ACE Surveillance: the next generation surveillance for long-term monitoring and activity summarization

Dmitry O. Gorodnichy
Institute for Information Technology (IIT-ITI)
National Research Council of Canada (NRC-CNRC)
Montreal Rd, M-50, Ottawa, Canada K1A 0R6
http://synapse.vit.iit.nrc.ca/ACE-surveillance

Abstract

This paper introduces a new concept for the area of video surveillance called Critical Evidence Snapshot. We show that automatic extraction and annotation of Critical Evidence Snapshots, which are defined as video snapshots that provide a piece of information that is both useful and new, is the key to improving the utility and efficiency of video surveillance systems. An implementation of an ACE (Annotated Critical Evidence) Surveillance system made from off-the-shelf cameras and a desktop is described. The results obtained on several real-life surveillance assignments confirm our vision for ACE Surveillance as the next generation technology for collecting and managing surveillance data.

1 Motivation

When shopping for a video surveillance system (Google: "video surveillance"), of which there are many available on the market, one gets easily informed on "advanced" video surveillance features which various surveillance manufactures list to attract customers. These include "the highest picture quality and resolution" (5Mbps), "most advance digital video compression technologies", "complete control of Pan, Tilt and the powerful 44X Zoom", "total remoteness", "wireless internet connection", "greater detail and clarity", "multi-channel support of up-to 32 cameras", "extra fast capture" (of 240 frames per second), etc.

The problem arises when you try to use such surveillance systems, with any of the above features, especially for a long-term assignment. — You quickly realize that, whatever great quality or quantity of video data you capture, you may just not have enough space to record it nor enough time to browse it all in order to detect that only piece of information that is important to you. Simple motion detection (or more exactly, video-frame differencing-based change detection), which is a default technique used by many surveillance systems to initiate video capture, does not resolve the problem. More complex foreground detection techniques are also not sufficient for the purpose.

This leads us to definition of what is considered to be the main bottleneck of the current surveillance technology.

1.1 Two main video surveillance problems

Storage space consumption problem

The first problem deals with the excessive amount of video data which is saved in a digital form to be analyzed later. This is the way commercial DVRs (Digital Video Recorders) work. – They digitize dozens of hours of video from each camera on a hard-drive, which can then be viewed and analyzed by a human when needed. The need to browse through the recorded surveillance data usually arises post-factum – after a criminal act has been committed.

To get an idea of how much hard-drive space is usually consumed by a regular surveillance system, consider a typical monitoring assignment, such as in a banks or a grocery store, where 2–16 of cameras are used to monitor the premises. For this type of monitoring, 7 days of video recording is usually required; however, archival of up to 30 days of video is also common.

Even with a compression rate of 10 Mb per 1 minute, it takes $10Mb * 24h * 60min/10 = 1.5$ GB per day per camera or about 20 – 700 Gb of hard-drive space for a

single assignment[1]. The factor of ten is generously chosen to approximate the motion-detection-based video capturing.

It is clearly seen that, while this amount of space is feasible to have by using dedicated DVR systems, it is not within a normal hard space used by an ordinary computer.

Data management problem. London bombing video backtracking experience

The problem however is not just the lack of hard-drive space, but not having enough time to go though all recorded data searching for what you need. Having too much stored data is just as bad as not having any data at all. — If the amount of data is so large that it cannot be managed within reasonable amount of time and efforts, it is useless.

After the London bombing, millions of hours of digitized video data from thousands of cameras were browsed by the Scotland Yard officers searching for the data which could lead to the identification of the bombers and their accomplices. Because of the extremely large number of search branching occurring every time a person disappears from a camera field of view, manual browsing of the recorded video data backtracking the suspect and everyone who resembled him was extremely time consuming and was not feasible to be performed within the desired time frame.

This example shows the difficulty one faces if no solution is provided for automating the information extraction from video. Therefore, *it is critical for a video surveillance system to store only that video data which is useful, i.e. the data containing new evidence.*

This is further illustrated with another example, which also happened in real life and which is used throughout the paper as a practical guide and benchmark for the surveillance systems we want to design.

2 "Who stole my bike" example

The proper name for the surveillance task example described below should be "What is happening while I am away", because it is not just about the very moment when a bike was stolen, but about the entire log of activities preceding the accident. Here is a description of the task.

[1]10 Mb per 1 minute is convenient rule-of-thump compression rate that can be used to estimate space required to store a video. More exactly, one of the best compression rates for digital video is achieved by Divx v5 codec — 9Mb per 1 minute of 640 x 464 digital video. Analog video requires about a quater of size needed for digital video, due to th fact that its resolution never acceeds 320 x 240.

I used to keep my bike in a hidden slot between my car and the wall inside my car-port. My house is located in a cul-de-sac and the bike left in the car-port is not seen, unless you come close to the house and sneak around, passing near the house windows. This is why, I never locked my bike there, assuming that it was safe enough. One morning the bike was gone.

Analyzing how this could happened, it was clear that someone was visiting my house prior to the incident – firstly, to find that there was a good unlocked bike there and secondly, to find out the best time when nobody is around. But when did this visit happen and who was that? This was left unclear. — I did not have any evidence which could have helped the police to identify the thief. This seemed ironic to me, since I do have a plenty of cameras in my house which could have been used to monitor the premises, and even more, I could have easily connected them to my desktop computer which has a variety of video capturing tools.

