACE Capture

ACE Capture software runs on a regular desk-top computer (with processor speed of at least 800MHz, RAM of at least 512 Mb and Hard drive space of at least 20Mb) under Windows 2000 or XP operating system. It runs with USB cameras or existing CCTV surveillance systems, where the video signal from the CCTV camera or monitor is send to a computer via an USB digitiser.

As a result of running the software: 1) the alarm sounds when a new object is detected or a new person arrives, and 2) Critical Evidence Snapshots (CES-es) are automatically detected and saved with time-stamps on a local machine or (and) remote server. Each CES is provided with visual and text annotation that enables instantaneous data comprehension and enhances manageability of archival data. CES Annotations include: a) the location and the size of the moving object (person) – shown by a rectangle, and b) the direction where the object(person) came from and its/his/her velocity - shown as a line on the image.

More specifically, for each video frame, ACE Capture module performs the following six tasks in real-time:

Task C1: Detect moving object(s) in video based on colour, motion and background information.
Task C2: Compute the attributes of the detected object(s) such as: object location, size, shape, velocity and colours distribution.
Task C3: Based on the attributes computed in C2, recognize object(s) or event(s) as either new or already seen.
Task C4: If either an object or an event is new (i.e. object attributes have changed in a non-linear fashion), label a frame as CES;
Task C5: For each CES, create graphical annotations based on detected object attributes, such as the outline around the objects and their velocity vectors. It also creates a text annotation, saved in a separate text file, to be optionally used by post-archival search techniques
Task C6: a) For each detected CES, the best quality high-resolution frame is saved along with the low-resolution copy of it augmented with the graphic annotations; b) For other frames with moving objects present, only resolution-reduced annotated image is saved. c) Frames where moving objects have not been detected are not saved.

The final output of the ACE Capture software is a succinct summarization of the entire surveillance video sequence, that can be efficiently browsed with the ACE Browser and using the *Zoom-on-Evidence* technique.

2.2 Zoom-on-the-evidence with ACE Browser

The ACE Browser is responsible for preparing the archived CES data for efficient browsing and searching. It is designed with the goal of making it easy to locate an abnormal event among the immense amount of stored data. The *Zoom-On-Evidence* browsing technique is developed for this purpose described below.

When a camera is selected for browsing, all activity detected in that camera is displayed in a hierarchy-summarized form: from day-by-day view (shown in the left column as number of CES-es captured for each day), to hourly view (shown in the second left column as number of CES-es captured for each hour for the selected day), to minute-by-minute graphical view shown as a window, the width and height of which corresponds to the number of minutes (rows) and number of seconds per minute (columns), filled with blue stripes that indicate the detected activity. The colours of the stripes indicate the type of event detected: the darker the box, the more consistent the velocity of the object. The CES corresponding to each stripe can be displayed on the left, by moving the mouse cursor over the stripe.

If there is more than the usual number of CES-es for any specific date, one can “zoom in” on that day by clicking on it, to see the number of CES-es for each hour of that day. One then can “zoom in” on a particular hour by clicking on that hour, to obtain the graphical presentation of evidence for this hour (See Figures 2-4). In this hourly summarization of captured activities shown as blue stripes, with the activities that are more likely to contain importance evidence (based on the attributes of the detected objects) shown in a darker blue colour.

This graphical hourly window-drawn activity summarization is very descriptive and is sufficient in many cases for the viewer to decide whether to go the next hour or click on any particular event slice. The coloured information helps to efficiently localize a minute when an event of interest has happened. The operator can then “zoom in” on a particular minute by clicking on the corresponding stripe, to see the annotated CES captured at that minute. – This image is of low-resolution (for faster browsing) and shows the detected objects (circumscribed by rectangles) and their velocity vectors. In case the detected objects or their actions need to be examined in further detail, the final “zoom-in” on the snapshot can be performed, by clicking on the annotated CES image, after which an original-size unannotated high-resolution image replaces the low-resolution annotated CES image.

