1

Known-item search
@
TRECVID 2012

Alan Smeaton

Dublin City University

Paul Over

NIST

Task

- **Use case:** You've seen a specific given video and want to find it again but don't know how to go directly to it. You remember some things about it. Its a natural, everyday scenario but you're not re-tracing history to re-find

- **System task:**
  - Given a test collection of short videos and a topic with:
    - some words and/or phrases describing the target video
    - a list of words and/or phrases indicating people, places, or things visible in the target video
  - Automatically return a list of up to 100 video IDs ranked according to the likelihood that the video is the target one,

    -- OR --

  - Interactively return a single video ID believed to be the target
    - Interactive runs could ask a web-based oracle if a video X is the target for topic Y. Simulates real user's ability to recognize the known-item. All oracle calls were logged. 24 topics in interactive KIS.

- Task is replicable, has low judging overhead and is appealing

Data

~ 291 hrs of Internet Archive available with a Creative Commons license

~8000 files

- Durations from 10s – 3.5 mins.
- Metadata available for most files (title, keywords, description, …)

813 development topics (initial sample topics, 2010 & 2011 test topics)

361 test topics created by NIST assessors, who …

- looked at a test video and tried to describe something unique about it;
- identified from the description some people, places, things, events visible in the video.

No video examples, no image examples, no audio; just a few words, phrases

Not YouTube in scale, but in nature. Its akin to a digital library

Example topics

- 891 1-5 KEY VISUAL CUES: geysers, bus, flags
- 891 QUERY: Find a video of yellow bus driving down winding road in front of building with flags on roof and driving past geysers

- 892 1-5 KEY VISUAL CUES: lake, trees, boats, buildings
- 892 QUERY: Find the video with panned scenes of a lake, tree-lined shoreline and dock with several boats and buildings in the background.

- 893 1-5 KEY VISUAL CUES: man, soccer ball, long hair, green jacket, parking lot, German
- 893 QUERY: Find the video of man speaking German with long hair and green jacket and soccer ball in a parking lot. [**NOT FOUND BY ANY RUN**]

- 894 1-5 KEY VISUAL CUES: Russian jet fighter, red star, white nose cone, sky rolls, burning airship
- 894 QUERY: Find the video of an advance Russian jet fighter with red star on wings and tail and a white nose cone that does rolls in the sky and depicts a burning airship

2012 Finishers

PicSOM +	Aalto University, Finland
AXES-DCU *	Access to Audiovisual Archives (EU-wide)
BUPT-MCPRL +	Beijing University of Posts & Telecom (MCPRL) China
ITI-CERTH *	Centre for Research and Technology Hellas, Greece
DCU-iAD-CLARITY *	Dublin City University, Ireland
KBVR +	KB Video Retrieval, US
ITEC_KLU * +	Klagenfurt University, Austria
NII * +	National Institute of Informatics, Japan
PKU_ICST * +	Peking Univ., Institute Computer Sc., China

* submitted interactive run(s) (6 groups)

Run conditions

Training type (TT):

A used only IACC training data

B used only non-IACC training data

C used both IACC and non-IACC TRECVID (S&V and/or Broadcast news) training data

D used both IACC and non-IACC non-TRECVID training data

Condition (C):

NO the run DID NOT use info (including the file name) from the IACC.1 *_meta.xml files