The problem was that, without anticipating a theft, would you record days and possibly weeks of video data with only a slight chance that it might be used at some point in future? How much space on your computer and how much time would you spend daily to maintain such system, provided that is meant to run in a *continuous* day-by-day basis?

Instead of answering these rhetoric questions and to make sure that next time I have a better idea on who and when is visiting my house while I am away, I have decided to use this scenario to develop and test a new type of video-recognition-based surveillance system that would automatically generate a summary of all activities visible by a camera over a period of time. With such a system, I should be able to let the system run for as many hours as I am away – usually for about 9-12 hours, and then simply press the play button upon my return and view a short video report produced by the system. I know that the report should be short – not more than five-ten minutes, because, as mentioned earlier, there is normally no much activity happening around my house.

Criteria for the system

The performance criteria required from the surveillance system are listed below.

1. It should be affordable, easily installed and operated – i.e. run on a desktop computer with off-the-shelf cameras: USB web-cameras, CCTV cameras or home-video hand-held cameras (analog or digital), either directly connected to a computer or connected via a wireless transmitter for viewing remote areas.

2. It should run real-time in a continuous mode (24/7, non-stop, everyday).

3. It should collect as much useful video evidence as possible. What type of evidence and how much of it to be collected is determined by the quality video data and the setup. For example, in a condition with bright illumination and close-range viewing (i.e. high resolution), the system can collect the information about the visitors faces, whereas in a monitoring assignment performed at night or at a distance, such information as number of passing cars and people would be collected.

4. It should be merciful with respect to the hard-drive space.

5. It should be as much automated as possible: both in recognizing the pieces of evidence in live video and retrieving them from saved data off-line.

6. The collected video evidence, besides being useful, has also to be easily manageable, i.e. it should be succinct and non-redundant.

The currently available surveillance technology does not meet these criteria. In this paper we present a new type of the video surveillance, called ACE Surveillance, that does.

ACE Surveillance can be considered as another example of automated video-recognition-based surveillance, also known as "Smart Video Surveillance" , "Intelligent Video Management", "Intelligent Visual Surveillance" ([12, 4, 3]). However, it not only attempts to identify in real-time the objects and actions in video, but also addresses the problems of the surveillance data compression and retrieval. These problems are tackled by introducing the concept of Critical Evidence Snapshot, which the ACE Surveillance system attempts to automatically extract from video.

ACE Surveillance is most useful in security domains where premises are monitored over a long period of time and are rarely attended. The affordability of ACE Surveillance, both in terms of the video-equipment cost and the computer requirements, also makes it very suitable for home monitoring with ordinary off-the-shelf cameras and desktop computers.

3 New concepts: Critical Evidence Snapshot and ACE Surveillance

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.*

Normally, the evidence is an event that is captured by a video camera. What is important is that the number of events that moving objects (such as people, cars, pets) perform is limited. These events are listed below:

1. Appear.

2. Move: left, right, closer, further.

3. Stay on the same location (with a certain range).

4. Disappear.

Another important observation that ACE Surveillance is based upon is that between the times when an object Appears and Disappears, there is no need to store all video frames from the surveillance video. Instead, what needs to be stored are the first snapshot – when the object appeared, and the last snapshot – before it disappeared, both of which must be time-stamped to provide a time interval within which the object was visible. In addition to these two snapshots, a few other (but not many!) snapshots of the object may also be stored. These can be stored intentionally – if a snapshot is qualified by the system as the one that either shows a better quality picture (or view) of the object or captures a new event performed by the object. They may also be stored unintentionally – due to the shortcomings of a video recognition algorithm, when a snapshot is erroneously qualified as the one described above. However, even if such snapshots are stored unintentionally, they still are useful as long as there are not too many of them, since they guide a viewer's attention to the detected changes and the better quality snapshots. *All of the extracted snapshots will make a pool of the extracted CES-es, of which some will be more useful, while some less. When combined together along with extra annotation, which describes them, they create a very useful and succinct representation of what has been observed by a camera.*

This is a key idea behind the ACE Surveillance. Regardless of how detailed the annotation of extracted snapshots is, the very fact that instead of storing 16 fps video it stores only selected number of frames, makes the output much more manageable and useful.

3.1 Annotations

The annotations that can be provided with each extracted CES depend on the level of difficulty of a monitoring assignment and on the complexity of the object