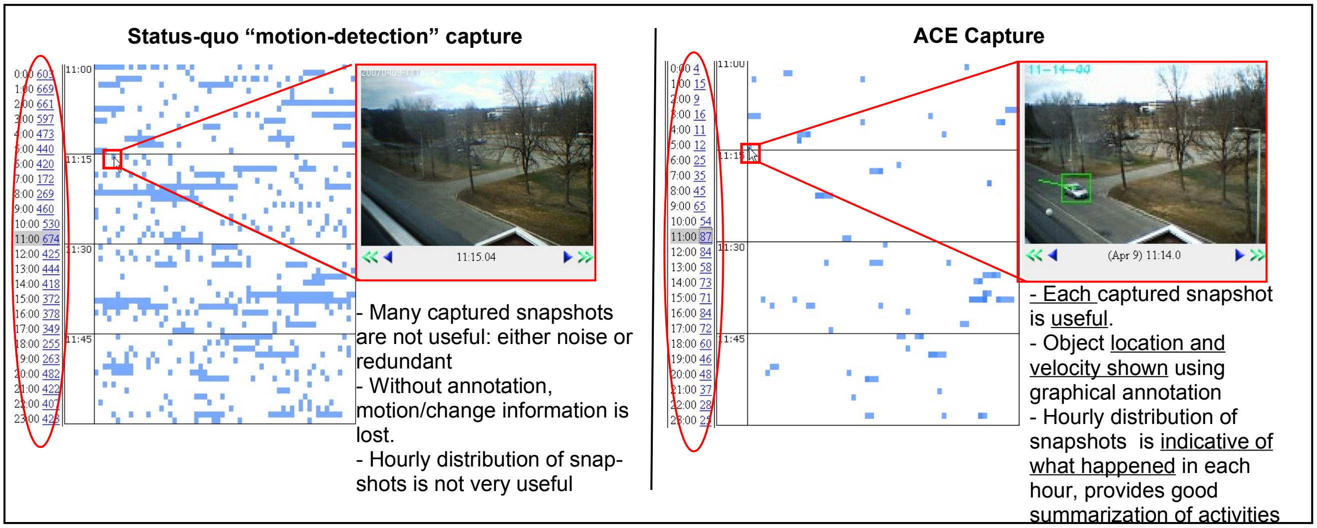
2.3 Critical features of ACE Surveillance

Critical difference #1: The critical difference between ACE Surveillance and conventional *motion-detection-triggered* video capturing devices is that ACE Surveillance is *object-detection-based*, rather then *motion-detection-based*. What ordinarily used motion-detection techniques do is, in fact, not a motion change detection, but *pixel brightness change detection* which *may* be attributed to the motion of an object, but which in many cases is caused by noise in the video image, environmental changes (such rain, tree waiving, light reflections etc) or illumination changes (due to weather, camera automatic adjustments etc). This is why they work poorly in outdoor environments.

In order to really detect the motion of an object, the software has to perform object detection followed by tracking of that object over several frames, which is algorithmically much more difficult than brightness change detection and which is what ACE Surveillance does.

Critical difference #2: The other critical difference is that ACE Surveillance replaces video files with a sequence of graphically annotated snapshots, where the annotations are designed in such a way as to provide an intuitive and natural substitution for the missing video data. Particularly, it can be proved using the theory of visual saliency driven attention that motion-based saliency can be replaced with colour based saliency. Motion-based saliency drives the visual attention of a human but requires lengthy video files to be stored. Alternatively, the colour based saliency can be created by adding extra visually attractive information such as lines and rectangles to the image.

This is demonstrated by the following comparative performance evaluation study.


