

TREC Vid Event Detection: Evaluation Protocol

R. Travis Rose

Jonathan Fiscus

Martial Michel

John Garofolo

Evaluation Goals

- ✓ Systems detect events in airport surveillance video (indoor)
- ✓ Evaluation protocol must:
 - ✓ Target a specific application for system tuning
 - ✓ Characterize performance over a wide range of operating conditions

Evaluation Protocol Synopsis

- ✓ Events are independent for eval. purposes
 - ✓ Systems output for each putative event occurrence:
 - The start/end times of the event
 - A confidence score
 - A binary decision about the event occurrence
- ✓ Detection performance is a tradeoff between missed detections and false alarms
 - ✓ Decision Error Tradeoff curves graphically depict performance
 - ✓ *A priori* application requirements unknown
 - A “Surrogate Application” needs to be defined
 - Optimization to be achieved using a “System Value Function”

The Role of Event Detection

Component vs. End-User Technology

- ✓ Event detection may be viewed as a component technology or an end-user application
 - ✓ Component technologies are generally more flexible
 - ✓ Components can be cascaded and fused

Metrics Discussion

- ✓ Precision/Recall have been considered
 - ✓ However, ranked lists are not well-suited to cascading/fusing components
- ✓ Decision scores:
 - ✓ Based on missed detection/false alarm
 - ✓ Allows tuning of the P_{miss} and P_{FA} probabilities
 - ✓ With decision scores, we can judge how well a system works for a specific accuracy tradeoff as well as a large range of accuracies

System Output

- ✓ Systems to output confidence scores as follows:
 - ✓ An ordinal value: a non-uniform density function that distinguishes close hits from far misses
 - ✓ Coherence across events: density function is consistent across events that facilitates down-stream processes using mutual information

System Development

- ✓ Problem:

- ✓ Difficult to know the error tradeoff requirements until a system is deployed
- ✓ But, system developers need a target tradeoff for tuning

- ✓ Solution:

- ✓ Pick a surrogate application that approximates the intended application
- ✓ This approach will permit some space for tuning around the surrogate