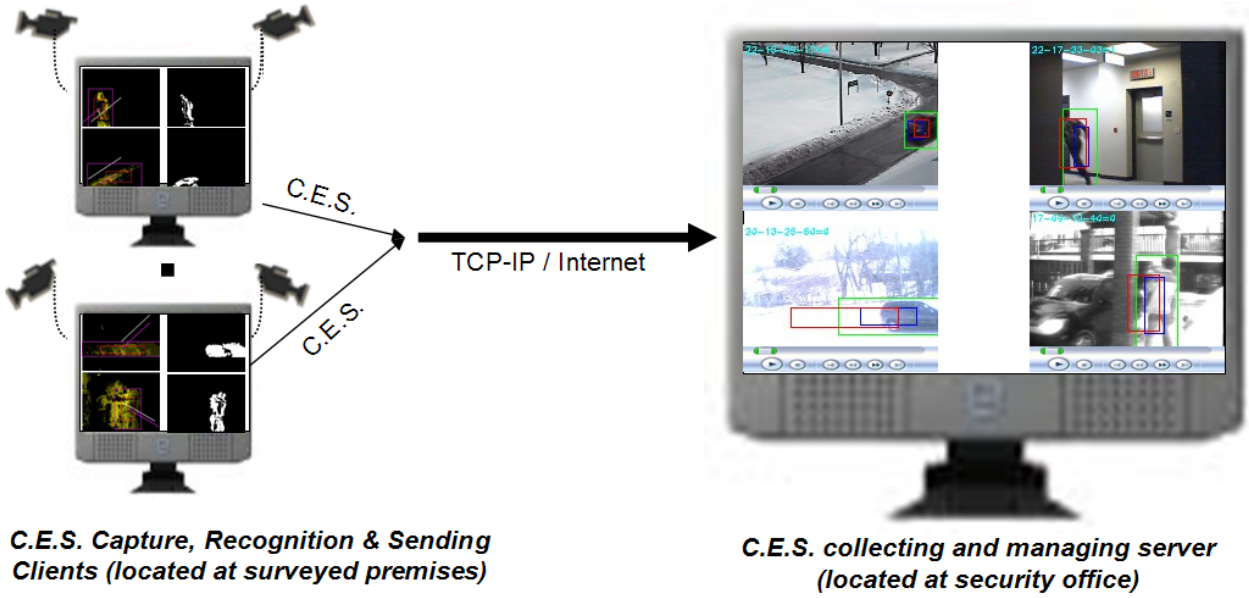

Figure 1. The ACE Surveillance client-server architecture. CES-es extracted by the clients are transmitted to the server along with their annotations.

tracking and recognition algorithms used for the assignment. In the currently used ACE Surveillance system (see Section 5), we use time-stamps, optional object motion activity labels (such as Left, Right, Closer, Further), optional object identification number (if an object is recognized as the already seen one) and object counts (such as number of pedestrians and cars passing by the window) to create a text-based annotation of the CES.

In many cases, such text-based CES annotation is already sufficient to know what was happening during the monitored time interval (in particular for cases when nothing unusual happened) without a need to browse the actual snapshots. This is illustrated in Figure 2.a. – When the same activity is observed every day, then approximately the same text-based log made of CES annotations is expected.

In addition to text-based annotations, graphical annotation of CES are displayed to facilitate their viewing. For example, the boxes circumscribing the moving objects and their trajectories, shown in Figures **??** - **??**, help the viewer to focus attention on the detected objects and their motion trajectories and/or boundaries.

3.2 Specific interest: faces in video

ACE Surveillance has a very good application for person identification from video. For forensic purposes, what is needed from a surveillance system is not a video or extensive number of snapshots somehow capturing a person, but a few snapshots that show a person in a best possible for the current setup view and resolution.

From our previous work on face recognition in video [7], we know that a face becomes recognizable, both by a human and a computer – for such tasks as classification and verification, when the image of his/her face has a resolution of at least 12 pixels between the eyes.

Therefore the key idea for the ACE Surveillance is to keep tracking an object (while taking CES-es and their annotations) until it is close enough to the camera for a face detection and memorization techniques to be employed. It can be seen however – by examining the quality of the video images from our surveillance tasks (Figures 2-6), that, while this idea definitely shows the right way to proceed in person identification from video, it does require a well lit environment or better quality video cameras so that high-quality video-snapshots can be taken when needed. One way of resolving the poor quality issue, is (as suggested in [8]) to use a pan-tilt digital photo-camera, the orientation of which is driven by the tracking results from the video camera and the trigger button of which is initiated by a detection of an appropriate CES from the video.

a)

Name	↑Ext	Size	Date
[..]		<DIR>	02/13/2006 10:06
_10-22-36-48_0	jpg	31,103	02/10/2006 22:36
_10-22-46-01_0_o1_OFF	jpg	25,741	02/10/2006 22:46
_10-23-30-09_0_o2_ON	jpg	28,095	02/10/2006 23:30
_10-23-30-11_0_o2_L	jpg	26,934	02/10/2006 23:30
_10-23-49-57_0_o3-o2_OFF-o4	jpg	24,451	02/10/2006 23:49
_11-02-43-31_0_o1_ON	jpg	27,228	02/11/2006 02:43
_11-02-43-32_0_o1_C	jpg	30,339	02/11/2006 02:43
_11-02-53-18_0_o1_OFF	jpg	24,315	02/11/2006 02:53
_11-06-36-58_0_o5_ON	jpg	31,555	02/11/2006 06:36
_11-06-36-59_0_o5_C	jpg	32,417	02/11/2006 06:36
_11-06-37-00_0_o5_C	jpg	32,757	02/11/2006 06:37
_11-06-37-01_0_o5_C	jpg	32,994	02/11/2006 06:37
_11-06-37-02_0_o5_CL	jpg	31,829	02/11/2006 06:37
_11-06-37-03_0_o5_L	jpg	31,936	02/11/2006 06:37
_11-06-37-04_0_o5_OFF	jpg	28,141	02/11/2006 06:37
_11-06-37-05_0_o5_ON	jpg	29,703	02/11/2006 06:37
_11-06-37-06_0_o5_R	jpg	32,142	02/11/2006 06:37
_11-06-37-07_0_o5_R	jpg	32,448	02/11/2006 06:37
_11-06-37-08_0_o5_FR	jpg	32,200	02/11/2006 06:37
_11-06-37-09_0_o5_F	jpg	32,090	02/11/2006 06:37
_11-06-37-10_0_o5_R	jpg	31,474	02/11/2006 06:37
_11-06-37-12_0_o5_R	jpg	31,412	02/11/2006 06:37
_11-07-03-18_0_o6_ON	jpg	56,228	02/11/2006 07:03

b)

