NRC·CNRC
From Discovery to Innovation...

A.C.E. Surveillance

(next generation surveillance for long-term monitoring and activity summarization)

Dmitry O. Gorodnichy
Computational Video Group
Institute for Information Technology
National Research Council Canada
http://synapse.vit.iit.nrc.ca/ACE-surveillance

The First International Workshop on *Video Processing for Security*
June 7-9, Quebec City, Canada

National Research Council Canada | Conseil national de recherches Canada | Canada

Two Big problems: on user side

1. Storage space consumption
   - Typical assignment:
     2--16 cameras, 7 or 30 days of recording, 2-10 Mb per min.
     ➔1.5 GB per day per camera / 20 -- 700 GB total !

2. Data management and retrieval
   - London bombing video backtracking experience:
     "Manual browsing of millions of hours of digitized video from thousands of cameras (by the Scotland Yard officers trying to back-track the suspect and everybody who resembled him) proved impossible within time-sensed period"

Conclusion:

"highest picture quality and resolution", "complete Pan/Tilt control", "powerful 44X Zoom", ``total remoteness", "multi-channel support of up-to 32 cameras", ``extra fast capture of 240 fps" is all good, **but** you may just not have enough time to browse it all in order to detect that only piece of information that is important to you.

Two Big problems: on research side

1. Low video quality:
   - low resolution, blurring, out-of-focus, interlacing, due to wireless - just see the snapshots from real surveillance monitoring!

2. Real-time constraint

Real-time: < 80 msec per frame (>12 fps) for Short-term & Short Range:
- objects close to camera (or captured at close zoom) → smaller image required → can be done faster !

Quasi real-time: < 500msec (2 fps) for Long-term & Long-range:
- objects away from camera (or wide zoom) → larger image required → can be done slower !

Besides vision recognition: is very ill-posed, and has high complexity and diversity of monitoring scenarios and data recognition and retrieval tasks that may have to be executed and integrated

Example and test-bed

"Who was visiting my house while I was away?"

Long-term monitoring of rarely visited premises with undedicated computers and off-the-shelf video-cameras

- 2 GHz computer with 60 GB hard-drive
- Time - priceless
- Webcams / Analog (8mm) / CCTV cameras (wireless and connected)
- *An unlocked bike to be stolen one day*

Monitoring the front of the house

Monitoring the back of the house

Required Criteria

"For video surveillance to be operational, it is critical to store only that video data which is useful, i.e. the data containing new evidence".

1. As much useful video evidence should be collected as possible.
2. However, the collected video evidence, besides being useful, has also to be easily manageable, i.e. it should be succinct and non-redundant.

What type of evidence and how much of it to be collected is determined by the quality video data and the setup.

1. in bright illumination and at close-range: information about visitors faces can be collected (when i.o.d. > 12 pixels)
2. at night or at a distance: the time and number of passing objects.

New concept: C.E.S.

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

Limited number of evidence types for moving objects:

- 1. Appear.
- 2. Move: left, right, closer, further.
- 3. Stay on the same location (with a certain range).
- 4. Disappear.

Between "Appear" and "Disappear" only a few shots are needed!

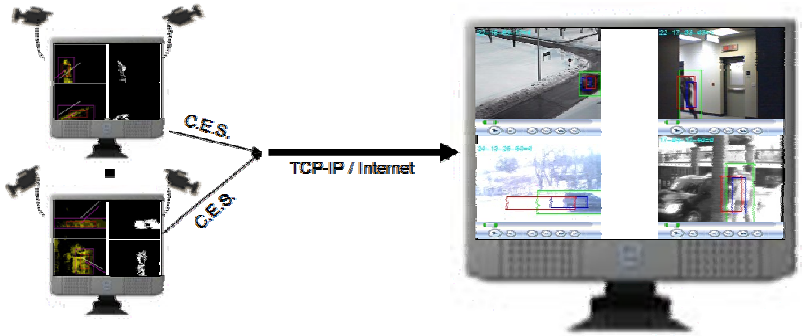
CES & their annotations

1. Saving CESes instead of .avi file, makes archived video data much easier to manage (inc. data retrieval and search)!
   - Even is some extracted CESes are not very useful.
2. Adding Annotations to CES makes them even more so!
   1. Text-based - for log reports
   2. Graphic image augmentations - as attention guide to viewer

