

# 2013 TRECVID MULTIMEDIA EVENT DETECTION EVALUATION PLAN

Jon Fiscus 6/27/13 9:07 AM

Deleted: DRAFT

## 1 Introduction

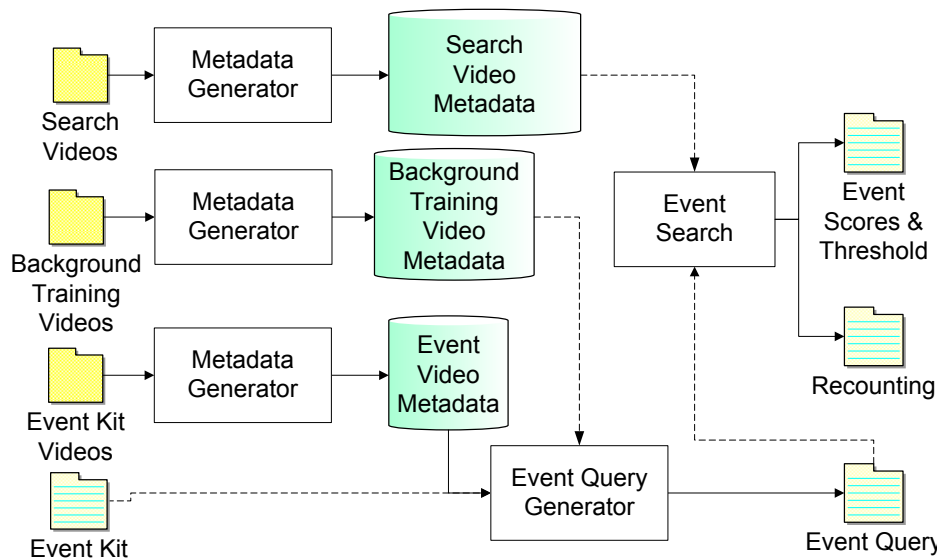
This is the evaluation plan for Multimedia Event Detection and Retrieval (MED) track of the 2013 TRECVID evaluation. The multi-year goal of MED is to support the creation of event detection and retrieval technologies that will permit users to define their own complex events and to quickly and accurately search large collections of multimedia clips.

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

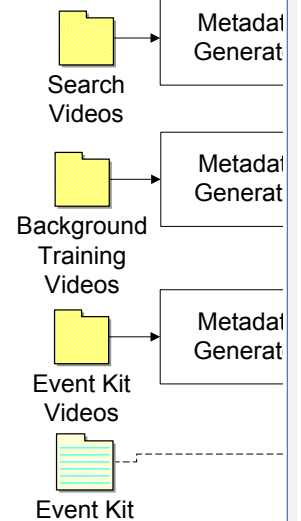
Jon Fiscus 6/27/13 9:07 AM

Deleted: which involves

A MED system (shown in Figure 1) is defined to have three separate modules: 1) metadata generation, 2) event query generation, and 3) event search.



Jonathan Fiscus 6/27/13 9:07 AM



Deleted:

Figure 1: MED System Block Diagram: The MED processing modules are shown in blocks and data is depicted as icons. Folder icons represent a collection of videos while text-folders are text/csv documents. Dashed lines in the diagram show that significant time can elapse between the computation of the data and its use in the following module.

The event query generator and event search modules process each event independently. The modules must be run in the following order and are limited to only the inputs shown in the block diagram and

described below:

- 1) The metadata generator extracts the video metadata for all videos in the input set. The metadata describes the content of the video that support the event search. Since this must be created prior to event query generation and event search, the metadata must be sufficiently rich such that it could be searched successfully for any event. The metadata generator is used to process 1) the search videos, 2) a common set of background training videos, and 3) each set of example event videos independently.
- 2) The event query generator uses the 1) event kit text, 2) background training video metadata, and 3) event video metadata to create an event query. The event query contains all information needed by the event search module. Participants are encouraged to make the event query human understandable so that, in a future system, a user could edit the event query to tune the event search.
- 3) The event search module searches the search video metadata for the event represented by the event query and creates an event confidence score and recounting for each search video. The event confidence score is a value between 0-1. The module also produces an event threshold score value, a score value above which the system identifies the video as containing the event.

NIST will provide all data inputs defining the event(s) to the MED system for testing and evaluation. **No other data is to be used as inputs to the system for evaluation or testing. This data is not allowed to be annotated or altered in any way, and the use of the data as inputs and outputs of the modules of the system must be strictly observed.** For example, the search video metadata store is not to be used as an input to the event query generator, only the event search.