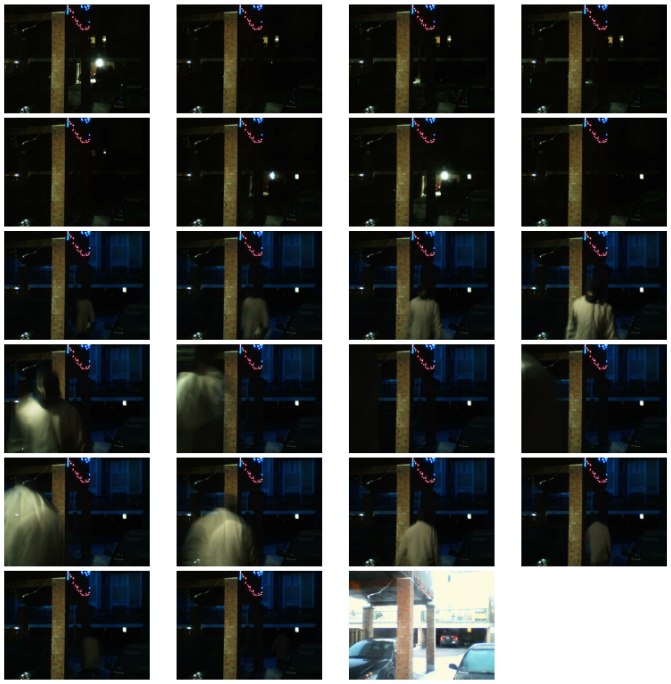

Figure 2. The entire 9-hour activity observed over night is summarized in 23 annotated CES-es: a) the text-based CES annotations, encoded in file names, and b) the thumbnail view of the extracted CES-es. In the file name _dd-hh-mm-ss_c_oX(_annot)-oY(_annot).jpg, dd-hh-mm-ss stand for date-hour-minute-second when CES was taken, C – camera number, oX – object #X is detected (If systems recognizes an object that it has already seen, it assigns # X corresponding to that object. Otherwise, a new # X is assigned). (_annot) – object annotation which is one of the following: L, R, C, F (object moves Left, Right, Closer, or Further away), On / Off (illumination change caused by object, which usually indicates either appearing/disappering of the object or switching on/off the lights).

4 ACE Surveillance architecture

The architecture of the ACE Surveillance system is driven by the goals of extracting and annotating the Critical Evidence Snapshots as presented below (See also Figure 1).

It consists of CES client module, of which there can be several – each connected to one or more video-cameras, and CES managing module, which can operate on an another server-based computer, normally located in a security monitoring office.

In the client-based CES registration module, video data are acquired and CES-es are extracted on-line from a video stream, annotated and prepared for transmission. In the server-based CES managing module, off-line post-processing, archival and retrieval of CES-es happens.

4.1 CES client architecture

CES clients capture video from one or more video sources, performs on-line video recognition of captured video data and then send video-frames along with the acquired CES and their annotation to the CES server.

For each video frame of each video source, a CES client performs the following six tasks in real-time (as it captures the video):

- Task C1: Detection of new or moving object(s) in video.
- Task C2: Computation of the attributes of the detected object(s)
- Task C3: Recognition of object(s) as either new or already seen, based on its attributes.
- Task C4: Classifying a frame as either being a CES (i.e providing new information) or not.

- Task C5: Extracting and creating CES annotations, such as: timestamps, augmentations, outlines, counters, contours.
- Task C6: a) If a video frame is CES, then it is sent to the CES server along with the annotations; b) It not, then resolution-reduced version of the frame is sent to the CES server.

There is an abundant set of literature addressing the computer vision techniques that can be employed for each of the above described tasks (see the current workshop [10]). The most important of results of these, in our opinion, of which we use in our work, are the described below (see also our discussion in [11]).

For task C1: A combination of change detection (based on last several frames only) [2], foreground detection (based on adaptive maintenance of the background model [16]), and motion pattern estimation (based on second-order change detection [6] or tracking the changed objects over several consecutive frames [13, 14]) allows one to achieve the best detection of only those objects that either move or appear/disappear and ignore all other changes observable in video, such as caused by wind blowing the tree branches or due to the wireless transmission. The presence of colour information is not critical for this task, which our experiments with black-n-white CCTV cameras confirm (see Section 5). Although, should colour be available, it should not be disregarded as it does improve change detection [1].

The techniques for isolating shadows from the object can be considered, but our experiments indicate that they are not critical. The reason is that the shadows follow the objects, and it is the object motion that determined whether a frame is a CES or not.

