Version 7 — April 19, 2013

TRECVID 2013 Multimedia Event Recounting (MER) Evaluation Plan

The Multimedia Event Recounting evaluation analyzes the recounting output by the MED system. The purpose of recounting is to enable a user to rapidly and accurately identify their events in the clips returned from the MED system. Participation in the MER evaluation is open to all MED participants in TRECVID 2013 whose system always produces a recounting.

BACKGROUND

Each event in this evaluation

- is a complex activity occurring at a specific place and time;
- involves people interacting with other people and/or objects;
- consists of a number of human actions, processes, and activities that are loosely or tightly organized and that have significant temporal and semantic relationships to the overarching activity; and
- is directly observable.

The following terminology will be helpful. Each event kit explicitly defines an *event*. A clip that is *positive* for an event contains an *instance* of that event. The recounting summarizes the *important evidence* that the clip contains the event. Ideally, the recounted evidence should be that which the MED system used to detect the event. The recounting can also state that no evidence for the event was found. It is important for the system to *locate* the important pieces of evidence — both temporally (where in the clip) and spatially (where in the frame) to help the user quickly locate the event within a clip. Evaluation of the recountings will focus on the ability of a human judge to rapidly and accurately decide which clips contain an actual instance of the event of interest; which we will call *triaging* the set of MED results.

The Event Kit text for 2013 will consist of

- Event name: a mnemonic title for the event
- Event definition: a textual definition of the event
- Evidence: a textual statement of evidence that indicates the existence of the event.

Evaluation may also include use of the 2012 Event Kits and/or other Event Kits of that type.

RECOUNTING Format

MED Systems will produce an output recounting file for each event and each clip. This file will be in XML format. The recounting is a list of observed evidence, where each piece of evidence on the list includes:

- id: a unique identifier
- Description: a concise textual summary of the piece of evidence. This would ideally be an English sentence or phrase meaningful to the user, but could alternatively be a code word
- Confidence: a score indicates how confident the system is in its detection and localization of this piece of evidence (range from least confident at 0.0 through most confident at 1.0).
- Importance: a score that indicates how important this evidence was in detecting the event (range from least important at 0.0 through most important at 1.0).
- Sources: a list of one or more sources of this piece of evidence, drawn from the following four possible values.
    - video (not involving O.C.R.)
    - visible_text (via O.C.R.)
    - speech (transcribed via A.S.R.)
    - non_speech_audio (sounds not involving A.S.R. textual transcription)
- Presentation Order: an ordinal number indicating the order in which the Triage Workstation should present the evidence to the user (i.e., 1, 2, 3, . . .).
- Pointer List: A list of pointers for each piece of evidence, to spatially and temporally locate the evidence within the clip. Each element on the list includes:
    - Start Time: an offset into the clip (either a time offset or a frame number) of the start of the piece of evidence
    - End Time: an offset into the clip to the end of the piece of evidence
    - Start Visual Bounding Box: defined at Start Time, the upper left (row, column), and lower right (row, column) pixel coordinates (in the video frame). The upper left corner of the video frame has coordinates 0:0 and the lower right corner of the video frame has coordinates that are the size of the video frame in pixels.
    - End Visual Bounding Box: defined at End Time, and the upper left (row, column) and lower right (row, column) pixel coordinates (in the video frame)

    Note: NIST will refer to the sub-clips that are defined by this list as “snippets.”

NIST will provide a DTD for this data format. The Start Visual Bounding Box and End Visual Bounding Box should be completely omitted for a snippet that is purely audio; otherwise both are required for each pointer in the Pointer List. A bounding box is allowed to enclose the entire video frame. The DTD will allow the recounting to optionally state the source(s) of evidence for each snippet.

