Canada Border Services Agency
Agence des services frontaliers du Canada

Université d'Ottawa | University of Ottawa

VIVA Research Lab

Interactive Surveillance Event Detection

uOttawa:
Chris Whiten, Robert Laganière,
Ehsan Fazl-Ersi, Feng Shi
CBSA Science & Engineering Directorate:
Dmitry Gorodnichy, Jean-Philippe Bergeron,
Ehren Choy, David Bissesser
Ecole Polytechnique Montreal:
Guillaume-Alexandre Bilodeau

Background

- First participation to SED task
- Limited submission results
  - *Person-runs event detection*
- Work in progress…

- uOttawa works on automatic the event detection part
- CBSA works on the interactive part

Design objectives

- Problem of high relevance to CBSA
- **To improve computational performance**
- Traditional framework:
  - to work with spatiotemporal features
    - Feature detector
    - Feature descriptor
  - Bag of words
  - SVM classifier
- Inspiration from MoSIFT (from CMU)
- Inspiration from recent fast image matching techniques
  - Fast feature detector
  - Binary descriptor

Operational need

CBSA ASFC

- Surveillance cameras are heavily used by CBSA (in particular, in Airports)
- Two modes of operation:
  - Real-time: eg. to send a traveler to secondary examination
  - Post-event: eg. evidence extraction

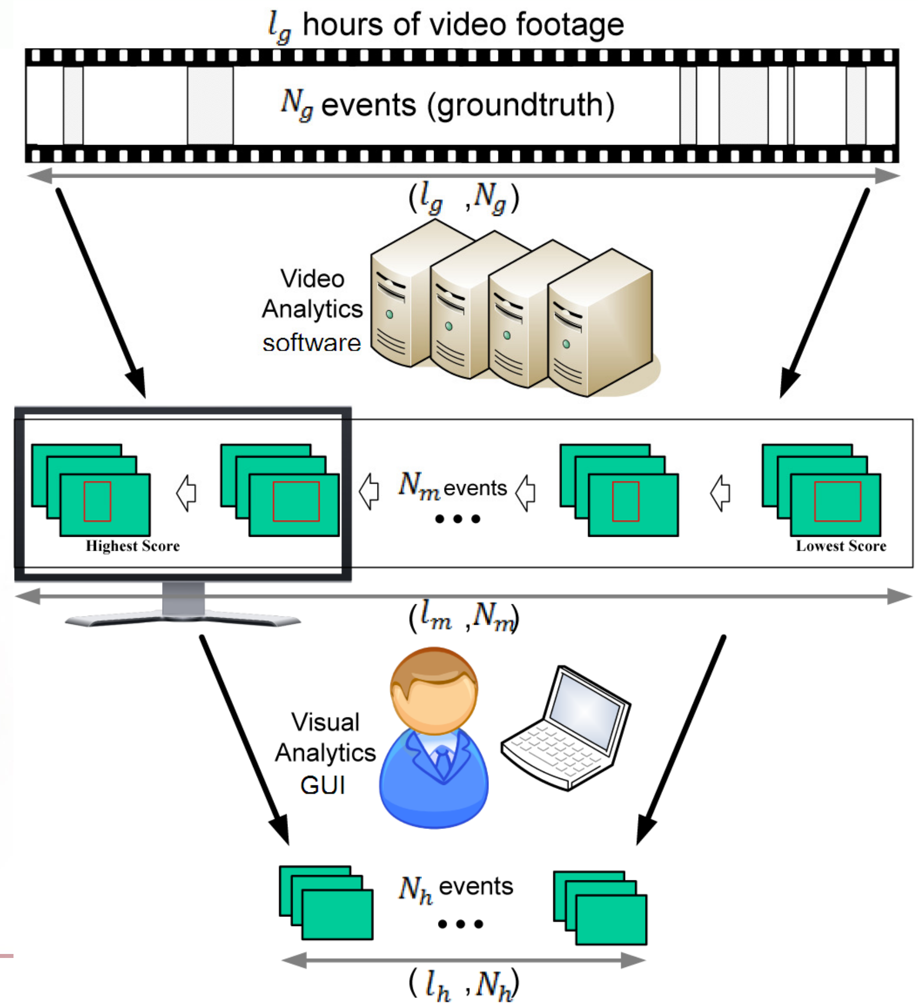
- In either mode, the decision - to trigger or not trigger alarm - needs to made within limited amount of time

Machine-Human approach

- Current *Video Analytics* algorithms produce lot of false alarms

- Filtering such amount of false alarms requires efficient *Visual Analytics* tools (GUI) …