For tasks C2 and C3: the attributes which allow us to track and identify an object as the already seen are:

- location and size, measured by the size of the detected motion blob (x,z,w,h)
- their derivatives (changes) in time: d(x,z,w,h)/dt
- colour: Pcolour(i,j)
- texture/edge: Ptexture(i,j),

Surveillance video usually shows objects in poor resolution and quality. Because of that feature-based tracking or recognition, such as the one based on selecting pixel-based object features as in Lucas-Kanade or SIFT-based approaches, may not work. We argue that the best approaches for surveillance video are those based on accumulation of data over time such as (listed in order of their discriminating power), histogram-based, correlogram-based [17], and associative memory based [9]. Histograms count the number of occurrences of different object values. Correlograms take also into account their geometrical location with respect to each other. Associative memories also provides a more intelligent update of the object representation based both on the currently obtained piece of data and the past data.

Using these approaches, the recognition of object(s) as either new or already seen is based on the computed object attributes and thresholds on allowed variations of objects size and location.

For colour-based recognition, recognition-driven (rather than sensor-driven) colour space should be used, such as UCS, HSC, or non-linearly transformed YCrCb, possibly with reduced range of values to count for the low quality of data. For texture/edge-based recognition, Local Binary Patterns [15] and Binary Consensus Transform [5] appear most promising.

For task C5: It has to be emphasized that the prime goal of the CES annotations is to provide a suggestion of what is observed, rather than to actually identify the activity or the object. This is one of the main lessons from the computer-vision: with so much variation in cameras quality, lighting conditions and motion types, it is very difficult for a computer to completely replace a human in an arbitrary recognition task.Therefore the prime task of the computerized surveillance is in helping a human to perform such recognition, rather than doing it instead of a human This is the idea behind CES annotations; they serve as a guide to facilitate the retrieval of the required data in a mass of archived data sets.

Out implementation of an ACE Surveillance system allows simultaneous monitoring of several locations using the same program on a single computer. A number of USB video-cameras (or video capture devices) can be connected to a computer. On a 2.4GHz Pentium IV computer, it takes about 60 msec to process a video frame from one camera, which allows processing up to four cameras in a close to real-time speed. Figures 2-4 show the moving object detection results, as well as CES and their annotations obtained from several real-life surveillance assignments.

4.2 CES server architecture

The CES server receives video-frames and CES-es from all CES clients (using either a TCP-IP protocol or secure ftp) and prepares them for viewing on a security desk monitor using a web-scripting code.

At any point in time, a security officer has the option of switching between the following two (or three)

viewing options.

- Task S1: Viewing live video from all cameras, which is shown as a flow of resolution-reduced video frames received from an CES client. Unlike CES frames, which are stored on a CES server hard-drive, regular (non-CES) frames are not stored and are deleted after being displayed.

- Task S2: Viewing Automatically obtained Critical Evidence Summarization (ACES) video for a particular camera or a time interval. ACES video is made of annotated resolution-reduced CES-es played in a sequence (as a movie) with possible augmentation (such as outlines/boxes around the detected objects) to guide the visual attention of the viewer. As ACES video is played, an officer has an option of seeing the actual (unmodified high-resolution) version of every CES. In addition, for each video-camera, the last acquired time-stamped CES as well as the a text-encoded log of all CES annotations plotted on a time-line are also made visible to the officer so that s/he always has a clear picture on what is and was happening in each camera field of view.

- Task S3: Viewing CES-es, by associative similarity. This is another advantage offered by the ACE Surveillance — it allows the viewing of the stored CES-es not only in the order they were captured (as done when viewing the summarization video), but also in the order of their similarity to each other or an query image. Histogram-based and associative-memory-based approaches are very suitable for such context-based snapshot ranking.

5 ACE Surveillance in action

Below we present typical results obtained using the ACE Surveillance on several real-life monitoring assignments. The focus of this presentation is on the low quality of video data one has to deal with while monitoring several premises using surveillance cameras and the utility of the CES-based video summarization obtained by the ACE Surveillance on these data.

Experiments were on conducted on two monitoring stations, called here "Office" station and "Home" station. For each monitoring assignment, the premises were surveyed over a long period of time (10 hours at least) Thus lighting conditions were changing from dusk to dark to sunlight and back to dark. Both indoor (through the door) and outdoor (though the window) viewpoints were tested.

In order to test and to tune the system, these monitoring assignments use several off-the-shelf video-camera and capture boards: including digital low-cost and high-cost CMOS web-cams, black-n-white wireless CCTV surveillance camera with and analog output and 8mm hand-held cameras with zoom and high-quality lenses connected directly and wireless (via a transmiter) to a computer. In case of a wireless video transmission, the receiver was located in the basement of the house, while the transmiter was located on the main floor of the house. This created extra noise due to wireless transmission, which, as our experiments show, does not change much the performance of the system.

The "Home" station had two cameras connected, one viewing the entrance of the house with the carport – at a close range, the other camera viewing the street on the back of the house – at a distance. The "Office" station had also two cameras, one viewing the corridor in front of the office, which has an exit stairwell door, the other one viewing the street from the third floor on which the office is located.

