

Minutes from TRECVID 2008 Event Detection Planning Telecon

January 30, 2008

Jon Fiscus, NIST
John Garofolo, NIST
Paul Over, NIST
Travis Rose, NIST
Francois Bremond, INRIA
Isaac Cohen, Honeywell
Mert Dikmen, Dennis, UIUC
David Eichmann, U. of Iowa
Sadiye Guler, intuVision
Murtaza Taj, QMUL
Michael Nechyba, Henry Schneiderman, Pitt Patt
Ram Nevatia, USC
Noel O'Conner, Dublin City U.
Jean-Marc Odobez, IDIAP
Aristodemos Pnevmatikakis, AIT
Rainer Stiefelhagen, Karlsruhe
Lexing Xie & Thi Lan, IBM

Agenda:

- Role call
 - Old business
 - Discussion of previous calls (if any)
 - New business
 - Data: Microcorpus discussion (one person per organization)
 - Schedule next telecon
-

A question & answer session was held at the beginning of the meeting, followed by a discussion of the microcorpus and candidate events.

Next telecon:

February 6th, 2008, role call at 10:55 am, telecon to begin promptly at 11 am (Eastern Time).

Will camera parameters be available?

A calibration target with approximately average human height (~1.6m) was used. Calibration data may be available for each camera. NIST will work on making these available.

Can systems use information from more than one camera to detect events?

- David Eichmann suggested allowing people to use multiple cameras, and have them report it as a condition. As a result, multiple-camera systems should be run as a contrastive condition.

- Community discussed that either single or multi-camera processing would be valid

- Straw poll taken:

Question: Will the primary condition use single camera input?

Outcome: Yes, majority in favor of the single-camera condition as the primary condition (i.e., a required condition).

Will multiple submissions per sequence be allowed?

Since system parameters can change, some small number of conditions will be possible per sequence (to be determined).

What contextual information is allowed?

For systems that use supervised training, this allows manual labeling of regions, objects, etc. Knowing that the test data was collected from the same camera views, annotation on training data is allowed, however systems must process the test data automatically

- Straw poll taken:

Question: Was the proposed definition above acceptable?

Outcome: Majority in favor.

Does the processing need to be causal, i.e., can past decisions change based on the results of additional

processing?

This relates to the online detection vs. retrospective analysis tasks. In online detection, past decisions cannot be changed. Online detection may be explored as a "freestyle" exercise.

- Straw poll taken:

Question: Will the primary task be retrospective or online detection?

Outcome: Majority in favor of retrospective analysis. The retrospective task allows past decisions to change; for example, this could be based on multiple processing passes.

What about anomalous events?

This can be explored as a freestyle exercise.

Is it possible to download the data?

Yes, this is possible; we anticipate a ~250 GB download for the entire corpus. David Eichmann volunteered to help set up a mirror to make the corpus available online. Others are welcome to be mirror sites.

What video format will be used for distribution?

No one objected to the MPEG 2 files distributed for the micro corpus so without objection, the data will be supplied in the same format.

Call for TRECVID goes out Friday, Feb. 1. Will include information such as:

- where to get forms
- request for mirror sites
- Please feel free to post additional comments/questions to the list.

Can we get a map or floor plan of the area?

We will post a sketch of the area.

How are the data synchronized?

We will show the synchronization approach that uses a time-code generator.

Events discussion (observations from the microcorpus). Additional questions appear below.

Aristodemos, AIT:

Dropping something

Arranging a piece of luggage

Pausing by doorway vs. hurrying across

Small group meetings

Dennis/Mert, UIUC:

kid walking

big bags

ATM use

door opening

running

crossing zones

cell phones

turning around

dropping items

picking up (object)

Personnel in yellow vests appear

sit/stand

pulling carry ons

stationary standing ("loitering")

meeting/hugging

separating

putting objects on display (in stores)

exchange bags

pointing

Some events are focused on objects vs. activity

Can an event occur off camera?

We will focus on the on-camera events in the first year.

Isaac Cohen, Honeywell:

Sometimes there are unexpected or "surprise" objects, such as a bicycle.

Jean-Marc, IDIAP:

Bike, large objects

pausing at a place that is unexpected

Noel, DCU:

- similar observations

David, Iowa:

Meandering -- some people stalling

Examples of "anti-events": exit block, person goes to elevator, does not use

Thi Lan, Lexing, IBM:

Also note the distinction between object recognition vs. events

A people counting task could be devised.

If we track direction of movements of people, e.g., moving towards an exit, then systems could detect "abnormal" movements (in the opposite direction of traffic flow).

Murtaza Taj, QMUL

In this corpus, we need to have a don't care region;

Group discussion on how to decide a "don't care" region?

Possible methods for designation:

- visibility of the entire body
- if a human can see it, it should be allowable, still need to address the far background (perspective)
- pixels per person is an option
- draw lines that are reasonable for each camera view.

Michael, Henry, Pitt Patt

There are a lot of events occurring all the time in the background on cameras 2, 5.

Need to delineate background vs. foreground (i.e., using a "don't care" region).

Implies the need to have a "do care" region.

In addition, it will be important to develop guidelines, i.e., how to explain the event to an annotator, etc.

Sadiye, intuVision:

There is one person pushing luggage & then leaves it to go shopping;

child running away;

milling around.

Ram, USC:

There is a lot of hugging, meeting; searching baggage/purse. Bike event may not be uncommon ...

Most events are rather subtle -- needs a sophisticated analysis of movement, such as hands, limbs, objects.

Want to choose events that are tractable.

As we come up with the candidate events, we can ask sites to assign difficulty level to each.

Group discussion regarding side information such as doors, elevators, environmental models, ATMs, etc.?

This can be provided as an initial annotation; we want to develop that and be able to share, which can be a community exercise.

Francois, INRIA:

We would like to have well-defined events, as there can be fuzziness surrounding some definitions. We expect some events occurring quite often in video vs. others that have lower frequency.

Pitt Patt:

Agree, next stage will be to go deeper in defining the events.

NIST:

bump event

taking a picture

Outcome of Event Discussion:

- NIST will post a survey of candidate events and ask sites to assess their difficulty.