Status-quo "motion-detection" capture

0:00	603
1:00	669
2:00	661
3:00	597
4:00	473
5:00	440
6:00	420
7:00	172
8:00	269
9:00	460
10:00	530
11:00	674
12:00	425
13:00	444
14:00	418
15:00	372
16:00	378
17:00	349
18:00	255
19:00	263
20:00	482
21:00	422
22:00	407
23:00	423

- Many captured snapshots are not useful: either noise or redundant
- Without annotation, motion/change information is lost.
- Hourly distribution of snapshots is not very useful

ACE Capture

0:00	4
1:00	15
2:00	9
3:00	16
4:00	11
5:00	12
6:00	25
7:00	35
8:00	45
9:00	65
10:00	54
11:00	87
12:00	84
13:00	58
14:00	73
15:00	71
16:00	84
17:00	72
18:00	60
19:00	46
20:00	48
21:00	37
22:00	28
23:00	25

- Each captured snapshot is useful.
- Object location and velocity shown using graphical annotation
- Hourly distribution of snapshots is indicative of what happened in each hour, provides good summarization of activities

Figure 2. Comparison of summarization results (on the same 24-hour outdoor monitoring assignment).

Comparative performance evaluation

As mentioned, the problem with the status-quo motion-detection-based surveillance systems is that they show poor performance when monitoring the outdoor environment. To compare the performance of ACE Capture with motion-detection-based video capture program, we conducted simultaneous several-day-long monitoring tests of the same environment with two identical cameras, one connected to ACE and the other to the competitor's software. Several available on the market motion-based capture programs were tried, the results shown correspond to the best performing of them.

Figure 2 shows the results of these comparative experiments. The figure also summarises the difference in the value of results obtained by either system. The drastic difference in number of captured frames per hour and per day between the two technologies can be seen. Upon closer viewing of the motion events saved by “motion-detection-based” software, one notices many false positives due to light reflection, precipitations and tree waving. Because of these many false positives, conventional “motion-detection-based” surveillance systems cannot be used for the purpose of summarizing the activities. Nor can they be used for alarm notification or event retrieval.

3. TESTING RESULTS

From the prospective of video analytics, the level of complexity of monitoring assignments varies significantly depending on lighting conditions, object motion patterns, camera location and environmental constraints. Figure 3 shows four layouts which differ significantly in the above mentioned factors. The most difficult are those located outdoors in unconstrained environments with little or no object motion consistency (as around a premise). The most easy are those working in controlled indoor environment where minimal direct sunlight is present and where all objects are of approximately the same size and exhibit similar motion pattern (as at access gate inside the premises). All of these layouts have been tested with ACE Surveillance.