The samples of the annotated CES-es extracted from several runs on these assignments are shown in Figures 2-4. The entire set of CES-based summarization obtained by ACE Surveillance on these assignments, which includes the original CES extracted from video, their annotated thumbnail versions, and ACES (CES-based summarization) video, are made available at the project web-site. Besides, an example of a live ACE assignment can also be accessed through a password-protected CES-server site: http://synapse.vit.iit.nrc.ca/acese-data/home-19_2200-20_1430.php (password: "vp4s-06"). Below is the summary of the results.

Figure 2 shows both the text annotation of the observed activities and the thumbnail CES-based summarization for an over-a-night monitoring (Hours of operation: 10 hours (22:00 - 8:00). Camera used: USB low-cost CMOS web-cam. Number of extracted CES: 23. ACES summarization video: 74Kb). The CES annotations automatically provided by the system are encoded in the file names.

Figures 3a and 4a show the results obtained on the same premises as in Figure 2, but obtained on different days with a different cameras: (20 hours (22:10 - 18:09) of monitoring with a wireless black-n-white CCTV cam + Video2USB converter. Number of extracted CES: 177 and 222. ACES video: 900Kb).

Figures 3b and 4c show the results obtained by a camera that overviews the road on the back of the house (20 hours of monitoring with a Creative USB web-cam, ACES video: 4Mb).

Figures 3 and 4 also show the results obtained from

the "Office" station: i)over a night (monitoring the corridor: (19 hours (16:45 - 10:00) of monitoring with a USB web-cam. Number of extracted CES: 148, ACES video: 600Kb), ii) and over a weekend monitoring the street (from Friday night to Monday morning).

Figures 3-4 show the annotated resolution-reduced CES-es extracted from video, where the annotation includes a timestamp (shown at top right of the video image) and highlighted boxes outlining the detected objects, the direction of their motion (blue box - the changing moving blob, red box - foreground history, green box - enclosing around the object).

For a comparison with a typical motion-trigged surveillance, which is also often called "automated" by the manufactures, we conducted a similar monitoring assignment in which a camera surveyed a street (as in Figure 4.d) over 2 hours with a commercially available automated surveillance program from Honest Technology called VideoPatrol. The number of video frames which was captured by the program (regardless of what motion threshold was choses) significantly exceeded the number (by the factor of up to 100) of actual number of frames that contained any useful information, such a passing car or a pedestrian. Most frames were captured simply because there was some inter-frame difference detected by the system, such as due to wind or non-perfect camera quality.

6 Conclusion

The paper defines a new type of automated surveillance, based on extracting and annotating pieces of Critical Evidence in live video-streams, termed the ACE Surveillance, which is shown to facilitate significantly the manageability of the surveillance video data, tackling both the data storage and the data retrieval problems.

The work presented in this paper is in its early stages. However even the results obtained so far on several real-life monitoring tests support the validity of our vision of the ACE Surveillance as the next generation surveillance. In this sense, the name "ACE Surveillance", which was originally coined as an acronym for Annotated Critical Evidence, can also signify Automated surveillanCE and imply an ace-like potential behind it.

Acknowledgement

The author acknowledges the help of Kris Woodbeck in the PHP coding of the CES-server for the described in the papers client-server architecture of the ACE surveillance system.

References

[1] J. Black and T. Ellis. Intelligent image surveillance and monitoring. In *Measurement and Control, 35(8): 204-208,*, 2002.

[2] E. Durucan and T. Ebrahimi. Change detection and background extraction by linear algebra. In *IEEE Proc. on Video Communications and Processing for Third Generation Surveillance Systems, 89(10)*, pages 1368–1381, 2001.

[3] M. V. Espina and S. Velastin. Intelligent distributed surveillance systems: A review. In *IEE Proceedings - Vision, Image and Signal Processing', 152(2) IEE, April 2005, pp. 192-204*, 2005.

[4] J. P. Freeman. Jp freeman ranks 3vr at top of intelligent video market. 2005.

[5] B. Froba and C. Kublbeck. Face tracking by means of continuous detection. In *Proc. of CVPR Workshop on Face Processing in Video (FPIV'04), Washington DC, 2004.*

[6] D. Gorodnichy. Second order change detection, and its application to blink-controlled perceptual interfaces. In *Proc. IASTED Conf. on Visualization, Imaging and Image Processing (VIIP 2003), pp. 140-145, Benalmadena, Spain, Sept.8-10*, 2003.

[7] D. Gorodnichy. Video-based framework for recognizing people in video. In *Second Workshop on Face Processing in Video (FPiV'05). Proceedings of Second Canadian Conference on Computer and Robot Vision (CRV'05), pp. 330-338, Victoria, BC, Canada, ISBN 0-7695-2319-6. NRC 48216*, 2005.

[8] D. Gorodnichy. Seeing faces in video by computers. Editorial for special issue on face processing in video sequences (nrc 48295). *Image and Video Computing*, 24(6):551–556, 2006.

[9] D. O. Gorodnichy. Associative neural networks as means for low-resolution video-based recognition. In *International Joint Conference on Neural Networks (IJCNN'05), Montreal, Quebec, Canada, NRC 48217*, 2005.

[10] D. O. Gorodnichy and L. Yin. Introduction to the first international workshop on video processing for security (vp4s-06). In *Proceedings of the Canadian conference Computer and Robot Vision (CRV'06), June 7-9, Quebec City, Canada. NRC 48492.*, 2006.