Participants may use any computer system architecture to run the metadata generator. However, participants must use ONLY a COTS standard personal computing platform to execute the event query generator and the event search. The event searches are to be performed locally and the input and the output of the searches, i.e., the event query, the event confidence scores and the recounting, are to be submitted to NIST as specified below for official scoring and analysis. NIST will be creating a [MER Workstation](#) (see the Multimedia Event Recounting track of TRECVID) for evaluation, but participants are encouraged to design their own modules for research and development and to help NIST refine [the MER Workstation](#).

The MED evaluation is open to all that find the task of interest and who are willing to abide by the rules of the evaluation.

## 2 Data Resources

Internet multimedia data (i.e., video clips containing both audio and video) will be provided to registered MED participants. The data resources, collected by the Linguistic Data Consortium (LDC), consist of publically available, user-generated content. The LDC will be the distribution point for the collection. See <http://www.nist.gov/itl/iad/mig/med13.cfm> for data format, licensing, and acquisition instructions.

MED data resources will consist of 1) test and evaluation event kits, 2) evaluation search videos, 3) two test search video collections, and 4) research resources. These are described [in more detail](#) in the following subsections.

Jon Fiscus 6/27/13 9:07 AM  
**Deleted:** an event triage module

Jon Fiscus 6/27/13 9:07 AM  
**Deleted:** their event triage module

Jon Fiscus 6/27/13 9:07 AM  
**Deleted:** posted to the Internet video hosting sites.

Jon Fiscus 6/27/13 9:07 AM  
**Deleted:** 3

Jon Fiscus 6/27/13 9:07 AM  
**Deleted:** fully

## 2.1 Event Kits

There will be 40 events in MED13 divided into two sets, pre-specified (**PS13**) and ad-hoc (**AH13**), each containing 20 events. The event names of the PS events are listed in Table 1. The AH13 events will be revealed according to the TRECVID master schedule.

Pre-Specified Events for MED13	
From MED11	From MED12
E006 - Birthday party	E021 - Bike trick
E007 - Changing a vehicle tire	E022 - Cleaning an appliance
E008 - Flash mob gathering	E023 - Dog show
E009 - Getting a vehicle unstuck	E024 - Giving directions
E010 - Grooming an animal	E025 - Marriage proposal
E011 - Making a sandwich	E026 - Renovating a home
E012 - Parade	E027 - Rock climbing
E013 - Parkour	E028 - Town hall meeting
E014 - Repairing an appliance	E029 - Winning race without a vehicle
E015 - Working on a sewing project	E030 - Working on a metal crafts project

Table 1: MED13 Evaluation Events

The event kit text as a whole defines the event. The event kit text consists of:

- Event name: A mnemonic title for the event
- Event definition: A textual definition of the event.
- Event explication: An expression of some event domain-specific knowledge needed by humans to understand the event definition.
- Evidential description: A textual listing of some attributes that are often indicative of an event instance. The evidential description provides a notion of some potential types of visual, acoustic, and temporal evidences indicating the event's existence. This is not an exhaustive list nor is it to be interpreted as required evidence.

Events kits may also be accompanied by a set of event videos examples that illustrate some, but not all, possible realizations of the event. The evaluation will support event queries with a varying number of positive event video examples.

1. **0Ex**: No example video clips will accompany the event kits.
2. **10Ex**: 10 positive clips and up to 5 miss clips (i.e., non-positive clips) will accompany the event kits. This is the eventual goal of a MED system.
3. **100Ex**: 100 positive clips, and up to 50 miss clips will accompany the event kits.

To support training of an event, a common set of background training videos will be defined as input to the event query generator. Note that both the set of example event videos and the background training videos will be run through the metadata generator before the event query generation is run.

Participants are allowed to use the pre-specified event kits, event example videos, and the background training videos for testing their system development. Therefore, participants are not allowed to alter, amend, or annotate these data in any way. The relevance labels will be provided in "reference database" (defined in section 3.1) containing only markings for positive and miss clips.

## 2.2 Test Search Video Collections

Two test search videos are to be used for teams to test their MED performance and publish results on a common data. To provide a common set, regardless of new event videos to be added in the future,

Jon Fiscus 6/27/13 9:07 AM

Deleted: Party

Jon Fiscus 6/27/13 9:07 AM

Deleted: As a new convention, bolded text in the event kit indicates required conditions for the clip to be positive<sup>1</sup> for the event while non-bold text indicates optional conditions.

Jon Fiscus 6/27/13 9:07 AM

Deleted: a

Jon Fiscus 6/27/13 9:07 AM

Deleted: a

Jon Fiscus 6/27/13 9:07 AM

Deleted: example

Jon Fiscus 6/27/13 9:07 AM

Deleted: NIST expects to release additional, experimental event kit that are similar to the existing event kits but texts that provide more detail definitions and revised set of example videos. ... [1]

teams must form a search video set before running the event search.

