TRECVid Event Detection: Proposed Evaluation Plan

R. Travis Rose

Jonathan Fiscus

Martial Michel

John Garofolo

Data

- Large video surveillance corpus
    - Gatwick Airport Multiple Camera Tracking scenario
    - 10 sessions, 5 cameras, 2 hours each
    - 100 hours total
    - Collected on non-consecutive dates
    - Division between training & test subsets (TBD)
- Digitized from PAL BetaCam to 25 fps, 720x576 PAL Digital Video
- Trans-coded to MPEG2 for distribution

Event Definitions

- Event
  - Example: bag drop
- Event instance
  - bag drop, start/end times (t1,t2)
- Event observation
  - bag drop, start/end times (t1,t2)
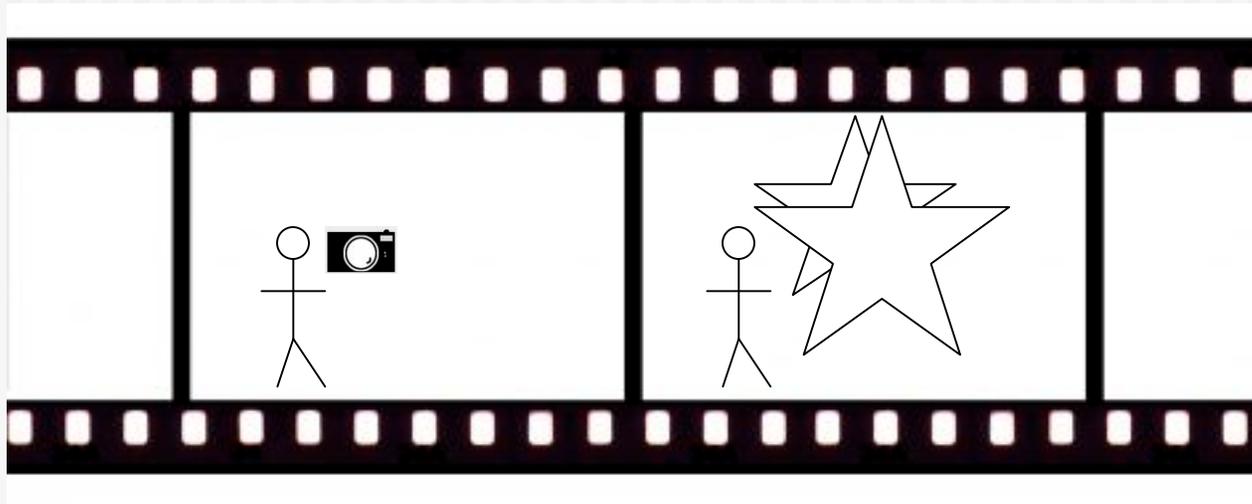 - at location (x,y) in image
  - seen in camera C_n

Event Types

- Instantaneous event
  - This event occurs within a small number of frames, e.g., 2 frames
- Spanning event
  - This event occurs over a longer time span, e.g., seconds to minutes

Instantaneous event

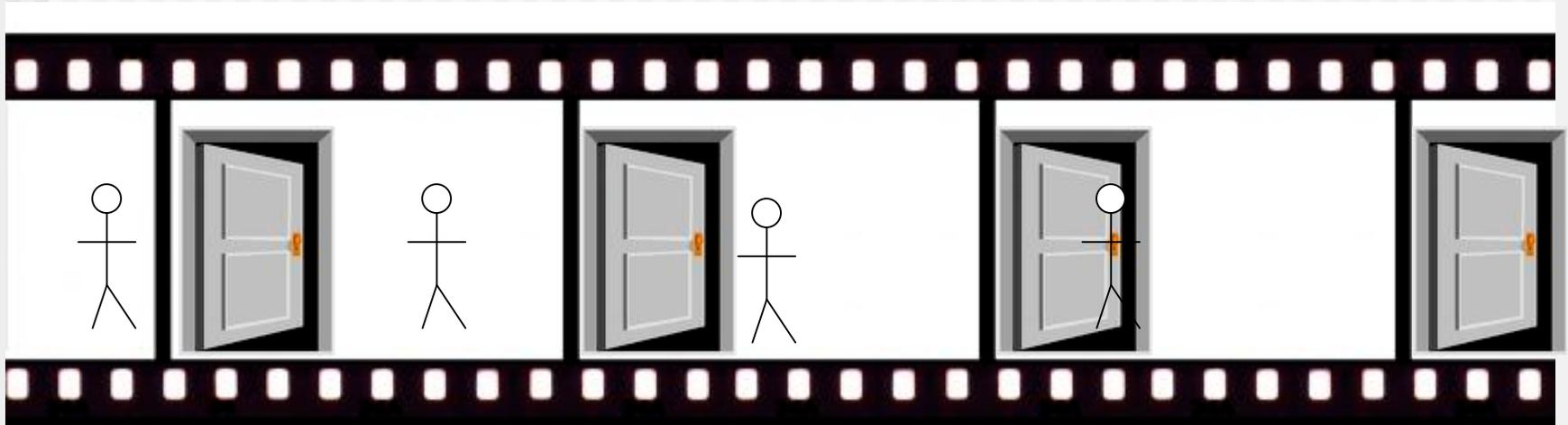ν Example:
  - ν Taking a picture
  - ν State change occurs in adjacent frames

Spanning event

- Example:
    - Entering a doorway
    - Requires more than 2 frames to detect

Annotation

- Annotations will include the following:
  - Label (event id), event, begin and end times
- Annotated training data may include a bounding box with coordinates (x1,y1,x2,y2)
- Formats may include XML or CSV (comma-separated values)