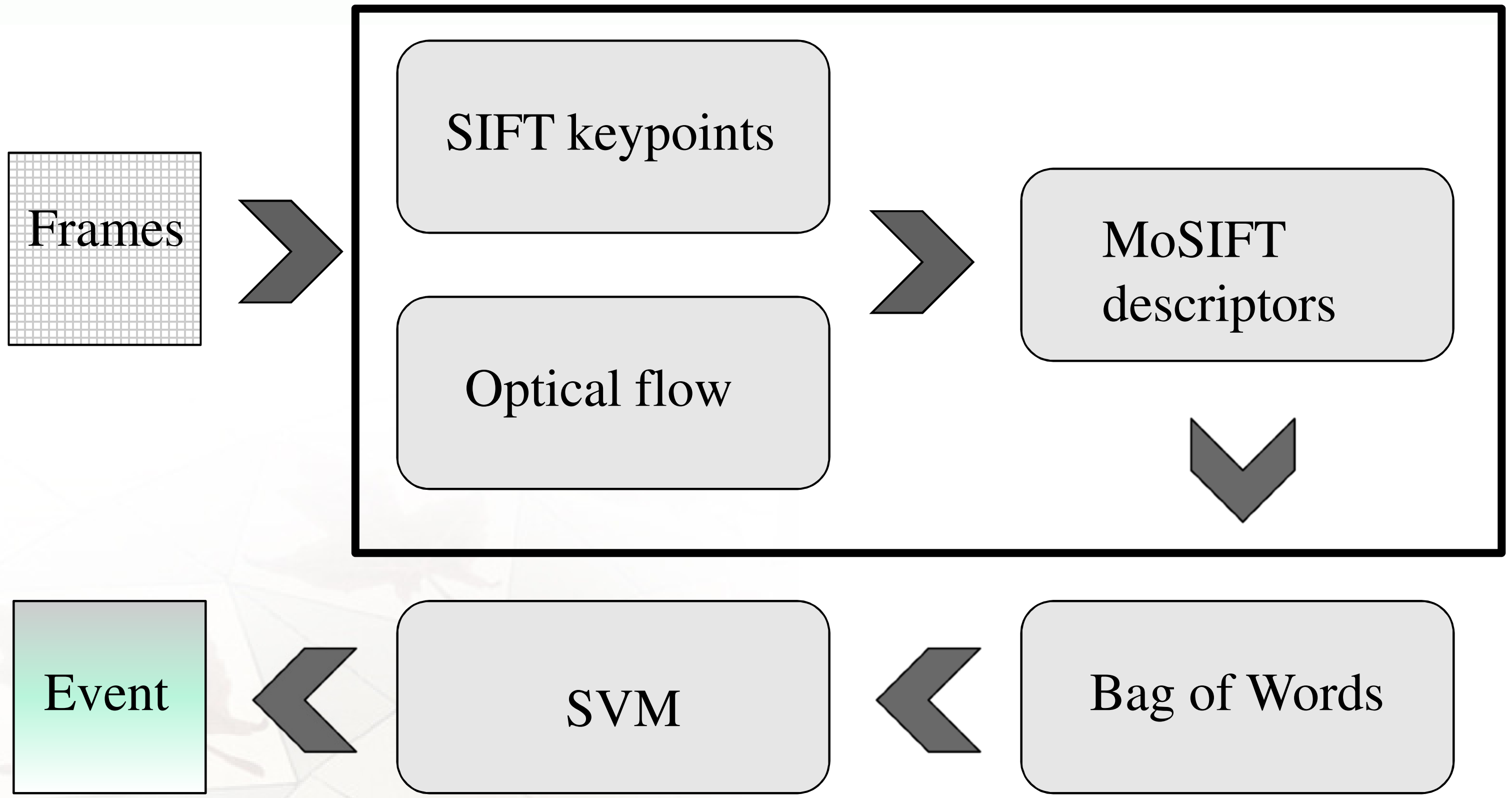
  … that makes use of humans visual recognition power for fast processing of large quantities of data