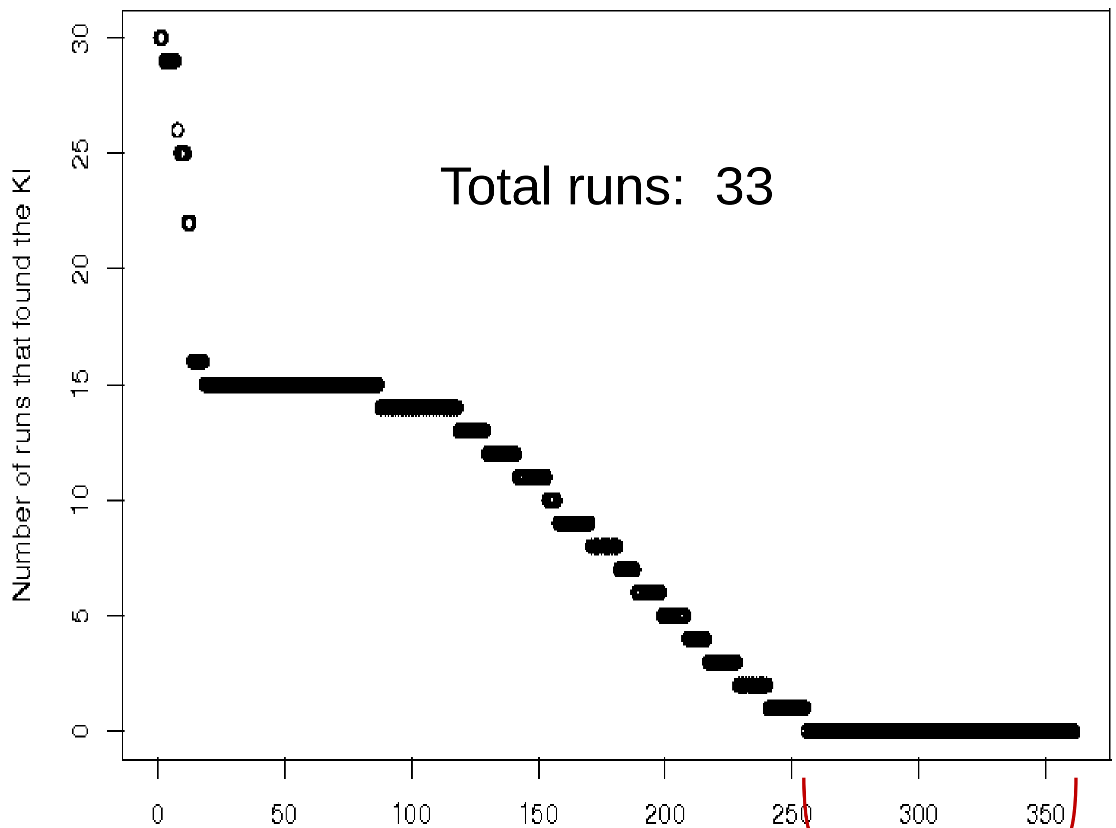
YES the run DID use info (including the file name) from the IACC.1 *_meta.xml files

Evaluation

- Three measures for each run across all topics (no NIST judging since we know the known item:
  - mean inverted rank of KI found (0 if not found)
    - for interactive (1 result per topic) == fraction of topics for which KI found
    - Calculated automatically using ground truth created with the topics
  - mean elapsed time (mins.)
  - user satisfaction (interactive) (1-7(best))

2012 Results – topic variability

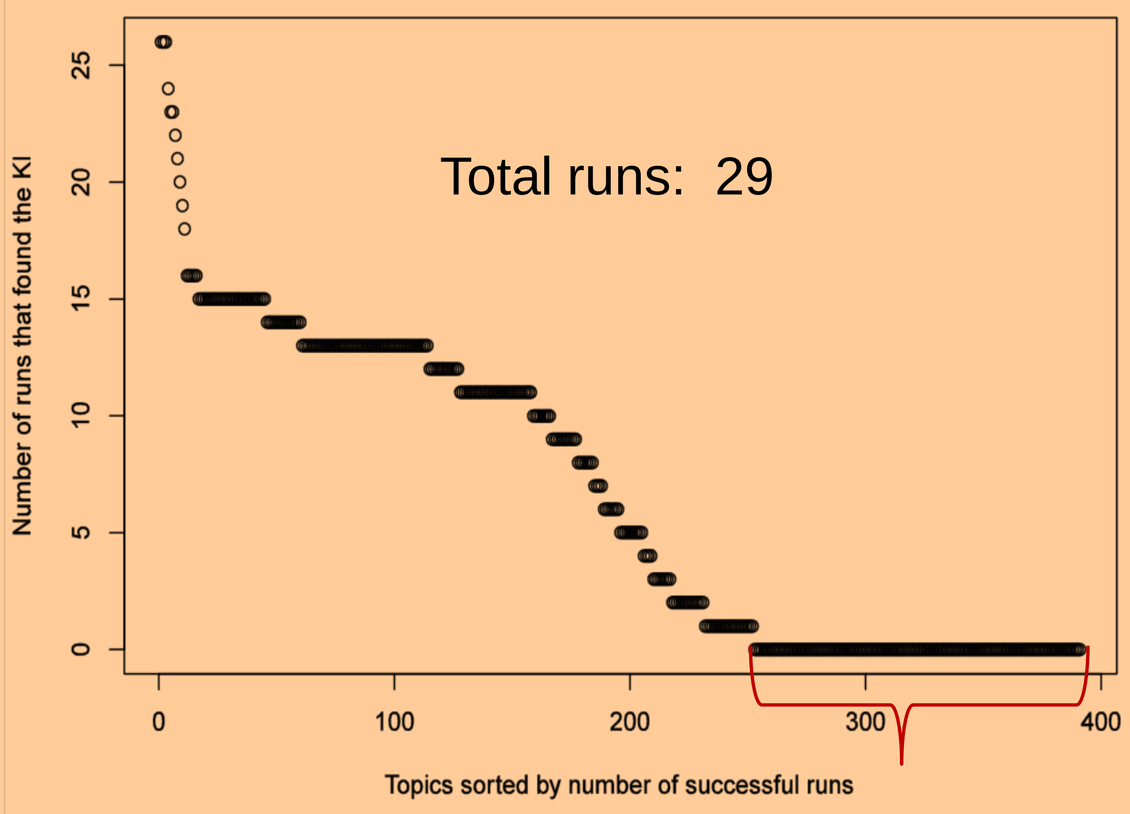
Topics sorted by number of runs that found the KI