   NB: Annotations are not answers, but suggestions and guides to a viewer

- Specific interest: faces* (*- some of this still to be done)
  - Track until person-looking blob is at least 20 pixels wide (IOD>12), after that discard or blobs/faces or worse resolution/quality
  - Video is very suitable for face classification task (e.g. similarity ranking of faces) [Gorodnichy'05]
  - ACE Surveillance makes it possible to browse and retrieve stored video-data, compressed as a collections of CES-es, by associative similarity.

ACE Surveillance architecture



C.E.S. Capture, Recognition & Sending Clients (located at surveyed premises)

C.E.S. collecting and managing server (located at security office)

- CES-Client(s):
  - real-time video processing and object/motion detection & recognition
  - extracting & annotating CESes
- CES-Server:
  - receives CESes from CES-Client(s),
  - tool for browsing CESes: by camera, by time, by context (associatively)

CES Client tasks

C1: Detect new or moving object(s) in video.

C2: Compute of the attributes of the detected object(s):

- location and size, measured by the size of the detected motion blob (x,z,w,h) and their derivatives (changes) in time: d(x,z,w,h)/dt
- colour (Histogrammed): Pcolour(i,j) * (2D H + Cb)
- texture/edge (Histogrammed): Ptexture(i,j)* (with Local Binary Patterns and Consensus Transform) *

* - replace histograms/correlograms with associative memory [Gorodnichy'06]

C3: Based on attributes, recognize object(s) as either new or already seen

C4: Classify frame as CES (i.e providing new information) or not.

C5: Create CES annotations: timestamps, outlines, counters, contours.

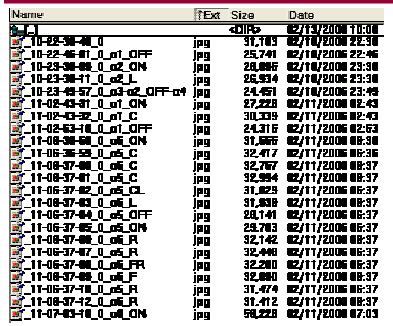
C6: If a video frame is CES, send it to CES server along with its annotations

Techniques:

For C1: Combination of change detection, background maintainance and dominant motion estimation [Espina'05, Tian'05, Magee'01, Toyama'99]

For C2 & C3: accumulation-over-time approaches. Proper colour space.

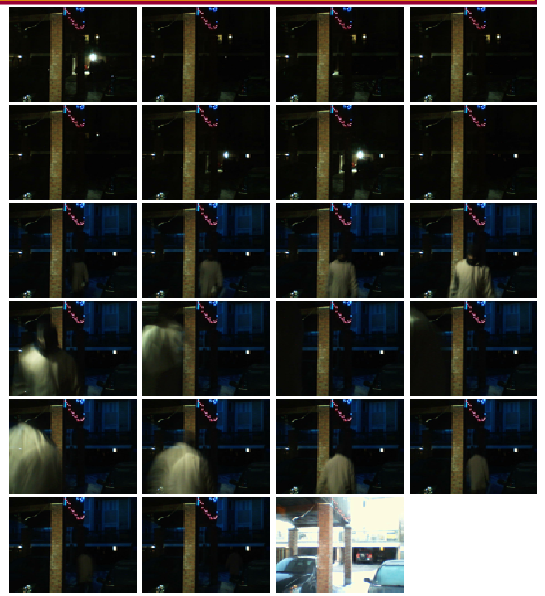
ACE Summarization (ACES)



a) By log of text-annotations (already sufficient to see that nothing unusual happened)

b) By CES thumbnails (8Kb each)