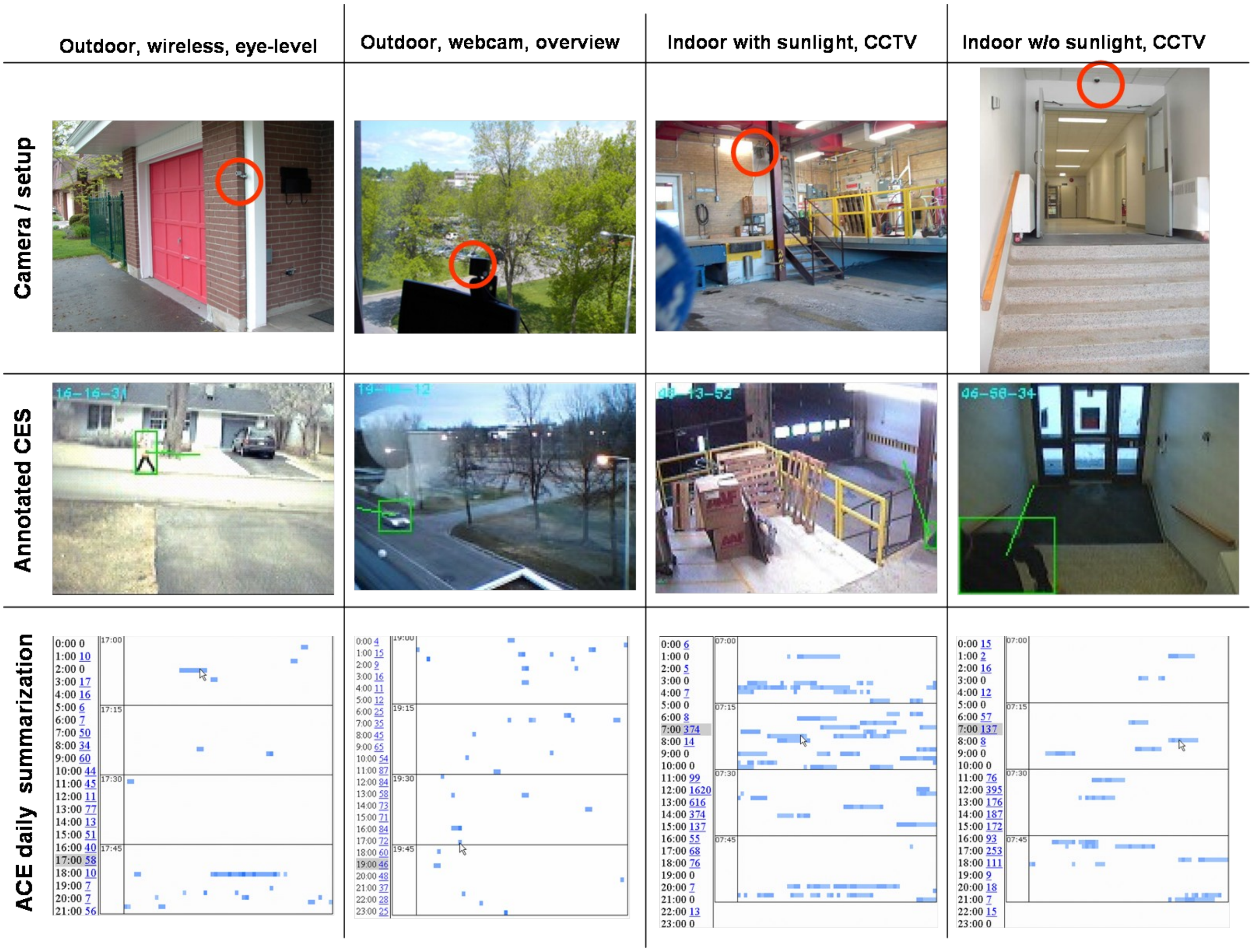
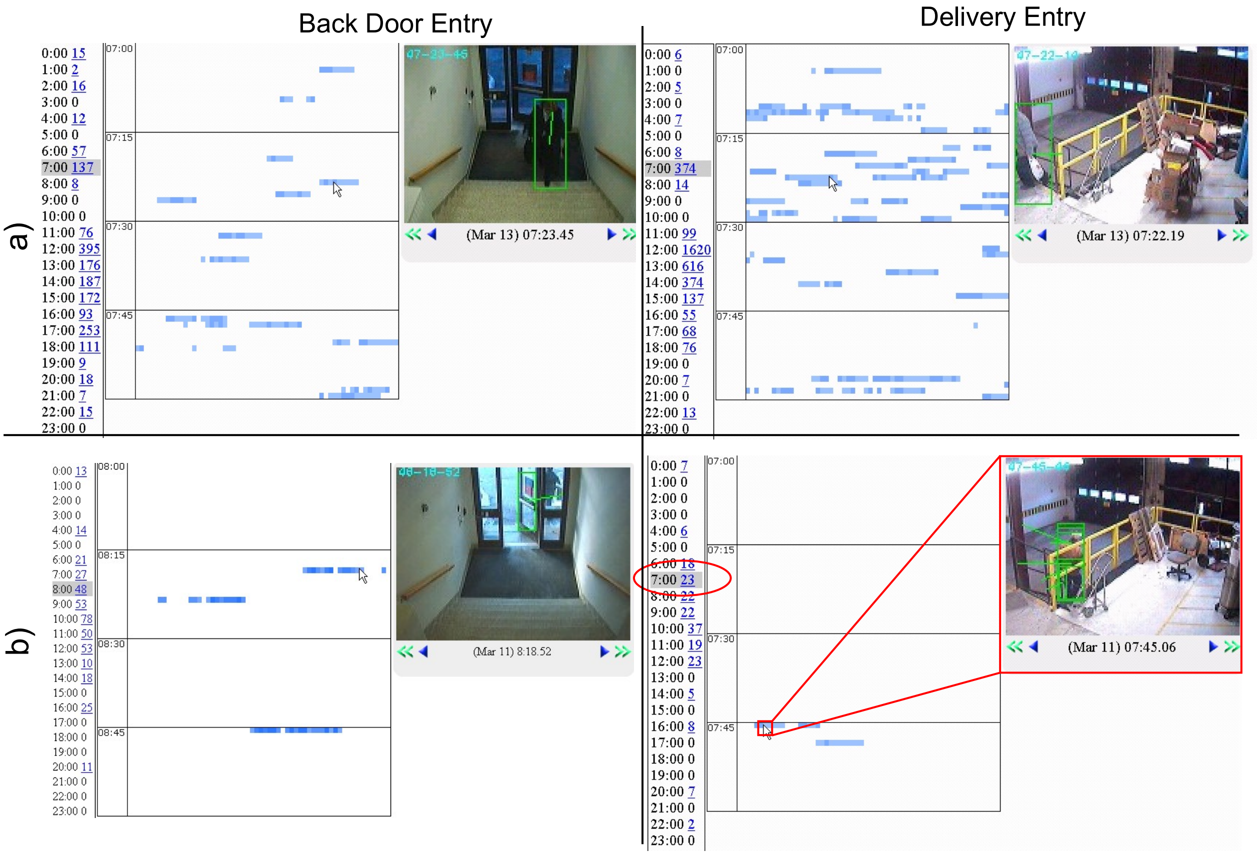