e.g., 106 of 361 topics (29%) were never successfully answered

2011 Results – topic variability

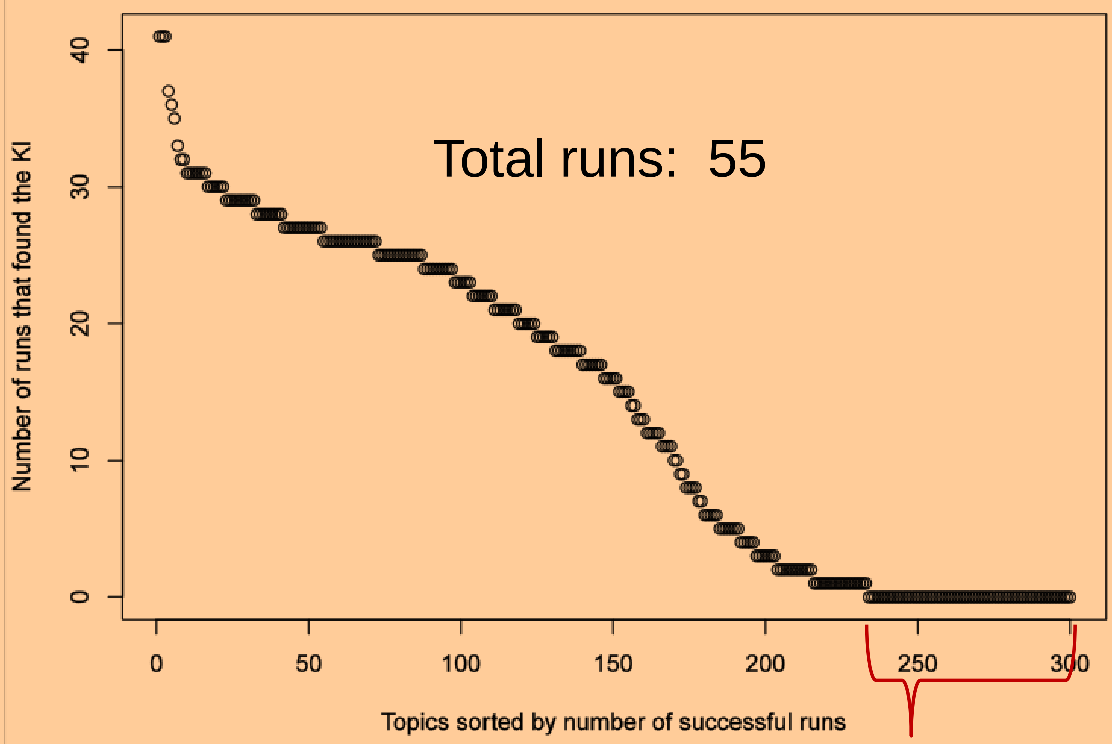
Topics sorted by number of runs that found the KI