The test search videos are structured into three components to make two collections:

1. MED Test Background Set (approximately 23,000 video clips)
2. Kindred Background Set (approximately 10,000 video clips), and
3. Event Positive Search Video Sets, one for each event, which are used with both the MED Test and Kindred Background sets.

To create the search set for testing, the frozen background video metadata will be combined with the frozen event search video's metadata and potentially re-indexed.

The metadata generator module should be run on each of these event sets independently. NIST will distribute sets of database files for the MED Test and Kindred Test collections that contain the merge of the background and positive search set videos.

In the system diagram, Figure 1, note that the search videos are used in the following way:

- Search video sets must be automatically processed to create the search video metadata which must be frozen (i.e., it may not be regenerated or augmented).
- The search videos are NOT input to the event query generation module and therefore cannot influence the event query generation.
- No training or tuning of the system parameters is permitted on the test or evaluation search videos. Therefore, the test search sets may not be annotated or altered in any fashion.

The teams should exercise the appropriate caution with the test search set to keep this independent of the research and training in order to provide a good rigorous test required for quality publications.

### 2.3 Evaluation Search Videos

Participants will be provided a set of evaluation search videos called the Progress Search Set. The Progress Search Set will be used for "blind" testing for MED12, MED13, MED14, and MED15 and therefore ground truth are not provided with this data.

Participants will have the choice of processing the full set of search videos or a smaller set:

1. **PROGSub:** Progress Evaluation Subset: a subset of the Progress Search Set consisting of approximately 32,000 videos
2. **PROGFull:** Progress Search Set: a set of approximately 98,000 videos

Participants must adhere to the same rules as outlined for the test data and the following additional rules for any evaluation search video sets:

- Participants must not attempt to gain knowledge of an evaluation search set properties or content by manually inspecting the video, clip metadata, output of the processing, or statistics developed during the processing.
- The evaluation search sets are only to be used as part of an official evaluation submission to NIST MED-MER evaluation.
- The evaluation search videos and their metadata stores must be deleted from systems (or made inaccessible) after each year's MED evaluation ends and re-copied and metadata generation applied the following year.

Jon Fiscus 6/27/13 9:07 AM

**Deleted:** that is

Jon Fiscus 6/27/13 9:07 AM

**Deleted:** Before running the event search on for any event, the teams will need to combine the background test metadata store (either HAVIC or Kindred) with event search video metadata.

Jon Fiscus 6/27/13 9:07 AM

**Deleted:**

Jon Fiscus 6/27/13 9:07 AM

**Deleted:** Participants must adhere to the same rules as outlined for the test data with the exception that teams will not need to combine the background with the event search videos for the test (these are already combined in the evaluation search se ... [2]

## 2.4 Research Resources

The goal of MED is to perform a general ad-hoc search for an event using few exemplars. Teams will conduct research to achieve this goal. To this end, NIST/LDC will provide some data for the purposes of research:

1. Research video data with ground truth,
2. Research event kits for E001-E005 and E016-E020.

Participants may use the research resources for research and development of their system software components. These resources may be altered, amended or annotated in any way participants need to facilitate their research. However, none of these resources may be annotated in a public forum, (e.g., Amazon Mechanical Turk) unless the data is shared with all participants in MED before used by the team collecting the annotations. Most importantly, none of the research resources or team added annotations may be used as direct inputs to the MED system.

Teams are also permitted to download their own research data from the Internet as long as they comply with all legal, licensing, and end-user agreements for the data. [Downloaded videos MUST NOT be used to extend the pool of event videos for search in any way.](#)

## 3 Evaluation Conditions

All evaluation conditions are options that the participants have in submitting results for evaluation. These options are:

1. Search Video Set (see Section 2.3)
2. Event Set (see Section 2.1)
3. Event Kit Type (see Section 2.1)

In addition, MED is also sponsoring several options for contrastive runs so that performers can evaluate their sub-systems on multimedia components: i.e.,

- Optical Character Recognition (**OCR**) only
- Automatic Speech Recognition (**ASR**) only
- Non-OCR visual (**Visual**) only
- Non-ASR audio (**Audio**) only

To minimally participate in MED, participants must create an encoded feature and concept based system<sup>2</sup>, process the minimal progress search set, provide results for the 20 pre-specified events provided in 2013, and used the extended (100EX) event kits. To fully participate in MED, teams must create all four types of component systems of their fully integrated system, process the full Progress Collection, and process both pre-specified and ad-hoc event kits with the 100EX, 10EX, and 0EX conditions.

<sup>2</sup> To the extent possible.

Jon Fiscus 6/27/13 9:07 AM

**Deleted:** Teams can download unlimited amounts of text, image, and sound files for their research. General videos may be downloaded and annotated. However, to maintain the eventual goal of ad-hoc event search, teams must not specifically search for or use event videos in order to not add to the event pool. Teams must delete any positive event videos from their research set unless it is provided by NIST/LDC. Team may retain any "miss" videos that are downloaded.