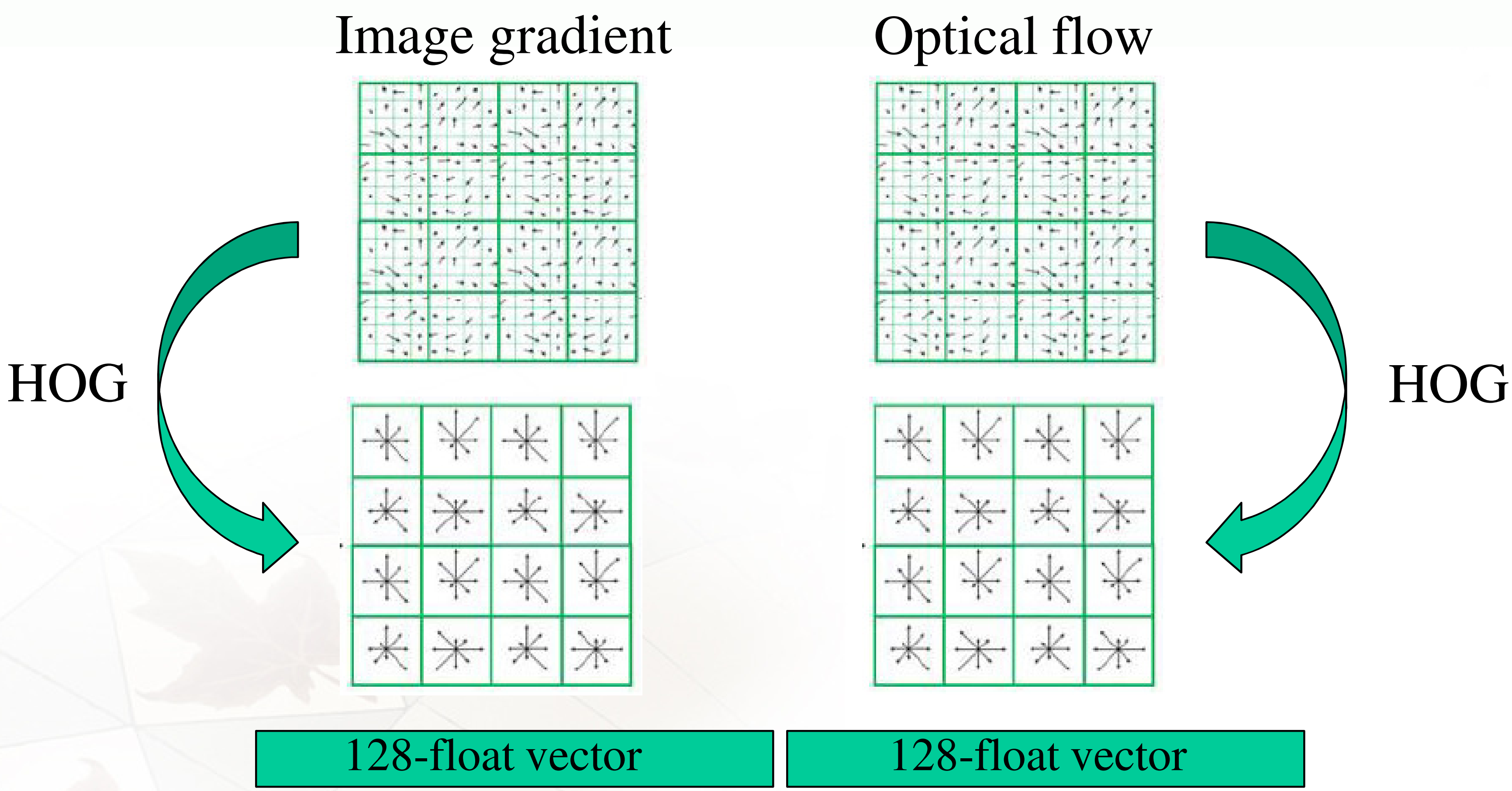
Event detection by Video Analytics

- Most Video Analytics approach are based on space-time points
- Historically, spatiotemporal descriptors have used gradient-based features (SIFT, Histogram of Oriented Gradients, etc..)
  - Slow to detect/compute/match
  - Difficult for the massive scale of surveillance data
- MoSIFT is a good example of such a space-time descriptor

