

Multimedia Event Recounting (MER) Evaluation Overview

Greg Sanders , Barbara Shaw-Clemens, David Joy

NIST

For this evaluation, an event:

- 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 processes, activities, and human actions that are loosely or tightly organized and that have significant temporal and semantic relationships to the overarching activity;
- Is directly observable

The MER Task, *Summarized*

Given an event kit and a test video clip that contains an instance of the event, the MER system must generate a recounting (MER) of the event summarizing the key evidence for the event in the clip. The recounting must state

- **When:** Interval(s) of time (or frames) when the event occurred in the clip
- **Where:** Spatial location in the clip (pixel coordinate or bounding polygon)
- **What:** A clear, concise textual recounting of the observation(s)
 - Must include the key observations (video and audio) that led the system to detect the occurrence:
 - Scene, people, objects, relationships, and activities pertaining to the event
- **Confidence,** as in MED: for each observation or relationship

Big Picture of Evaluating MER

- In large part, the MER evaluation is intended to help developers, rather than as a summative evaluation
 - A bake-off with winners and losers is **NOT** an evaluation goal
- We evaluated the “what” (the textual recounting of the event)
 - Exactly what was in this textual recounting was up to the system developers
 - The DTD for the system output was designed to provide a place for anything a system might choose to include in its MER output
- The temporal “when” and spatial “where” were not scored for TRECVID
 - An actual user could use the temporal “when” and the spatial “where” to find (in the clip) what is recounted

The “events” in this evaluation

(a description of the clips that were chosen)

- **E022: Cleaning an appliance**
 - One or more people clean an appliance
- **E026: Renovating a home**
 - One or more people work to improve a home by rebuilding, finishing, and/or destroying parts of the structure
- **E027: Rock climbing**
 - One or more people climb up or across rock formations or artificial rock walls
- **E028: Town hall meeting**
 - An informal meeting . . . to hear discussion (and sometimes share opinions) on shared subjects of interest
- **E030: Metal crafts project**
 - One or more people fashion an object out of metal

The Three MER-related Datasets

- Each dataset consists of six clips for each of the five MER events (total of 30 clips)
 - **The Development Dataset:** to support research and a dry run
 - **The 2012 Evaluation Dataset:** a set of clips known to the developers, used for evaluation where the developer teams can compare the clips, their MER outputs, and the detailed scoring
 - These clips were not used for MED
 - **The Progress Dataset:** Selected from the MED pre-specified evaluation
 - These are true positive clips that all MED systems deemed positive
 - The 2012 MER systems generated MER output for these clips
 - Identity of these 30 clips will not be disclosed
 - Can be used to compare MER outputs in future years to MER outputs from the current year

The Two MER Assessments

We focused on how **informative** the MER outputs were, with two evaluation subtasks