System Output

- ν Systems to provide output for each detected event, which includes:
    - ν Label (event id), event, begin and end times
- ν Systems may optionally include a bounding box in the form (x1,y1,x2,y2) for each event
- ν Formats may include XML or CSV (comma-separated values)
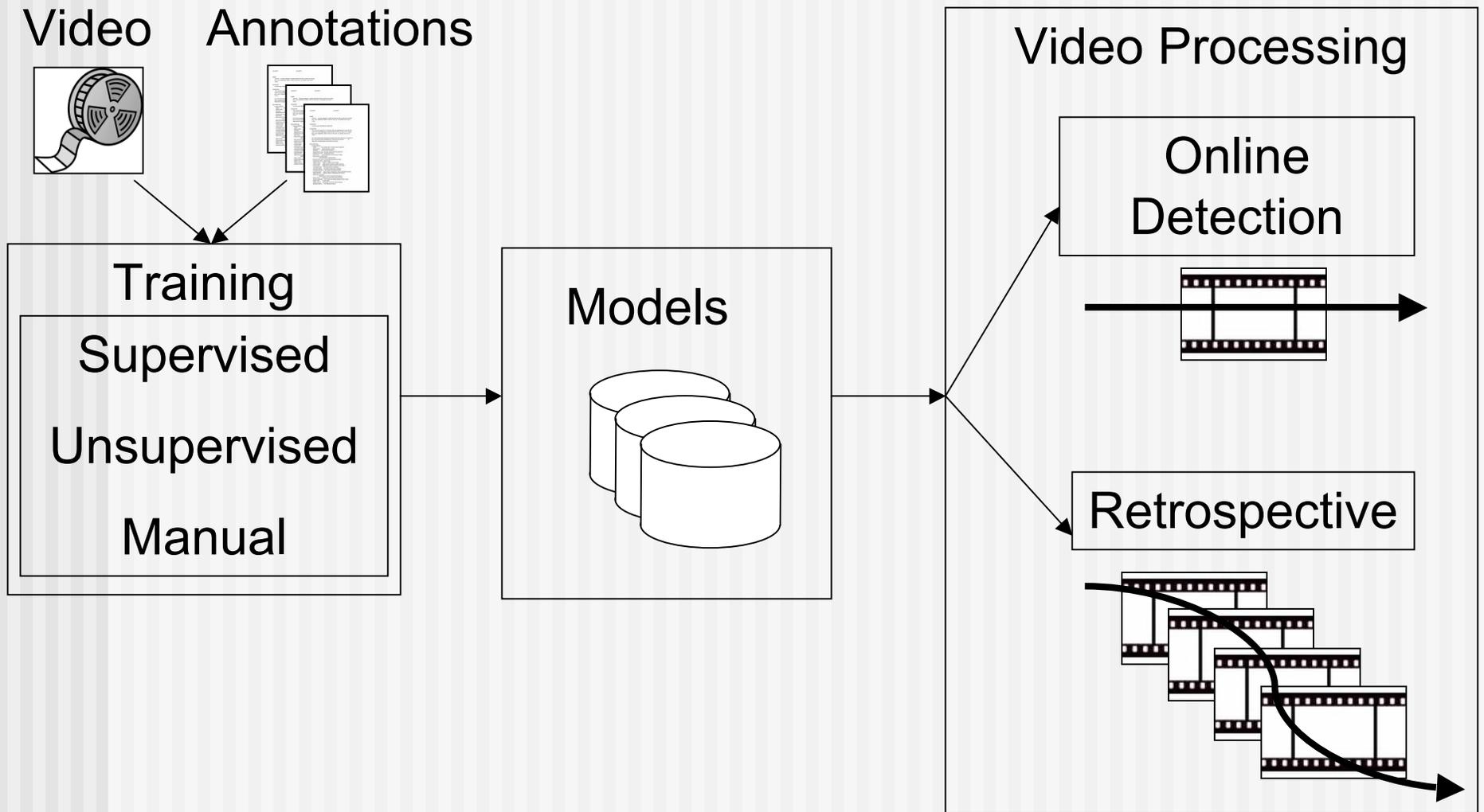
Evaluation Conditions

Testing Video Processing Variations

	Online	Retrospective
Unsupervised	Supported	Supported
Semi-supervised	Supported?	Supported?
Manual	Supported?	Supported?

Training Variations

Training and Video Processing are Decoupled

Proposed System Operation

Online Detection

- ν Videos are processed as streams
- ν One pass for each kind of event
- ν Detected events recorded as observations in separate metadata stream
- ν Can look at frames up to currently processed frame, cannot change past decisions
- ν Videos to be processed using real-time algorithms

Retrospective Analysis

- v Videos may be processed iteratively
- v Multiple passes allowed
- v Detected events recorded as observations in separate metadata stream
- v Videos may be processed using real-time or slower algorithms, up to a maximum time, e.g., 100X real-time (TBD)
- v Plan to define specifications for a hardware platform

Additional Training Questions

ν Training on public datasets allowed (?)

ν Proprietary datasets allowed (?)

ν What will be the division between training and test data?

    ν Preferred:

        • Divide by day (of collection)?

        • Divide by hour?

    ν Alternative:

        • No withheld data?

Outstanding Evaluation Work Items

- ν What will the events be?
- ν What will be provided for event training data?
- ν What metrics will be used?
- ν How can researchers share annotations and/or other intermediates?

Proposed Measurements

- Since this is a pilot evaluation, alternative metrics are being considered, including:
  - False alarm/missed detection metric
    - Systems would emit a confidence score for each detected event instance
    - Decision Error Tradeoff curves
  - Precision&Recall --> Average Precision
    - Ranked list similar to TREC
- Training time, processing time(s) for each kind of event
- Others, TBD