Figure 3: Four types of 24/7 video surveillance layouts (in order of decreased complexity) that have been monitored using the ACE Surveillance. A typical Critical Evidence Snapshot (CES) automatically extracted and annotated by the system as well as a daily event summarization, represented by the distribution of CES-es per hour and within an hour, obtained by the system are shown.

Figures 2-3 show representative results from the tests conducted. For each test, ACE surveillance was run 24/7 non-stop for a long period of time (over 2 or more months), which allowed to see its performance at different times of the day (daylight, dusk, night) and different time of the week (busy vs. less busy days) and year (with different levels of precipitation and sunlight).

To test the operational quality of ACE Surveillance, this technology has been installed at the NRC Commissionaires desktop computers and Commissionaires were given the training on how to use it. These tests have been conducted as a pilot project endorsed by Security Operation of the NRC Administrative Services and Property Management Branch, with the already existing CCTV surveillance cameras (Figure 1 and 3).

Since Commissionaires already have the desktop computers which they use for their daily work and the CCTV monitors have been already installed, the cost of the ACE installation was less the $150 for monitoring two ADT-installed analogue RCA CCTV surveillance cameras that monitored two entrances to the building.



0:00	15
1:00	2
2:00	16
3:00	0
4:00	12
5:00	0
6:00	57
7:00	137
8:00	8
9:00	0
10:00	0
11:00	76
12:00	395
13:00	176
14:00	187
15:00	172
16:00	93
17:00	253
18:00	111
19:00	9
20:00	18
21:00	7
22:00	15
23:00	0

0:00	6
1:00	0
2:00	5
3:00	0
4:00	7
5:00	0
6:00	8
7:00	374
8:00	14
9:00	0
10:00	0
11:00	99
12:00	1620
13:00	616
14:00	374
15:00	137
16:00	55
17:00	68
18:00	76
19:00	0
20:00	7
21:00	0
22:00	13
23:00	0

0:00	13
1:00	0
2:00	0
3:00	0
4:00	14
5:00	0
6:00	21
7:00	27
8:00	48
9:00	53
10:00	78
11:00	50
12:00	53
13:00	10
14:00	18
15:00	0
16:00	25
17:00	0
18:00	0
19:00	0
20:00	11
21:00	0
22:00	0
23:00	0

0:00	7
1:00	0
2:00	0
3:00	0
4:00	6
5:00	0
6:00	18
7:00	23
8:00	22
9:00	22
10:00	37
11:00	19
12:00	23
13:00	0
14:00	5
15:00	0
16:00	8
17:00	0
18:00	0
19:00	0
20:00	7
21:00	0
22:00	2
23:00	0

Figure 4: Summarizations obtained by ACE Surveillance in monitoring the business property: (a,c) on a working day, when many people pass by, and (b,d) on a week-end, when few people should be visiting the premises. Screenshots from ACE-Browser showing the events captured at the main building entry (upper row) and at the delivery entry (lower row) are shown. Consistency in the number of captured evidence for each hour and each day throughout the week, seen in vertical columns of the pictures, is the characteristic of ACE Surveillance technology that makes it very efficient for detecting abnormal events.