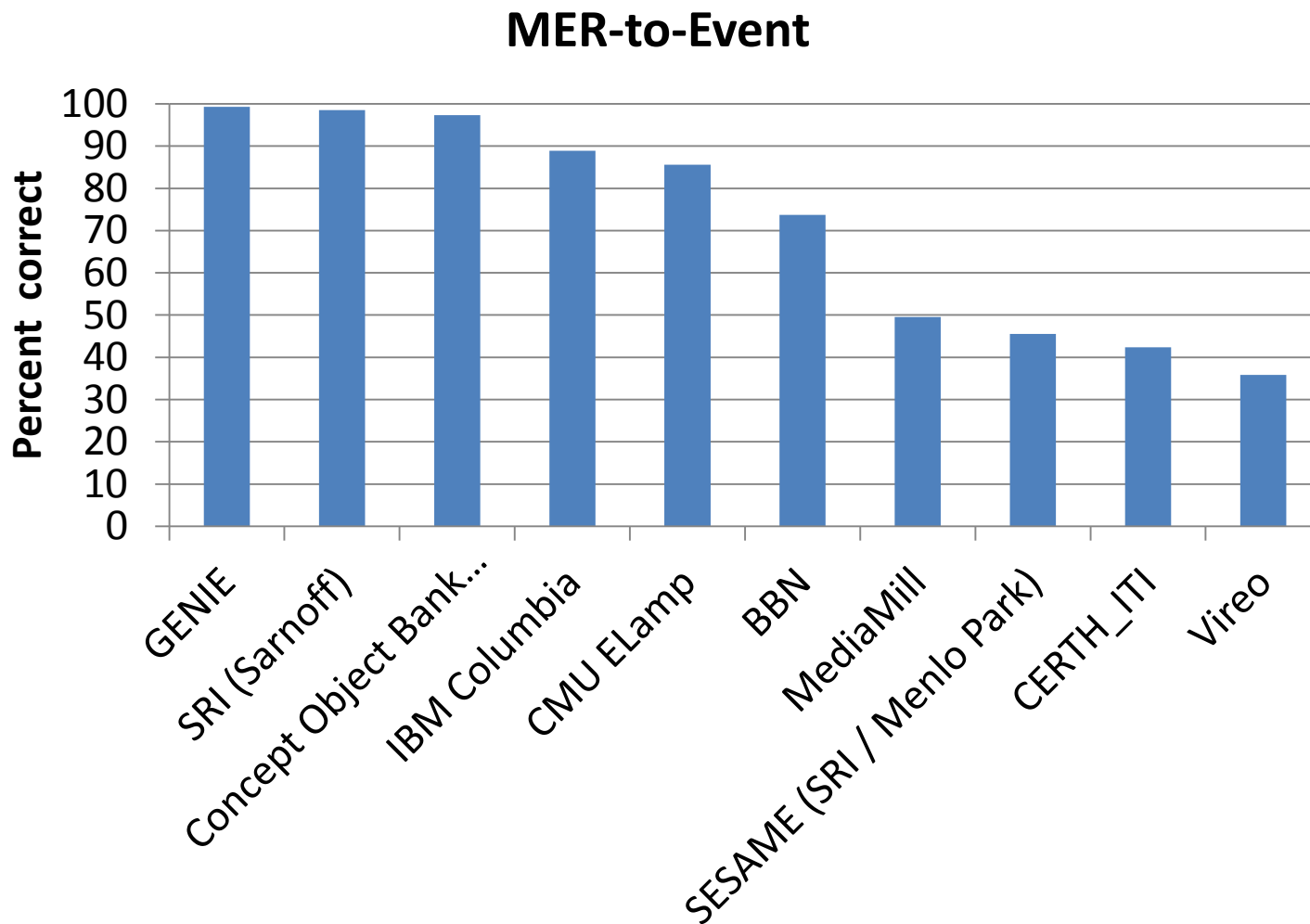
-- future years will use different assessments

MER-to-Event: Was it possible for a human judge to read a recounting and figure out which of the five events it describes?

MER-to-Clip: Was it possible for the judge to view the six clips for an event and figure out which MER output went with which clip?

Summary of MER-to-Event, by System

(% of judgments that got the event correct)