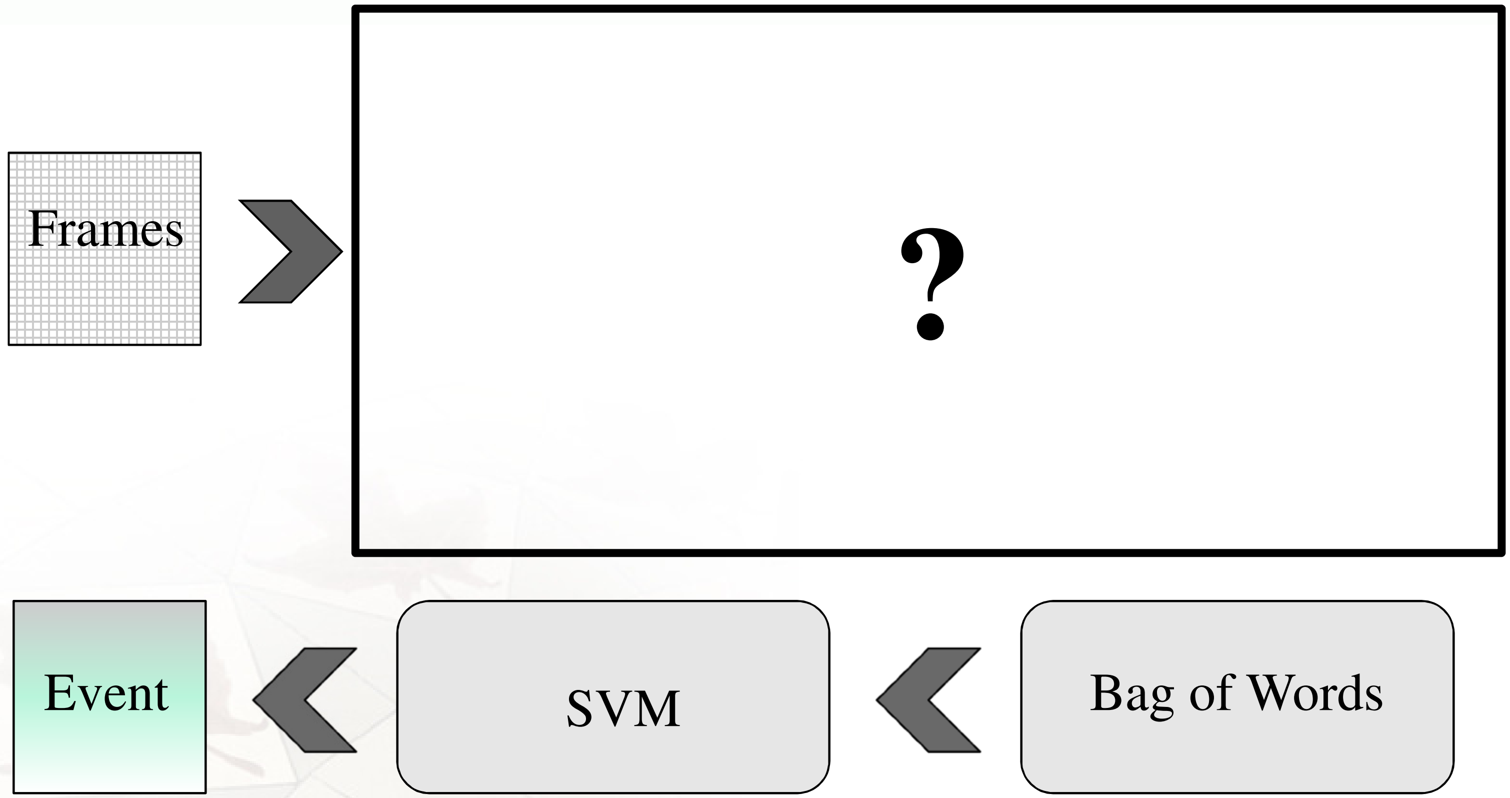
MoSIFT Approach



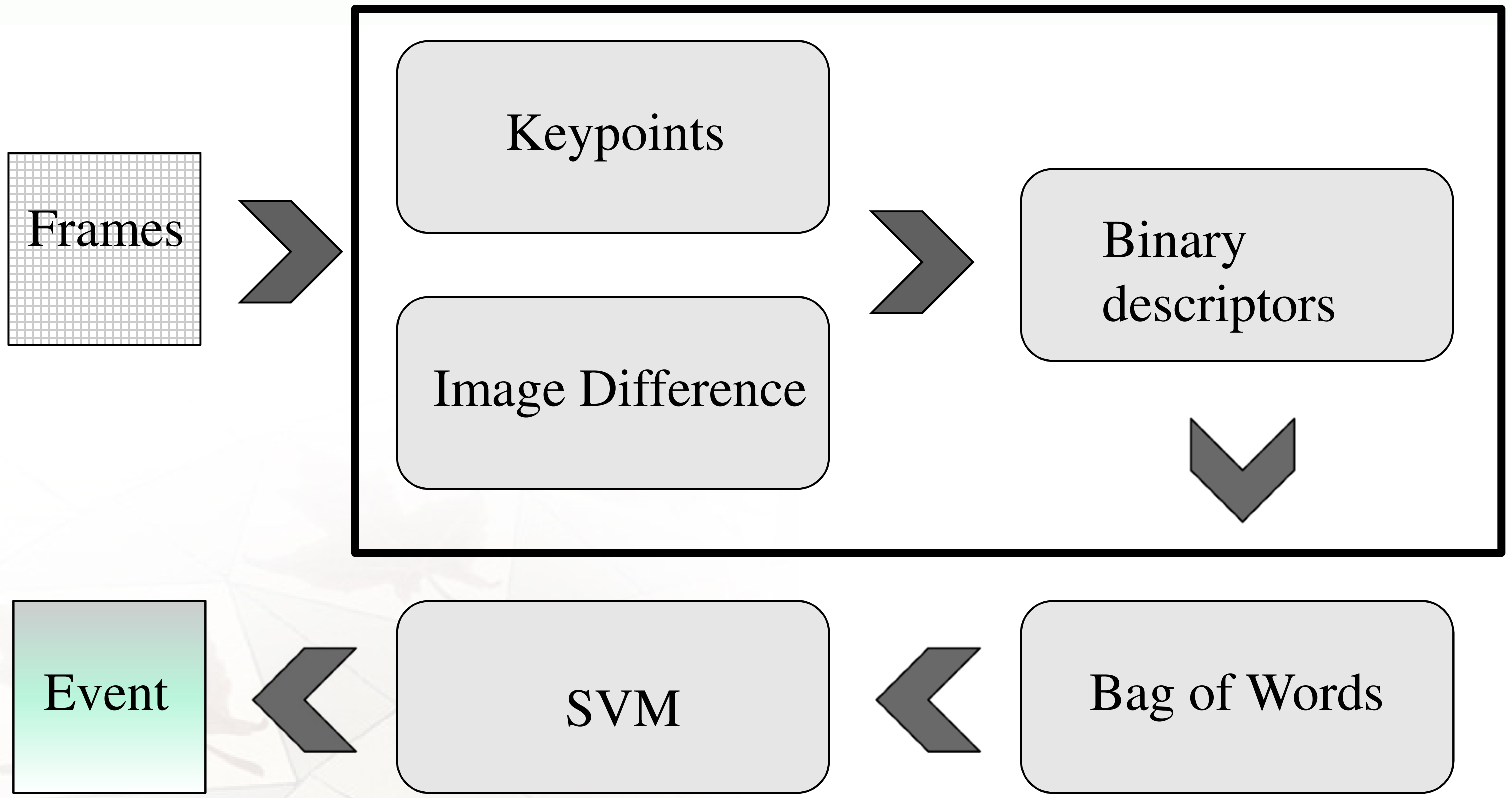
MoSIFT space-time descriptor



Another approach



Our Approach



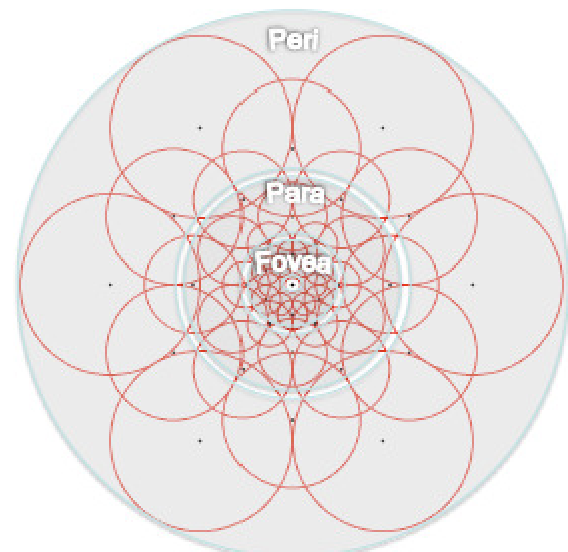
uOttawa

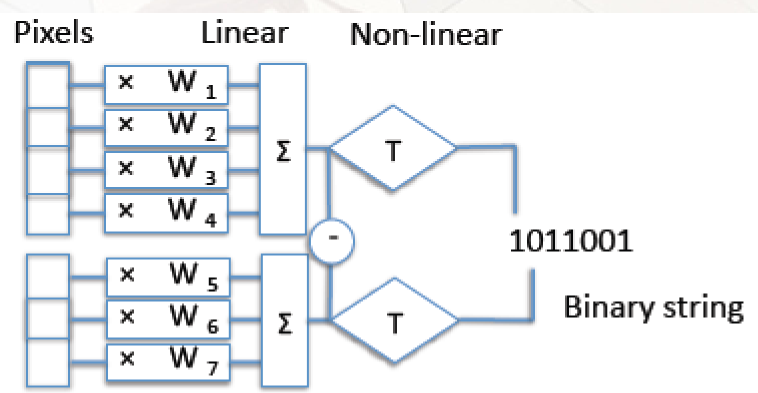
Extracting space-time descriptors

- We elect to use the recently proposed FREAK descriptor
  - Represents local keypoint with a binary string
  - Efficient to detect/compute/match
- The bytes in the FREAK descriptor follow a coarse-to-fine ordering

  First 16 bytes correspond to a human's peripheral vision

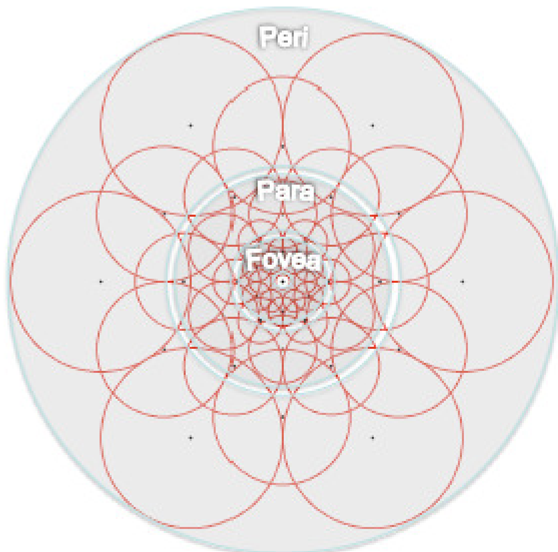
  Remaining 48 bytes encode finer details