Jon Fiscus 6/27/13 9:07 AM

**Deleted:**

Jon Fiscus 6/27/13 9:07 AM

**Deleted:**

The following table indicates the required, optional and full conditions for MED:

Condition	Required	All Possible
<b>Subsystem Type</b>	FullSys	FullSys, OCRSys, ASRSys, Visualsys, AudioSys
<b>Search Video Set</b>	PROGSub or PROGFull	PROGFull
<b>Events Processed</b>	PS13	PS13, AH13
<b>Event Kit Videos</b>	100Ex	100Ex, 10Ex, 0Ex

Therefore, participants that are participating minimally in MED provide the MED run to the NIST T&E Team:

- FullSys-PROGSub-PS13-100Ex or
- FullSys-PROGFull-PS13-100Ex

Full participation would mean teams would submit 30 runs, i.e., for each system type (SysType = {FullSys, OCRSys, ASRSys, VisualSys, AudioSys}):

- SysType-PROGFull-PS13-100Ex
- SysType-PROGFull-PS13-10Ex
- SysType-PROGFull-PS13-0Ex
- SysType-PROGFull-AH13-100Ex
- SysType-PROGFull-AH13-10Ex
- SysType-PROGFull-AH13-0Ex

After minimal participation, teams are invited to do any of the full participation runs.

### 3.1 Event Training Inputs

Event training inputs will be specified through a set of three Comma Separated Value (CSV)<sup>3</sup> tables which will be provided (and named) for each evaluation condition (e.g., 100EX, 10Ex, and 0EX)<sup>4</sup>:

- The Event Database - \*\_EventDB.csv files  
This two-column table defines the **EventID** and **EventName**.
- The Clip Database - \*\_ClipMD.csv files  
This five-column table contains the metadata for each clip in the collection. The fields are **ClipID**, **MEDIA\_FILE**, **CODEC**, **MD5SUM**, **DURATION**.
- Judgment Database - \*\_JudgmentDB.csv  
This table specifies the exemplars to be associated with event kits. The three columns with data are ClipID, EventID, and INSTANCE\_TYPE (either "positive" or "near\_miss"). Note that 10Ex

<sup>3</sup> See Appendix C for the CSV file format specification.

<sup>4</sup> These tables are the authoritative sources for system inputs. Participants should avoid using directory listings of the collections as inputs because the tables are an experimental control mechanism.

Jon Fiscus 6/27/13 9:07 AM

Deleted: 24

Jon Fiscus 6/27/13 9:07 AM

Formatted: Underline

exemplars are also included in the 100Ex condition but the 100Ex videos cannot be included in the 10Ex condition.

### 3.2 System Outputs and Documentation

Teams are allowed to submit runs as specified above. For each submitted run to be considered complete and valid, participants must provide the information requested in this section. The requested information is a crucial element required for the research community to properly interpret the performance results.

#### 3.2.1 System Description

The purpose of the system description document is to provide a list of the resources and techniques used to build the MED system and identify the computing resources and time required to process the test set.

Each submission must include a system description which describes:

- the algorithms used for each of the modules,
- the hardware components used and computation times of the metadata generation, event query generation, and event search modules
- the metadata store size

See Appendix B section B.1 for the template that covers the *minimum* requirements of the system description document.

#### 3.2.2 Event Search Reporting

A MED system processes the metadata store detecting instances of each event independently. For each system submission, two output files must be created using Experiment Identifiers (**EXP-ID**) which describe the characteristics of the run (see Appendix B for the definition of *EXP-ID*):

1. <EXP-ID>.threshold.csv

Each line will contain information pertaining to the processing of a single event. Events not processed should not be included in the file. The 5 fields in this file are as follows.

- **EventID**: The Event ID processed - copied from the event database file.
- **DetectionThreshold**: A value of the system-supplied threshold.
- **DetectionTPT**: A value indicating the number of hours used during the *event search phase* for the event.
- **EAGTPT**: A value indicating the number of hours used during the *event query generation phase* for the event.
- **EMDTPT**: A value indicating the number of hours used to generate the metadata for the “Event Kit Videos”.
- **EBGMDTPT**: A value indicating the number of hours used to generate the metadata for the “Background Training Videos”. The values for this column may be identical if the clips processed are re-used across events.
- **SEARCHMDTPT**: A value indicating the number of hours used to generate the metadata for the “Search Videos”. The value **MUST** be identical for all events.

Jon Fiscus 6/27/13 9:07 AM

Deleted:

Jon Fiscus 6/27/13 9:07 AM

Deleted: 4

Jon Fiscus 6/27/13 9:07 AM

Deleted: indication

Jon Fiscus 6/27/13 9:07 AM

Deleted: event

Jon Fiscus 6/27/13 9:07 AM

Deleted: store

Jon Fiscus 6/27/13 9:07 AM

Deleted: event

2. <EXP-ID>.detection.csv

Each line will contain information pertaining to the processing of a single trial. The 2 fields in this file are as follows.