[11] D. O. Gorodnichy and A. Yogeswaran. Detection and tracking of pianist hands and fingers. In *In Proc. of the Canadian conference Computer and Robot Vision (CRV'06), Quebec, Canada, June 7-9. NRC 48492.*, 2006.

[12] A. Hampapur, L. M. Brown, J. Connell, M. Lu, H. Merkl, S. Pankanti, A. W. Senior, C. fe Shu, and Y. li Tian. Multi-scale tracking for smart video surveillance. In *IEEE Transactions on Signal Processing, Vol. 22, No. 2*, 2005.

[13] Y. li Tian and A. Hampapur. Robust salient motion detection with complex background for real-time

video surveillance. In *IEEE Computer Society Workshop on Motion and Video Computing, Breckenridge, Colorado*, 2005.

[14] D. Magee. Tracking multiple vehicles using foreground, background and motion models. In *In ECCV Workshop on Statistical Methods in Video Processing, pages 7-12*, 2001.

[15] T. Ojala, M. Pietaikinen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(7):971–987, 2002.

[16] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers. Wallflower: Principles and practice of background maintenance. In *International Conference on Computer Vision, Greece.*, 1999.

[17] Q. Zhao and H. Tao. Object tracking using color correlogram. In *2nd Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, Page(s): 263 - 270, 15-16 Oct*, 2005.

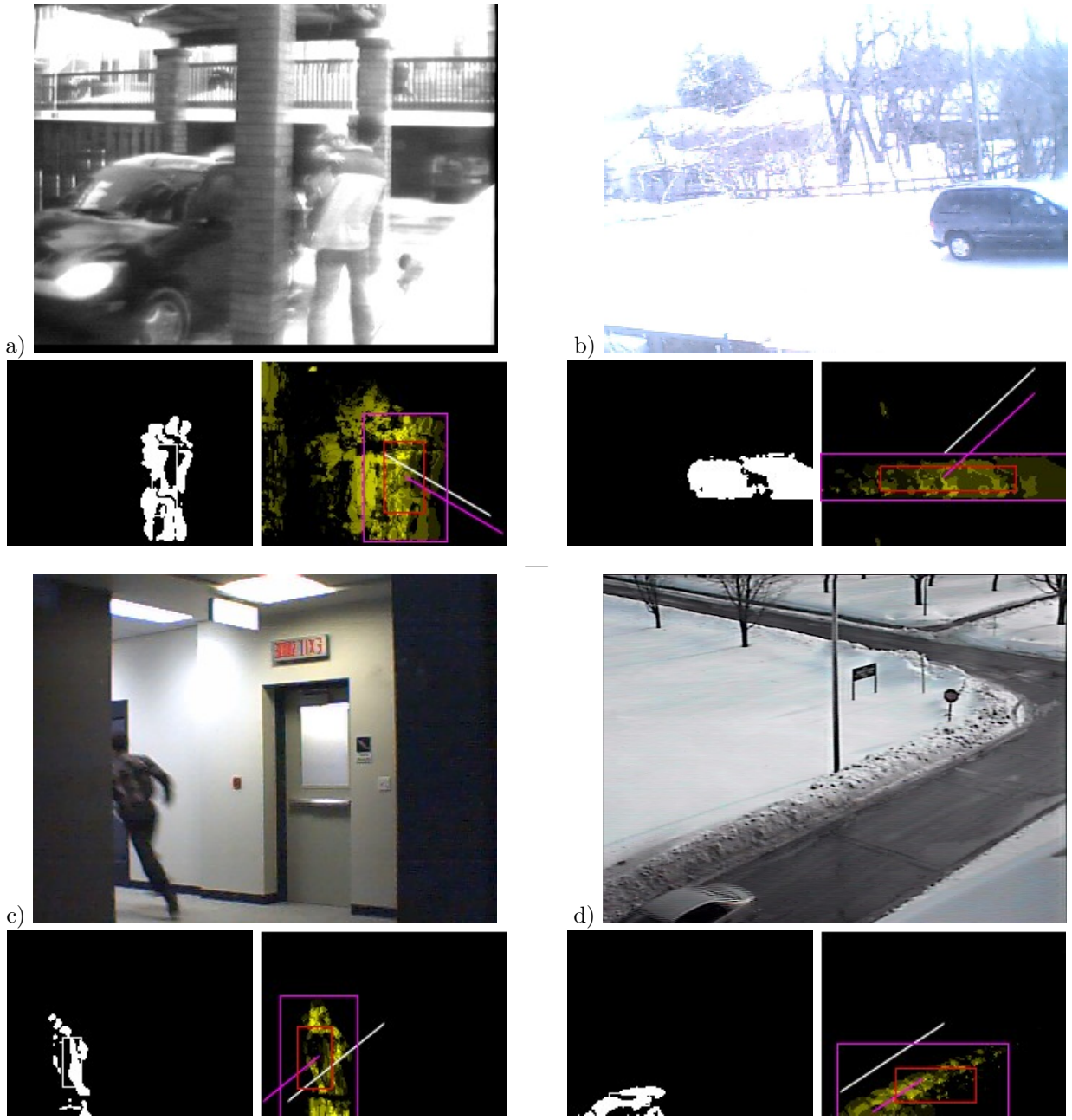

Figure 3. Examples of a 640x480 CES-es obtained from "Home" (a,b) and "Office" (c,d) surveillance stations: a) overnight monitoring of the car-port in front of the house (by wireless CCTV camera), and b) over-day monitoring of a street on the back of the house (by a USB webcam). c) overnight monitoring of the corner corridor and exit door in front of the office (by a USB2 webcam), and d) over-week-end monitoring of a street seen from the office window (by Sharp 8mm hand-held camera). The moving blob image and the foreground motion history images are shown on the bottom of each CES. The detected objects and their the direction/range of the motion are shown. Low quality of video images can be seen.