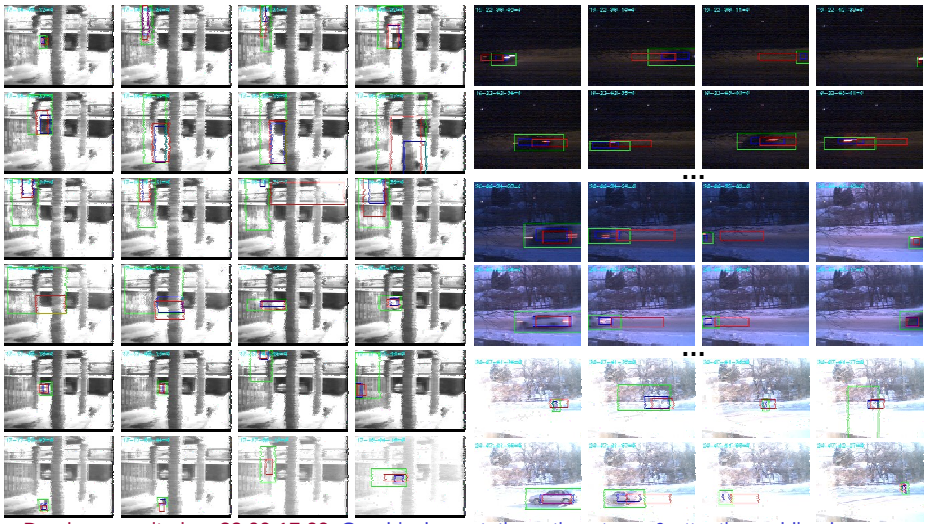
c) By ACES video



Overnight monitoring: 22:00-8:00

Cam: USB webcam. # CESes=23. ACES video = 360 Kb

CES extraction results: "Home" station



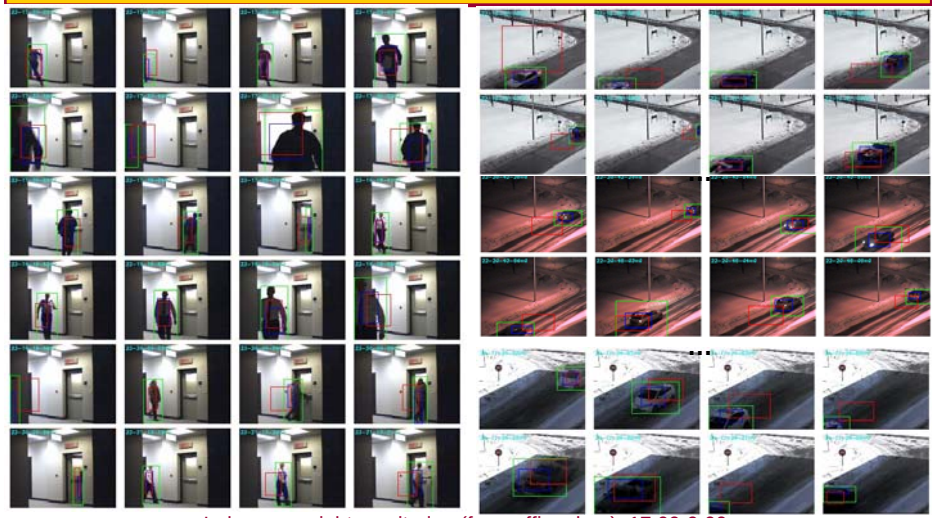
Day-long monitoring: 22:00-17:00. Graphical annotations: timestamp & attention guiding boxes

Cam1 (monitoring front): wireless black-n-white CCTV camera. # CES=177. ACES video = 360 Kb

Cam2 (monitoring back of the house): webcam. #CES=~2000.

See archived day: http://synapse.vit.iit.nrc.ca/acese-data/home-19_2200-20_1430.php (password: vp4s-06)

CES extraction results: "Office" station



Indoor overnight monitoring (from office door): 17:00-9:00.

Over weekend monitoring (from office window): 17:00 on Fri - 9:00 on Mon.

Cam1: webcam. # CES=148. ACES video = 360 Kb

Cam2: 8mm SHARP camera. # CES=~2000. ACES video = ~5 Mb

(Blue - changing moving blob. Red - foreground history. Green - around the object).