1. **TrialID**: The Trial ID processed - copied from the input trial index file.
2. **Score**: A probability value between 0 (low) and 1 (high) representing the system's confidence that the event is present in the clip.

#### 4 Evaluation Measures

System output will be evaluated by how well the system retrieves and detects MED events in evaluation search video metadata and by the computing resources used to do so. The determination of correct detection will be at the clip level, i.e. systems will provide a response for each clip in the evaluation search video set. Participants must process each event independently in order to ensure each event will be scored independently.

For each event, a MED system produces a threshold value and an event confidence score between 0.0 and 1.0 for each video in the search set. This score will be used to rank the videos, in descending order. Using ground truth, the rank of each true-positive event video will be computed by NIST forming a rank vector,  $\text{rank}(tp)$  where  $tp = 1$  to  $P_E$  and  $P_E$  is the total number of positives for the event. Videos with identical scores will be ranked randomly.

Precision and Recall can be computed for each position in the rank vector. Recall,  $\text{Recall}(tp)$  is the index of the rank vector  $tp$  divided by the total number of positive videos  $P_E$ . Precision,  $\text{Prec}(tp)$ , is the index of the rank vector,  $tp$ , divided by the rank of that positive,  $\text{rank}(tp)$ .

#### 4.1 Primary Measures

**Retrieval Metric: Mean Average Precision, MAP**

For a set of events (i.e., PS13 and AH13), the mean average precision (MAP) score will be computed as:

$$MAP = \frac{1}{Q} \sum_{q=1}^Q AP(q)$$

where Q denotes the number of different events and  $AP(q)$  is the average precision for event q.

$$AP(q) = \frac{1}{P_E} \sum_{tp=1}^{P_E} \text{Prec}(tp) = \frac{1}{P_E} \sum_{tp=1}^{P_E} \frac{tp}{\text{rank}(tp)}$$

where  $P_E$  denotes the number positives of event q.

**Detection Metric: Minimum Acceptable Recall,  $R_q$**

For a set of events (i.e., PS13 and AH13), the mean minimal acceptable recall,  $MR_q$  will be computed as:

Jon Fiscus 6/27/13 9:07 AM  
**Deleted:** 0.5 must represent equal confidence between event and non-event.

Jon Fiscus 6/27/13 9:07 AM  
**Deleted:** At a minimum, for...each trial, t, (... [3])

Jon Fiscus 6/27/13 9:07 AM  
**Deleted:** normalized ...ank vector. Recall, (... [4])

Jon Fiscus 6/27/13 9:07 AM  
**Deleted:** Normalized

Jon Fiscus 6/27/13 9:07 AM  
**Deleted:** collection size-normalized ...ean a (... [5])

Jon Fiscus 6/27/13 9:07 AM  
**Deleted:**  $MAP' = \frac{1}{Q} \sum_{q=1}^Q AP'(q)$  (... [6])

Jon Fiscus 6/27/13 9:07 AM  
**Formatted:** Font:+Theme Body

Jon Fiscus 6/27/13 9:07 AM  
**Deleted:**  $AP'$

Jon Fiscus 6/27/13 9:07 AM  
**Formatted:** Font:+Theme Body

Jon Fiscus 6/27/13 9:07 AM  
**Deleted:** collection size normalized

Jon Fiscus 6/27/13 9:07 AM  
**Formatted:** Font:+Theme Body

Jon Fiscus 6/27/13 9:07 AM  
**Formatted:** Left

Jon Fiscus 6/27/13 9:07 AM  
**Deleted:**  $AP' = \frac{1}{P_E} \sum_{tp=1}^{P_E} \gamma \frac{R_t}{\rho_t}$ ;  $\gamma = 0.001$  (... [7])

Jon Fiscus 6/27/13 9:07 AM  
**Deleted:** .

Jon Fiscus 6/27/13 9:07 AM  
**Deleted:**  $R_Q$

Jon Fiscus 6/27/13 9:07 AM  
**Deleted:** The detection metric that...will be (... [8])

Jon Fiscus 6/27/13 9:07 AM  
**Moved (insertion) [1]**



$$MR_0 = \frac{1}{Q} \sum_{q=1}^Q R_0(q)$$

where Q denotes the number of different events and  $R_0(q)$  is the minimal acceptable recall for event q and is computed as:

$$R_0(q) = \text{Recall}(T_q) - 12.5 * \frac{\text{rank}(T_q)}{V}$$

Where  $\text{Recall}(T_q)$  and  $\text{rank}(T_q)$  is the recall and rank at the threshold T for event q, and V is the total number of videos in the search set.

$$R_0 = R_T - m\rho_T; \quad m = 12.5$$

$R_0$  can be visualized in as the intercept with the percent rank = 0 axis of a line through the threshold point with a slope of 12.5. Teams will try to maximize their mean minimum acceptable recall score across all events of the evaluation, i.e.,

#### Metadata Generation Processing Speed

NIST will report the **real-time** factor to complete all steps necessary to build the metadata store.