A recounting will be maximally useful if it can help a user to triage a collection of clips more rapidly and more accurately. Therefore, the recounting should (1) include only the key pieces of

evidence (omitting anything unimportant), (2) textually state each such piece of evidence accurately and concisely, (3) maximally narrow down the time span where each of those pieces of evidence occurs in the clip, and (4) maximally narrow down the bounding boxes that identify where the piece of evidence occurs in the frame.

DATA TO BE PROCESSED AND JUDGED IN THE 2013 MER EVALUATION

MER participants should provide recountings for all events and all clips processed by their MED system. NIST will select a subset of the recountings for evaluation.

All submissions will be judged on their recountings for the selected set of (clip, event) pairs.

EVALUATION

The system's recountings will be evaluated by a panel of judges. NIST will create a Triage Workstation and provide it to participants and judges. Using this workstation, the judge will be instructed to study the event kit text (not the example videos) and then to triage the clip by:

1. Reading the entire list of textual descriptions
2. Viewing/hearing ***all*** the snippets defined by the spatiotemporal pointers
3. Classifying the clip as one of the following:
    - The clip *contains an instance* of the event
    - The clip *does not* contain an instance of the event
    - The *recounting* does not allow me to tell whether the clip contains an instance of the event
    - The *event kit* does not allow me to tell whether the clip contains an instance of the event

The event kit text will be available to the judge for reference.

The Triage Workstation will *not* display the source(s) of the information — neither for each piece of evidence, nor for the snippet(s) associated with each piece of information. The stated sources of information will be used for post-hoc understanding of the system.

For each submission and each event, NIST will measure the following characteristics of the recountings:

- Percent Triage Time: percentage of clip time the judges took to perform steps 1–3. above

$$Percent\ Triage\ Time = \frac{Total\ time\ needed\ to\ triage\ the\ event}{Total\ time\ of\ clips\ to\ be\ assessed}$$

- Accuracy: the degree to which the judges' assessments (step 3 above) agree with the MED ground truth.

$$Accuracy = \frac{Number\ of\ correctly\ labeled\ clips}{Number\ of\ clips\ to\ be\ assessed}$$

MER submissions whose recountings enable the judges to assess recountings the most rapidly and accurately will be considered the best. NIST will be evaluating the recountings, not evaluating the judges.

SOFTWARE RESOURCES

NIST will provide a Triage Workstation to be used in judging. The triage workstation will be implemented in Ruby on Rails and will run over the internet using a typical web browser. Participants are encouraged to improve the Triage Workstation and to provide feedback to NIST. The best suggestions will be integrated into the final Triage Workstation used by the judges for the evaluation.

DRY RUN

All participants will be required to participate in the MED dry run and generate recountings. The purpose of the dry run is to ensure both that the system outputs are being generated as expected and that they can be parsed by the evaluation pipeline. This exercise will also provide the evaluation team with insight into how the recounting will be used on the Triage Workstation for the judges in the formal evaluation. No evaluation of the recountings in the dry run submission will occur, and the only feedback that will be provided to performers will relate to any problems with data formatting, conformance to the DTD, and so forth. Teams are encouraged to use the Triage Workstation provided with their own judges to assess their performance.

SCHEDULE

Date (*all dates 2013*)	**Milestone**
April 1	Final design published in the TRECVID Guidelines on the web, and MER DTD available
May 1	Initial draft version of Triage Workstation available to participants.
June 5	2013 MED Progress Set disk drives available (for new participants)
July 1	MED/MER dry run begins
July 31	MED/MER dry run ends (MER dry run submissions due at NIST)
Sept 10	MED/MER participants submit Pre-specified Event runs (including recountings)
Sept 13	NIST releases Adjudicated MED results for the Pre-specified Events
Sept 27	NIST releases pre-adjudication MER results on the Pre-specified Events
Oct 2	NIST releases post-adjudication MER results on the Pre-specified Events
Oct 4	TRECVID speaker proposals due at NIST by noon (Gaithersburg time)
Nov 20 – 22	TRECVID workshop at NIST in Gaithersburg, MD