Minutes from TRECVid 2008 Event Detection Planning Telecon
February 6, 2008

Jon Fiscus, NIST
John Garofolo, NIST
Paul Over, NIST
Travis Rose, NIST
Francois Bremond, INRIA
Larry Davis, UMD
Mert Dikmen, Dennis, UIUC
David Eichmann, U. of Iowa
Sadiye Guler, intuVision
Ram Nevatia, USC
Ciaran O'Conaire, DCU
Henry Schneiderman, Pitt Patt
Heather Simpson, LDC
Andrew Senior, Lexing Xing, IBM
Murtaza Taj, QMUL

Agenda:

- Role call
- Old business
  - Room schematic (posted on the web site)
  - TRECVID Call for Participation announced
- New business
  - Discuss the results of the event survey (see the web site)
  - Open discussion to specify the contextual/side information given for each camera, e.g., door locations, ATM locations, normal flow, etc.
  - No-score regions implementation & discussion

- Schedule next telecon

Next telecon:
February 13th, 2008, role call to begin promptly at 11 am (Eastern Time).

Review old business.

Review event survey.
Fiscus: We note that the definition of each event would need to be made more specific. Based on the ratings, there appears to be a spread of difficulty in the events.
- examples of difficult events: embrace, cell phone, object exchange
- many of the events appear to be tractable

Editorial Summary: There was a long discussion about the expected levels of performance and the interaction of several factors like camera view, background clutter, event difficulty, complexity of motion, number of people, type of event (people tracking, object tracking, etc.) [A more complete dialogue appears below.]

During the discussion, two options surfaced to control for the variability: reduce the size of the test set or keep the full test set and select a small subset to annotate with detailed factors. Since the evaluation seeks to push the technology in terms of corpus size, reducing the test set size was discounted. NIST does not have the budget for further annotation (beyond temporal localization) so a straw poll was taken to determine if a community-based, post-evaluation annotation exercise would be feasible.

- Straw poll taken:
  Question: Are sites interested in contributing to community annotation?
  Outcome: Majority in favor of sharing annotations.

Group discussion of community annotation effort.
Many expressed interest in contributing to annotation, provided it is regarded as a pilot exercise. People also wanted to know about factors such as the estimated time required. As a result, there was a suggestion that what to annotate and how much should be defined. Others pointed out that processing the full 100 hours would take several months, and as such the corpus is a tremendous amount of data.

There was a proposal that the community use an appropriate annotation tool that is shared across sites. One candidate is Viper, which may be hard to use for this evaluation, but it is being considered.

Action item (all):
If sites know of similar tools that may be appropriate, they can suggest these by posting to the list.

Group discussion about the required contextual information.
Francois:
typically use a lot of contextual info -- if you can draw on the image to mark what is an interesting area, etc.
- 2D region
- depends on the event definitions, e.g., "sitting" would need where sitting usually occurs.

Francois:
May need to wait until the required events are defined, e.g., ATM location is only important for ATM-related events
It's easiest to draw a 2D region and make it part of the system input.

Jon:
However, the regions by themselves do not preclude events happening outside of the specified region.
Also, it's hard to merge all of the contextual requirements.
-- suggested approach: sites can annotate what they need, then post online

Dave:
-- obvious things like bench, ATM, elevator can be shared; then groups can easily extend the metadata if needed.

Action item (all):
-- determine what types of contextual information developers may need as "side" information
-- important to limit the specification of "context" to the camera views
-- post suggestions for contextual information requirements to the list

- Straw poll taken:
  Question: Should there be a deadline for the community annotation to be completed?
  Outcome: Majority in favor of having a deadline.

Group discussion of the no score regions.
- Lexing - should we let annotators see all of the video data?
- Jon Fiscus - We prefer to black out regions of the video so that annotators are not affected by visual information in the no-score region.
- Paul Over- also, we want to bound computational complexity and lower annotation cost
- Heather Simpson: To reduce annotation load, we would like the annotators to not see the no-score regions.
- **outcome:** NIST will propose no-score regions for each camera view and post them for next week's discussion.

---------- End of Minutes ----------

---------- Dialogue from event survey discussion ----------
Francois:
-- note that the difficulty of detection also depends on the quality of the source video, clutter, and sizes of crowds.
-- some events may become hard depending on the number of people.

Larry Davis:
Two problems:
-- complexity of scene
-- complexity of detection
This can vary depending on whether the detection, tracking, and segmentation are also hard.

David Eichmann:
This task should not be more difficult than current retrieval evaluations because of homogeneity of the footage.

Ram:
There is some variance in the assessment (I think). For example, my ratings were less optimistic.
Doing detection and tracking of objects for events will be hard.
- depends on the environment, how cluttered, etc. In addition, this will depend on ability to find people, e.g., pushing a trolley.

Sadiye:
Consider also there is bias introduced by which video camera you sample. Should look at assessing the difficulty of detecting events in particular views.

Jon Fiscus:
We want to identify factors that affect performance. Rather than specifying all of them in advance, we should be able to discern several of these from the evaluation.

Ram:
-- may actually be looking in ROC region that is near the origin
-- currently exploring to see how feasible the task is
Suggest a reasonable experiment to try is using a larger microcorpus.

Jon Fiscus:
Let's plan to do this with the training corpus, which will be a much larger dataset (i.e., 50 hrs).

Dave:
Notes there is hesitation about tasks that may not have "good" performance (e.g., 80% accuracy), but we want to baseline the technology capability with realistic video.

Travis:
Is this partly addressed by having a range of difficulty of events? Dave responded yes.

Paul:
Because of the variability in several of these factors, it's hard to make predictions.
-- Regarding the event survey, interpret the result as most participants willing to give most events a try
-- Since this is a pilot evaluation, systems and approaches will be stressed by the data. The evaluation will show whether these approaches scale, can be extended, etc.

Francois:
It depends also on how the evaluation is going to be done; important to explain from beginning.
Maybe try to define the difficulty of detection in terms of clutter, which view, etc. otherwise in the evaluation it will be very hard to get good performance. If just a score may not have enough data to do further analysis of failure modes e.g., if an event is easy to see in foreground vs. hard to see due to background clutter

Jon:
Some approaches:
- score everything
- do annotation based on distance, amount of clutter, etc.
- may need to sub-sample; importantly, we need to use an approach that is cost-effective

John:
- sites may also contribute annotations. These may be further refined as a post-facto annotation exercise (after results are submitted). Can also look at samples of the data where there is high vs. low agreement among systems.

Jon:
- One option would be ask systems to spatially identify where the events occur, then map the spatial locations onto the reference annotations. With the inferred spatial annotations, the systems could be conditionally scored on various factors.

Ram:
Doesn't seem the stress is so much the volume, rather it should be the complexity of the data;
can do a smaller amount of data processing, for example.

John:
The challenge also is to preserve the density of the events in the data, which we feel is important.