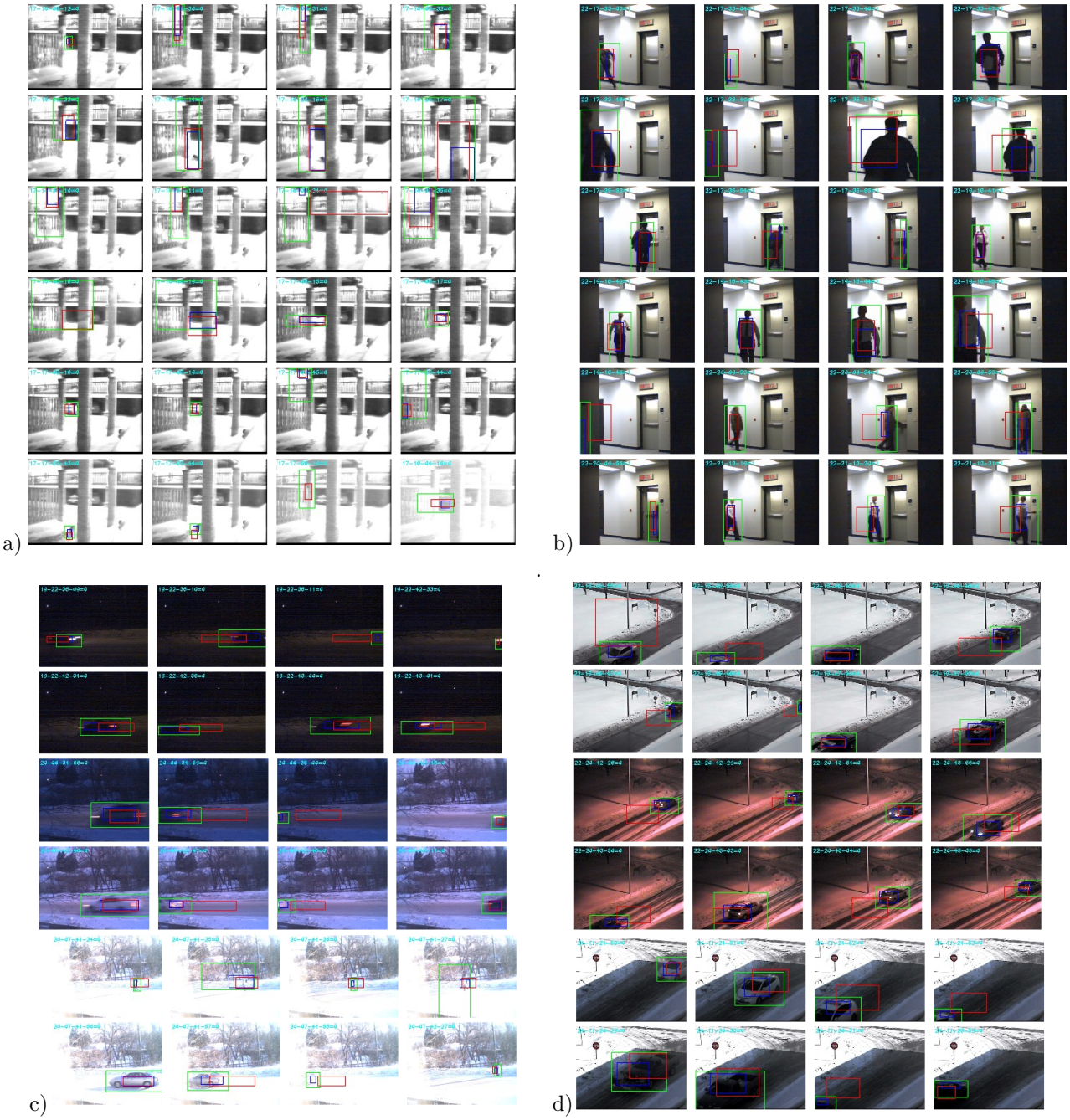

Figure 4. Examples of Annotated CES-es from several monitoring assignments: a) overnight monitoring of the carport in front of a house (with a black-and-white wireless CCTV camera on "Home" station), b) overnight monitoring of the corner corridor and exit door in front of the office (by a USB2 webcam on "Office" station), c) over-day monitoring of a street on the back of a house (with a USB webcam on "Home" station), d) over-week-end monitoring of a street from the office window (by Sharp 8mm hand-held camera on "Office" station). Timestamp is shown at top right of CES. Highlighted boxes indicate: blue box - the changing moving blob, red box - foreground history, green box - enclosing around the object.