FREAK descriptor

CBSA ASFC



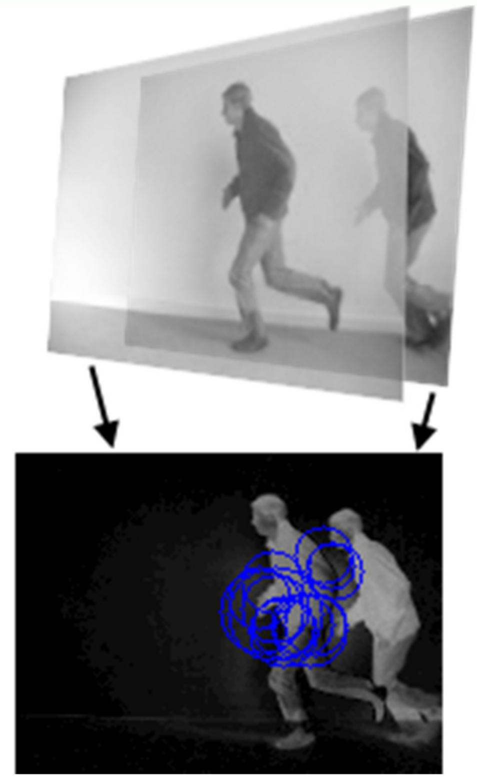
intensity comparisons

512 bits

uOttawa

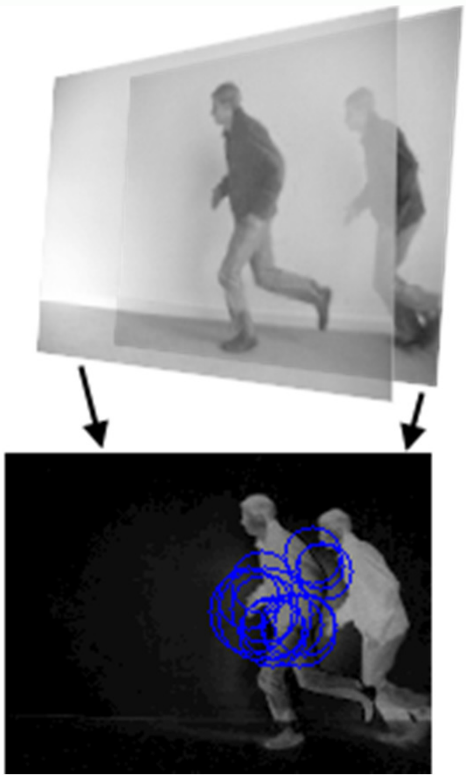
Extracting space-time descriptors

- At frame t, we compute the difference image between frame t and t – 5
  - Implicitly encode motion in the difference image
  - Avoid costly optical flow computations