e.g., 139 of 391 topics (35%) were never successfully answered

2010 Results – topic variability

Topics sorted by number of runs that found the KI



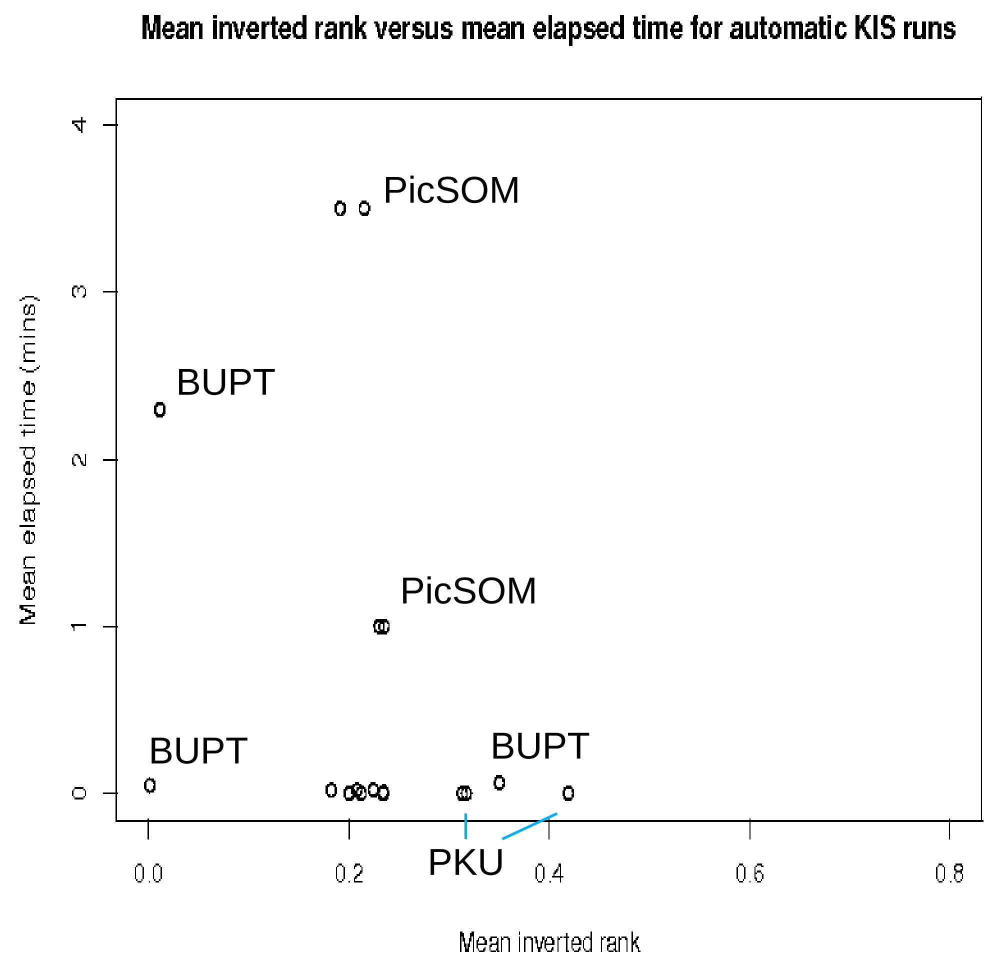
e.g., 67 of 300 topics (22%) were never successfully answered

Known items not found by any run

	Interactive		Automatic	
2012	2/24	17%	108/361	29%
2011	6/25	24%	142/391	36%
2010	5/24	21%	69/300	22%