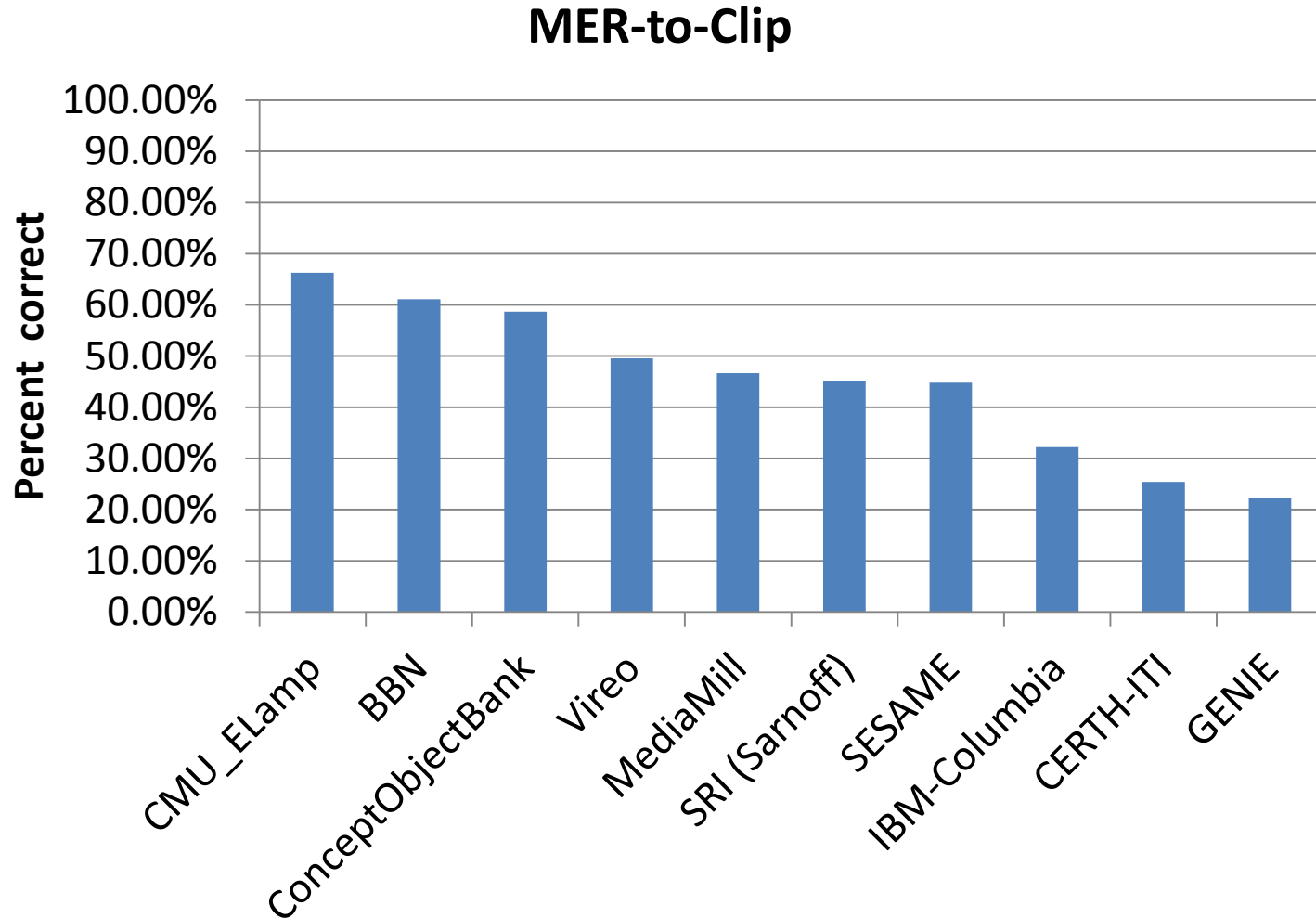


Summary of MER-to-Event, by Event

(% of judgments that got the event correct)

70.02% (median)	E022: Cleaning an appliance most commonly confused as E026 or E028
67.70%	E026: Renovating a home most commonly confused as E022 or E028
77.36%	E027: Rock climbing most commonly confused as E026 or E028
86.75%	E028: Town hall meeting most commonly confused as E026
60.38%	E030: Metal crafts project most commonly confused as E026 (very common) or as E022

Summary of MER-to-Clip, by System



Summary of MER-to-Clip, by Event

53.67% E022: Cleaning an appliance

40.99% E026: Renovating a home

49.27% E027: Rock climbing

36.23% E028: Town hall meeting

46.12% E030: Metal crafts project

(median)

Conclusions

- The intended use of MER systems:
 - Allow human user to identify clips of interest
 - Data for downstream automatic processes, such as IR
- This evaluation was not a direct match to intended use of MER outputs
 - MER-to-Event and MER-to-Clip are **not** the intended use
 - The evaluation subtasks were intended to give a fair evaluation of how informative the MER outputs are, across very different approaches to MER
- MER-to-Clip is much more difficult than MER-to-Event

Future Plans

- What to do for TRECVID-2013 is not yet decided
- We may well merge the MED and MER tasks
- We do not plan to repeat the 2012 evaluation
 - The evaluation was not a close match to the envisioned use of a system that does MED/MER
- We might investigate the ability of humans to do the MED task from the MER outputs, or other very different evaluations.
- It could be interesting to see whether MER could be performed by a system that works from the Event Kit exemplar clips but does not have any Event Kit text to refer to in deciding what evidence is salient or to incorporate into textual MER output.

Thank You