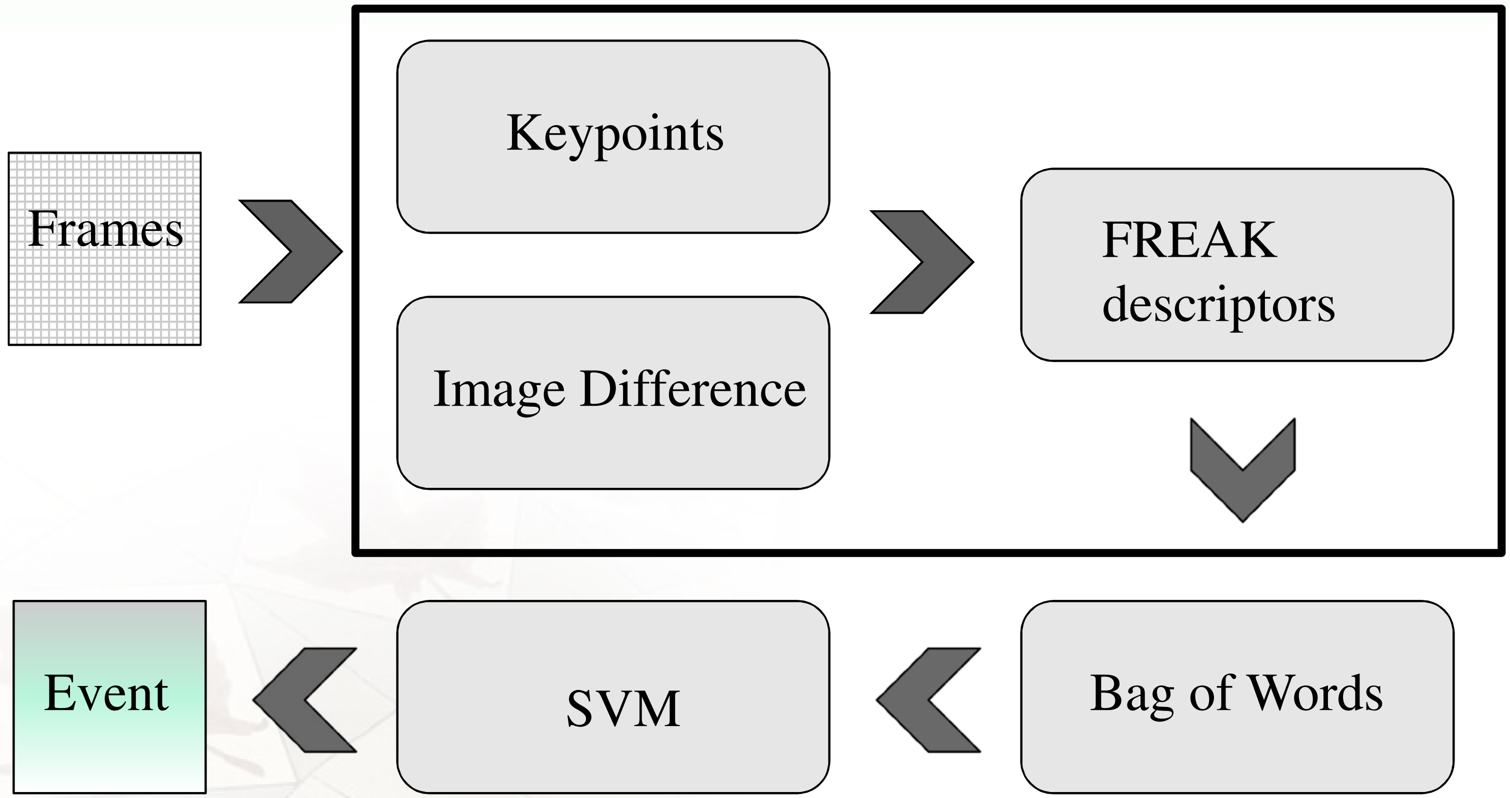


Extracting space-time descriptors

- For event recognition, we want to learn the action, not the actor
  - Avoid “finer detail” bytes
- We choose to keep only the first 8 bytes of the FREAK descriptor
  - Compact
  - Efficient
  - Encodes action in a more generic way
- 64 bits



MoFREAK Approach



Bag of Words in Hamming space

CBSA ASFC

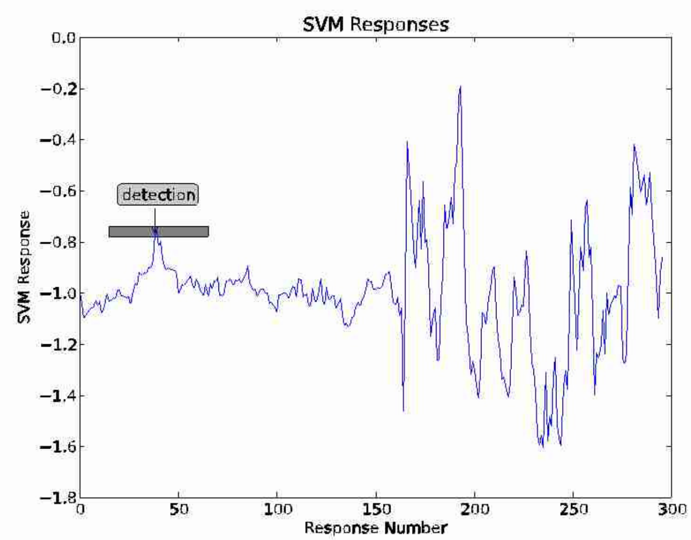
- We work with a binary descriptor
  - which allows us to avoid Euclidean distance
  - and instead use more efficient Hamming distance

- In addition
  - We use random clusters
  - Perform as well as K-means

uOttawa

Automated Event Detection

- Each bag-of-words feature is fed into an SVM
  - The SVM uses the histogram intersection kernel
- Each classified BOW feature returns a float
- The set of all classifications gives a distribution with many peaks and valleys
- Sufficiently large local maxima = event