*Real-time* factor is the Total Processing Time (TPT) for the process (as reported in the system description) divided by the number of hours of video in the test collection. TPT is the wall clock time (in hours) used during a computation phase, including I/O, from start to finish. The processing time for parallelized sub-steps adds to TPT as a single step. The processing time for metadata “shared” across sites, e.g., speech transcription, person tracking, etc. (including time used to incorporated data into the metadata store) adds to TPT as a single-step.

#### Event Query Generation Processing Speed

NIST will report the *real-time* factor for each event processed during the event query generation phase.

#### Event Search Processing Speed

NIST will report the *real-time* for each event processed during event search.

### 4.2 Diagnostic Measures

Diagnostic measures will be provided in software for participants to run on the test search videos. Diagnostic measures will not be provided as part of the evaluation to keep the content of the evaluation search videos blind. These diagnostic measures include:

#### Recall-Percent Rank Curves

Graphical performance assessment uses the recall as a function of the system’s percent rank. Percent rank is computed by dividing rank by the total number of videos in the search set, V.

### 4.3 Evaluation Tools and Command Line Example

Jon Fiscus 6/27/13 9:07 AM  
Formatted: Page Number

Jon Fiscus 6/27/13 9:07 AM  
Deleted: at system

Jon Fiscus 6/27/13 9:07 AM  
Formatted: Font:Not Italic

Jon Fiscus 6/27/13 9:07 AM  
Deleted: , R<sub>0</sub>, i.e.,

Jon Fiscus 6/27/13 9:07 AM  
Moved up [1]: -

Jon Fiscus 6/27/13 9:07 AM  
Deleted: -

NIST will use the Detection EVALuation (DEVA) tools within the NIST Framework for Detection Evaluation (F4DE) toolkit to score the evaluation submissions. The TRECVID MED11 Scoring Primer document (DEVA/doc/TRECVID-MED11-ScoringPrimer.html within the F4DE release [and is essentially identical to the tool usage for MED13](#)) contains instructions for how to use the scorer. NIST will use the following command line to evaluate MED systems.

```
% DEVA_cli\  
--profile MED13\  
--outdir <OUTPUTDIR> \  
--refcsv <DATA>_Ref.csv \  
--syscsv <EXP-ID>.detection.csv:detection \  
--syscsv <EXP-ID>.threshold.csv:threshold \  
<DATA>_TrialIndex.csv:TrialIndex \  
<DATA>_ClipMD.csv:ClipMD \  
<DATA>_JudgementMD.csv:JudgementMD \  
<DATA>_EventDB.csv:EventDB
```

Where,

- <OUTPUTDIR> is the pre-existing output directory
- <EXP-ID> is the Experimental ID defined in Appendix B
- <DATA> is the data set as defined in Appendix B

## 5 Result Submission Instructions

[See Appendix B](#)

## 6 Schedule

For TRECVID related schedule information please consult the main schedule on the TRECVID 2013 web site.

## 7 References

[None](#)

Jon Fiscus 6/27/13 9:07 AM

**Deleted:** \*\*\* THIS SECTION WILL BE UPDAT ... [9]

Jon Fiscus 6/27/13 9:07 AM

**Deleted:** MED11

Jon Fiscus 6/27/13 9:07 AM

**Deleted:** /\* We shouldn't give performers these files \*/

Jon Fiscus 6/27/13 9:07 AM

**Deleted:** \*\*\* THIS SECTION WILL BE UPDATED \*\*\*

Jon Fiscus 6/27/13 9:07 AM

**Deleted:** <#>Harold W. Kuhn, "The Hungarian Method for the assignment problem", *Naval Research Logistic Quarterly*, 2:83-97, 1955. .

## Appendix B: Submission Instructions

The packaging and file naming conventions for [MED13](#), relies on **Experiment Identifiers** (EXP-ID) to organize and identify the files for each evaluation condition and link the system inputs to system outputs. Since EXP-IDs may be used in multiple contexts, some fields contain default values. The following section describes the EXP-IDs.