Functionality-wise, ACE installation proved to be an excellent value for adding extra security features to the existing CCTV system. The descriptiveness and robustness of the activity summarization as well as detection results can be seen in Figure 4, which shows summarizations obtained by ACE Surveillance in monitoring the business property: (a,c) on a working day, when many people pass by, and (b,d) on a week-end, when few people should be visiting the premises. Screenshots from ACE-Browser showing the events captured at the main building entry (upper row) and at the delivery entry (lower row) are shown. Consistency in the number of captured evidence for each hour and each day throughout the week, seen in vertical columns of the pictures, is the characteristic of ACE Surveillance technology that makes it very efficient for detecting abnormal events.

The main cost however in running the system was found to be the cost of training the personnel and maintaining the system. First, the security officers are generally not required to be computer proficient, and working with more than single Word processing software was found to be difficult for many of them. Illustrative enough is the example when an officer was closing the ACE software window instead of minimizing it every time he needed to switch to another program. Second, since the ACE software is run under Windows operating system, it is susceptible to all problems related to working with this operating system, such as automated restarting of the computer to perform updates etc. As a result, there was no guarantee that the software would work continuously 24/7. The data was often captured with many gaps due to the unintentional termination of the software, and when a break-in into the building happened there was unfortunately no data recorded at that time of the day.

While the ACE Surveillance software GUI can be further improved to take human factors into account, the made observations demonstrated that security personnel are not ready for software-based improvement of their work. Due to the nature of their work, they are prepared to work with all-in-one-box type solutions, which are operated by tangible sliders or buttons, rather than by software program.

4. DISCUSSION

4.1 Lessons learnt

ACE Surveillance is demonstrated as a technology that offers an affordable solution to automated monitoring and surveillance. The benefits of using ACE Surveillance are clear. In case of NRC commissioners, they do not have to look at the monitor at all times and have an easy to browse 24/7 log of all captured activities over the several months. However several problems with the new approach have also been exposed.

The first problem relates to a human factor. – Unless security officers are trained in computers, it might be difficult for them to operate a computer program. The second problem relates to handling and treatment of evidence captured by the ACE surveillance. Can one trust the evidence extracted by a computer and not by a human? From the forensic prospective, data that are not original and have been processed by a computer can not be considered as evidence. There is also a question related to privacy policies. - Surveillance data are normally not saved for a long period of time (usually not more than 1 month), due to their size. ACE Surveillance however allows one to store on a single hard drive many months (and even years) of evidence data, which implies that new protocols for managing such evidence have to be established.

4.2 Importance of the ACE Surveillance study

The ACE surveillance technology presented in this paper demonstrates that affordable and efficient automated video surveillance is achievable. This technology however is not ready for the market yet, since it was created by a research lab rather than by industry. The technology software is developed as a prototype for the purpose of showing what *could* be done in the area of Intelligent Video surveillance, rather than for purpose of selling or deploying it.

A very important objective of this study, which we believe has been achieved, is to provide a reference standard against which can be measured other intelligent video solutions such as those coming from private industry, as well as to help video surveillance users, of which there are many within various federal departments, to deal with common misconceptions and problems related to deploying intelligent video surveillance systems. Some of these are listed below.

The main problem affecting video surveillance users within the federal departments, as reported on the First Federal departments meeting on Deploying Video Technologies for National Security hold last year in Ottawa [3] is the lack of standards and coordination in this area. While there are many departments relying on video surveillance technology, there is practically no department or working group within the government that develops this technology or conducts benchmarking or testing of its performance. For comparison, the working groups on Privacy Issues and Biometrics have been established almost five years ago.