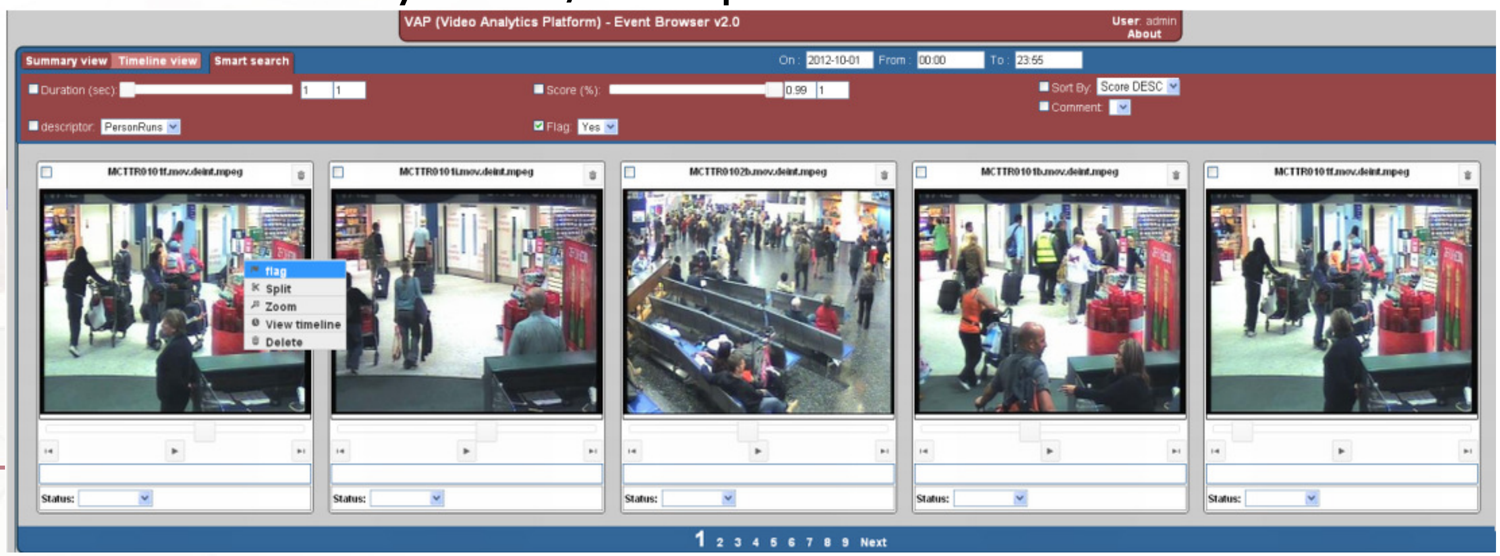


Manually Filtering False Positives

CBSA ASFC

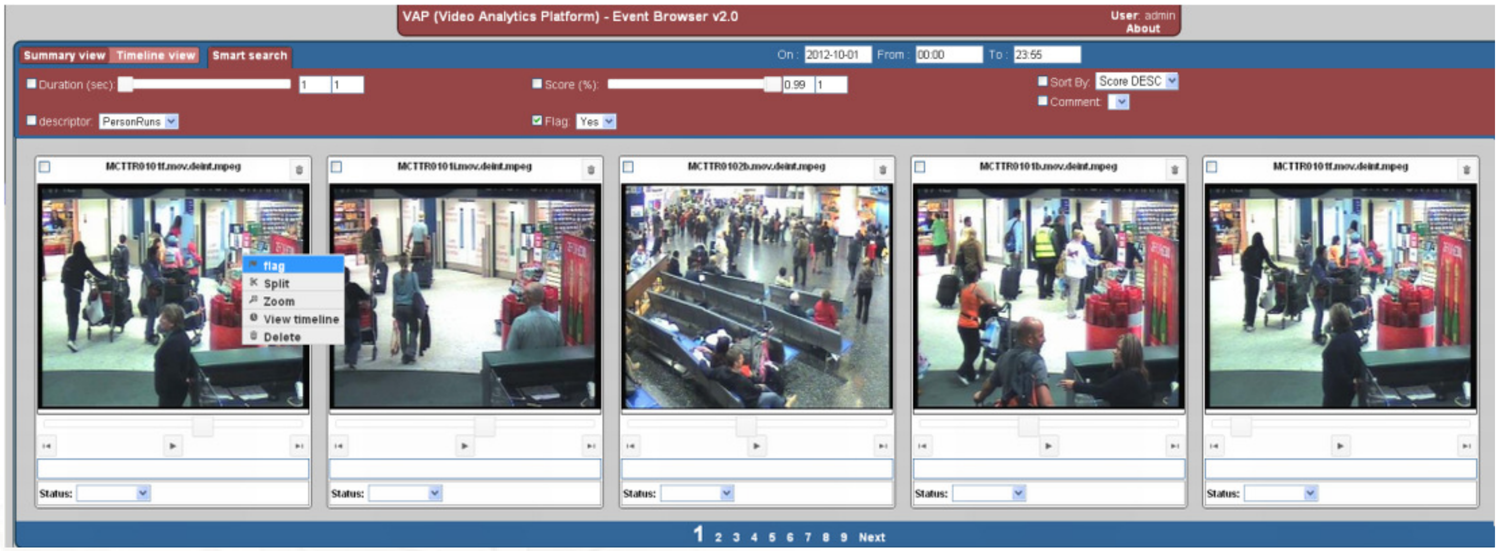
- The event detection system yields many false positives
  - Requires human feedback to know which detected events are legitimate
- Visual analytics system:
  - Events are presented in order of SVM response
  - to allow a user to efficiently navigate detected events to identify false/true positive events.

CBSA VAP platform



uOttawa

VAP Browser interface



- Using this visual analytics platform, a human operator is able to process over 600 detected events in a 25 minute time-window (24 events per minute)

TRECVID submission

CBSA ASFC

- We submit the results for the *person-run* event
  - Events were detected using MoFREAK approach
  - Events were filtered using VAP browser

  - 15 events were extracted

uOttawa

Person-runs Detections

http://www.site.uottawa.ca/~laganier/video/runs.avi

Conclusion

- Using recent advances in binary descriptors, rather than gradient-based descriptors, makes processing surveillance footage much more feasible
  - Currently 3 times faster

- Machine-human approach should however prevail:

  Video Analytic component allows to detect alarms automatically

  Visual Analytic interface is critical for efficient filtering of false alarms.