The following EBNF describes the EXP-ID structure:

```
EXP-ID ::= <TEAM>_MED13_<SYS>_<SEARCH>_<EVENTSET>_<EKTYPE>_<VERSION>
where
```

```
<TEAM> ::= your Short TRECVID Team Name. "+" or "_" characters are not allowed.
<SYS> ::= either "FullSys", "OCRSys", "ASRSys", "VisualSys", "AudioSys"
<SEARCH> ::= either "MED13DRYRUN", "PROGSub" or "PROGAll"
<EVENTSET> ::= either "PS" or "AH" for Pre-Specified or Ad-Hoc tasks respectively as
defined in Section 2.1.
<EKTYPE> ::= either "100Ex", "10Ex", "0Ex"
<VERSION> ::= 1..n (with values greater than 1 indicating multiple runs of the same
experiment/system)
```

Note: there can be only one system submission per TEAM.

In order to facilitate transmission to NIST and subsequent scoring, submissions must be made using the following protocol, consisting of three steps: (1) preparing a system description, (2) packaging system outputs and system descriptions, and (3) transmitting the data to NIST.

### B.1 System Descriptions

Documenting each system is vital to interpreting evaluation results. As such, each submitted system, (determined by unique experiment identifiers), must be accompanied by a system description with the following information.

#### Section 1 Experiment Identifier(s)

List all the experiment IDs for which system outputs were submitted. Experiment IDs are described in further detail above.

#### Section 2 System Description

A brief technical description of your system.

#### Section 3 Metadata Generation System Hardware Description and Runtime Computation

Describe the hardware setup(s) to perform the metadata generation phase and report the Total Processing Time (TPT) as defined in Section 4.1. The phase may be broken down into sub-steps in which case the hardware and processing time of each sub-step must be documented in the system description. The TPT must also be placed in the system's \*.threshold.csv file per Section 3.2.2.

A hardware setup is the aggregate of all computational components used to perform this phase. Examples of a system might be: a 16-node, Dual Quad Core 2.26 GHz Intel Xeon, 24GB RAM per node, with a 10TB Data Server.

#### Section 4 Event Query Generation Hardware Description and Runtime Computation

Jon Fiscus 6/27/13 9:07 AM

Deleted: MED12

Jon Fiscus 6/27/13 9:07 AM

Deleted: Visualsys

Jon Fiscus 6/27/13 9:07 AM

Deleted: XXX.

Jon Fiscus 6/27/13 9:07 AM

Deleted: and number of CPU cores

Jon Fiscus 6/27/13 9:07 AM

Deleted: .

Jon Fiscus 6/27/13 9:07 AM

Deleted: , and CPU core counts

Jon Fiscus 6/27/13 9:07 AM

Deleted: .

Describes the computing hardware, including the number of CPU cores, used to generate the event queries. The Total Processing Time for each event [and the Background Training Videos](#) will be placed in the system's \*.threshold.csv file [per Section 3.2.2](#).

The hardware setup is the aggregate of all computational components used to perform this phase.

#### Section 5 *Event Search Hardware Description*

Describes the computing hardware, including the number of CPU cores, used for executing the event search(s). The Total Processing Time for each event will be placed in the system's \*.threshold.csv file [per Section 3.2.2](#).

The hardware setup is the aggregate of all computational components used to perform this phase. This hardware platform must be limited to a COTS standard personal computing platform.

#### Section 6 *Training data and knowledge sources*

Lists the resources used for system development, and runtime knowledge sources beyond the provided MED corpora.

#### Section 7 *References*

A list of pertinent references.

### B.2 Packaging Submissions

[All system output submissions must be formatted according to the following directory structure:](#)

- [output/<EXPID>/<EXPID>.txt](#)
- [output/<EXPID>/<EXPID>.detection.csv](#)
- [output/<EXPID>/<EXPID>.threshold.csv](#)

where,

- [EXPID is the experiment identifier as described in Section B.1,](#)
- [<EXPID>.txt is the system description file as specified above \(Section B.2\),](#)
- [<EXPID>.detection.csv is the CSV-formatted system output file containing the detection scores for each TrialID \(see Section 2.2.2\).](#)
- [<EXPID>.threshold.csv is the CSV-formatted system output file containing the detection thresholds and processing speed measurements \(see Section 2.2.2\)](#)

[Multiple EXPIDs may be present under the "output/" directory.](#)

### B.3 Validating the Submission

The F4DE distribution contains a submission checker that validates the submission both at a syntactic and semantic level. Participants should check their submission prior to sending it to NIST. NIST will reject submissions that do not pass validation. The TRECVID [MED13](#) Scoring Primer document (DEVA/doc/TRECVID-MED13-ScoringPrimer.html within the F4DE release) contains instructions for how to use the validator. NIST will use the following command line to validate MED submission files.

```
%TV13MED-SubmissionChecker --TrialIndex <DATA>_TrialIndex.csv \  
MED13_testTEAM_DRYRUN_2.tar.bz2
```

Jon Fiscus 6/27/13 9:07 AM  
**Deleted:** \*\*\* THIS SECTION WILL BE UPDATED \*\*\*

Jon Fiscus 6/27/13 9:07 AM  
**Deleted:** MED12

Jon Fiscus 6/27/13 9:07 AM  
**Deleted:** MED12

Jon Fiscus 6/27/13 9:07 AM  
**Deleted:** TV12MED

Jon Fiscus 6/27/13 9:07 AM  
**Deleted:** MED12

## B.4 Transmitting Submissions

To prepare your submission, first create the previously described file/directory structure. The following instructions assume that you are using the UNIX operating system. If you do not have access to UNIX utilities or ftp, please contact NIST to make alternate arrangements.