As a result, in the absence of coordinated effort in the area, there have been many local initiatives over the last few years related to deploying video surveillance systems, practically all of which have been influenced by the industry advertising campaigns, rather than by the proper performance evaluation. As a result, it not uncommon to see over 30 different video systems deployed within the same department, with the performance of those systems falling much short of the original expectations.

Because of the influence of market-driven advertisements and solutions, there appeared several misconceptions related to the area of Intelligent Video as well as an overall lack of trust to this area from the current user of video surveillance technologies.

One of the most common misconceptions relates to the notion of “motion detection”, which has been coined by the industry to mislead the user into believing that object motion in a video sequence is performed, which would be an example of the intelligence but which is actually not true, since the only thing that is detected in most cases is the change of brightness in a particular area of the image. Consequently, the so called “motion-triggered” capture commonly used in surveillance and mentioned in CCTV manuals [2] is in fact not an example of an intelligent video solution.

“One-fit-all” solution is another misconception about Intelligent Video solutions. Because of the diversity of scenarios and environmental conditions, there is no and never be a “box-prepackaged” solution for one’s needs. Instead Intelligent Video solution is to be provided with configurable software or a library, for tuning of which some video recognition expertise will be required.

The peculiarity of Intelligent Video solutions comes from the fact that for humans performing video recognition tasks is very easy. Therefore, as many people think, for computers these tasks should also be easy to perform. To avoid the fate of “artificial intelligence” which has become a dishonoured term in late sixties, after it has been realized that computers are very far behind humans in intelligence, it is important to understand what Intelligent Video can do and what it cannot.

More video data (better resolution or compression) do not imply easier or better video intelligence decisions. In fact, it is the opposite. While better quality of video image is needed for forensic purposes as evidence, it does not help make video solutions more intelligent. The results presented in this paper serve to clarify this and other misconceptions related to Intelligent Video.

4.3 Future lookout

As a result of realignment of its programs, National Research Council of Canada has terminated the Video Recognition Systems (VRS) project under which ACE Surveillance was being developed. As a result, Canada Border Services Agency which considers Video Technology very important for its needs, is currently taking an initiative to establish a Video Technology section within its Technology and Innovation Branch that would capitalize on the expertise acquired within the VRS project. Specifically, the plans are to integrate ACE Surveillance system, as a multi-modal component with the radiation detection technology RADNET developed by the agency [6].

In addition, CBSA possesses an extensive video surveillance infrastructure – many CCTV installations have been done over the last several years in many Land and Air Points of Entries (POE). Many more will follow soon due to the increased security mandate all over the world. It is therefore critical for the efficient deployment and operation of these systems to have proper expertise-driven rather than market-driven performance evaluation standards within the agency. In this context, the methodology and the results presented by the current ACE Surveillance study are considered very important.

REFERENCES

[1]Transport Canada Rail & Urban Transit Security Workshop proceedings, Montreal, November 2007.

[2]Transport Canada CCTV Reference Manual for Security Application, Draft of August 31, 2007

[3]Federal departments meeting on Deploying Video Technologies for National Security, Organized by NRC/IIT Video Recognition Systems group in coordination with Disruptive Technology Office, Video Analysis and Content Extraction program, Ottawa, , June 2007: www.videorecogntiton.com/vt4ns (http://vrs.iit.nrc.ca/vt4ns).

[4]NRC/IIT ACE Surveillance (Annotated Critical Evidence) technology.
Website: www.videorecogntiton.com/ACE-surveillance (http://vrs.iit.nrc.ca/ACE-surveillance).
Poster: http://www.computer-vision.org/VideoRec07/pdf/ace_poster.pdf ,

[5]Gorodnichy, D., Mohammad, A. A., Dubrofsky, E., Woodbeck, K. "**Zoom on Evidence with the ACE Surveillance**," International Workshop on Video Processing and Recognition (VideoRec'07). Montréal, Québec, Canada. May 28-30, 2007. NRC 49349.

[6]Tony Mungham, Paul Arseneault, Bruce Rosenquist, Grant Gallant. Integration of Sensors for Automated Radiation Detection, in current proceedings of the NATO SET-125 Symposium on "Sensor and Technology for Defence against Terrorism", Germany, 2008