First, change directory to the parent directory of your “output/” directory. Next, type the following command:

```
tar -cvf - ./output | gzip > MED13 <TEAM> <SEARCH> <EVENTSET> <SUBNUM>.tgz
```

where,

- <TEAM>, <SEARCH>, and <EVENTSET> are the same as defined above.
- <SUBNUM> is an integer 1 to n, where 1 identifies your first submission, 2 your second, etc.

Important: only the latest submission will be used for scoring, but a submission file can contain multiple EXPIDs.

This command creates a single tar/gzip file containing all of your results. After shipment to NIST (as described in the next step), NIST will validate your submission with a syntactic and semantic validator.

Next, ftp to [jaguar.ncsl.nist.gov](http://jaguar.ncsl.nist.gov) giving the username 'anonymous' and (if requested) your e-mail address as the password. After you are logged in, issue the following set of commands, (the prompt will be 'ftp>'):

```
ftp> cd incoming
ftp> binary
ftp> epsv4 off
ftp> passive auto
ftp> put MED13 <TEAM> <SEARCH> <EVENTSET> <SUBNUM>.tgz
ftp> quit
```

Note: the “epsv4 off” is designed to bypass a limitation of the FTP server, and should return: “EPSV/EPRT on IPv4 off.”. If your prompt returns something different it is likely that your system does not support this feature, and therefore it is not needed.

Note: the “passive auto” should return “Passive mode: on; fallback to active mode: on”.

Note that because the “incoming” ftp directory (where you just ftp-ed your submission) is write protected, you will not be able to overwrite any existing file by the same name (you will get an error message if you try), and you will not be able to list the incoming directory (i.e., with the “ls” or “dir” commands). Please note whether you get any error messages from the ftp process when you execute the ftp commands stated above and report them to NIST.

The last thing you need to do is send an e-mail message to [med\\_poc@nist.gov](mailto:med_poc@nist.gov) to notify NIST of your submission. The following information should be included in your email:

- the name of your submission file,
- the file size,
- a listing of each of your submitted experiment IDs.

Please submit your files in time for us to deal with any transmission errors that might occur well before

the due date if possible. Note that submissions received after the stated due dates for any reason will be marked late.

## Appendix C: Comma Separated Value File Format Specifications

The MED evaluation infrastructure uses Comma Separated Value (CSV) formatted files with an initial field header line as the data interchange format for all textual data. The EBNF structure the infrastructure uses is as follows:

```
CSVFILE ::= <HEADER> <DATA>*  
  
    <HEADER>      ::=    <VALUE> {"," <VALUE> }* <NEWLINE>  
    <DATA>        ::=    <VALUE> {"," <VALUE> }* <NEWLINE>  
    <VALUE>       ::=    <DOUBLEQUOTE><TEXT_STRING><DOUBLEQUOTE>
```

The first data record in the files is a header line. The header lines are required by the evaluation infrastructure and the field names for the trial index file and the system output file are dictated by Sections 4.1 and 4.2.

Each header and data record in the table is one line of the text file. Each field value is delimited by double quotes and is separated from the next value with a comma.

An example trial index is (\*\_TrialIndex.csv):

```
"TrialID", "ClipID", "EventID"  
"72.P001", "72", "P001"  
"72.P002", "72", "P002"  
"72.P003", "72", "P003"  
"285.P001", "285", "P001"  
"285.P002", "285", "P002"  
"285.P003", "285", "P003"
```

An example system output file is (\*.detection.csv):

```
"TrialID", "Score"  
"72.P001", "0.062712"  
"72.P002", "0.978791"  
"72.P003", "0.115392"  
"285.P001", "0.801007"  
"285.P002", "0.861036"  
"285.P003", "0.120700"
```

An example threshold system output file is (\*.threshold.csv):

```
"EventID", "DetectionThreshold", "DetectionTPT"  
"P001", "0.54", "5923.3"  
"P002", "0.74", "9204.3"
```

Jon Fiscus 6/27/13 9:07 AM

Deleted: \*\*\* THIS SECTION WILL BE UPDA... [10]

Jon Fiscus 6/27/13 9:07 AM

Deleted: DectectionThreshold