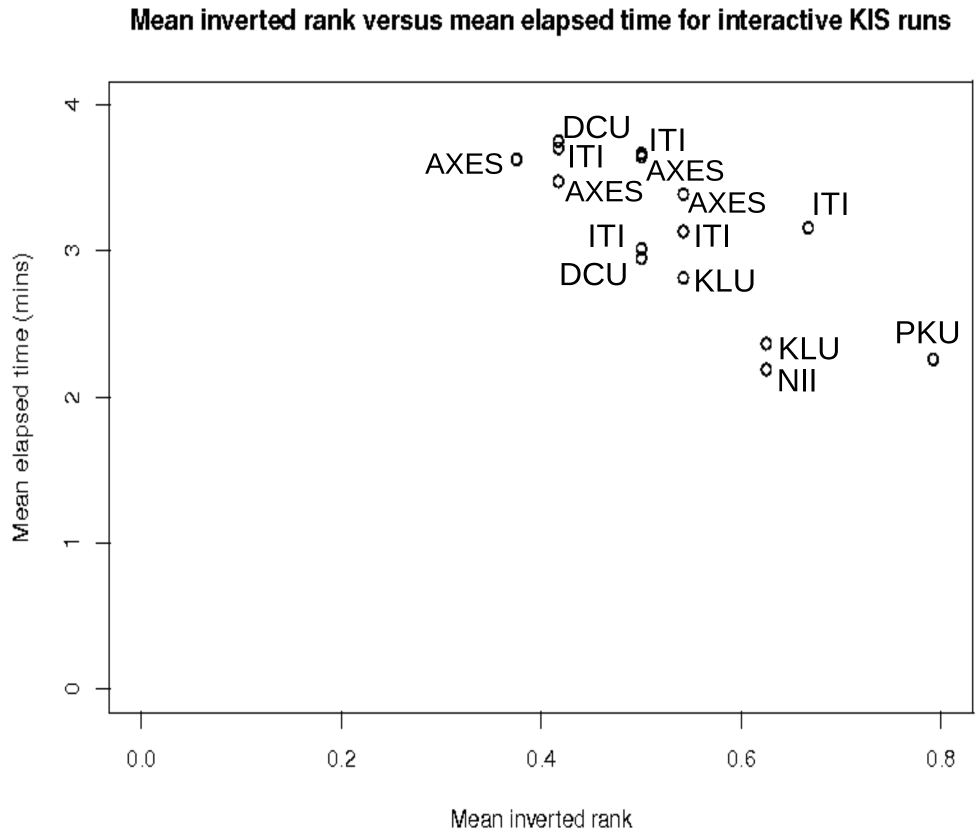
2012: Results – automatic runs

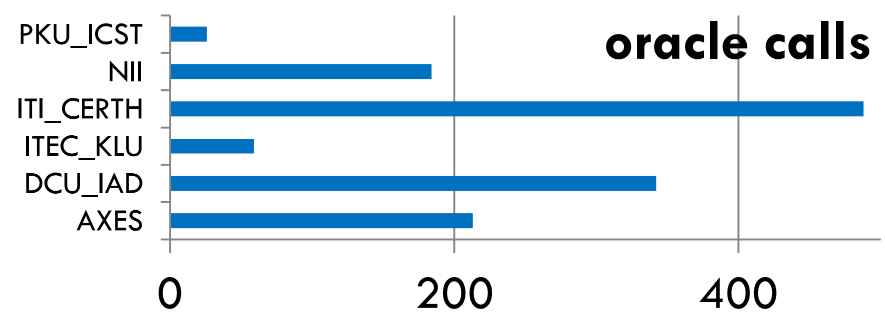
	Mean Time	Mean IR	Mean Sat
F_A_YES_PKU-ICST-MIPL_2	0.001	0.419	7.000
F_A_YES_MCPRBUPT4_4	0.065	0.350	3.000
F_A_YES_PKU-ICST-MIPL_3	0.001	0.317	7.000
F_A_YES_PKU-ICST-MIPL_4	0.001	0.313	7.000
F_A_YES_PicSOM_2_3	1.000	0.235	7.000
F_A_YES_ITEC_KLU_A2_2	0.009	0.234	5.000
F_A_YES_ITEC_KLU_A1_1	0.000	0.234	5.000
F_A_YES_PicSOM_1_4	1.000	0.230	7.000
F_D_YES_KBVR_1	0.021	0.224	5.000
F_D_YES_PicSOM_3_2	3.500	0.215	7.000
F_A_YES_NII1_1	0.001	0.212	5.000
F_D_YES_KBVR_3	0.020	0.208	5.000
F_A_YES_NII3_3	0.001	0.200	5.000
F_D_YES_PicSOM_4_1	3.500	0.191	7.000
F_D_YES_KBVR_2	0.020	0.182	5.000
F_A_NO_MCPRBUPT3_3	2.298	0.011	3.000
F_A_YES_MCPRBUPT2_2	0.049	0.001	3.000
F_A_YES_MCPRBUPT1_1	0.049	0.001	3.000



2012: Results – interactive runs

	Mean Time	Mean IR	Mean Sat
I_A_YES_PKU-ICST-MIPL_1	2.258	0.792	7.000
I_A_YES_ITI_CERTH_3	3.158	0.667	6.000
I_A_YES_NII4_4	2.188	0.625	5.000
I_A_YES_ITEC_KLU1_3	2.365	0.625	4.000
I_D_YES_AXES_1_1	3.388	0.542	7.000
I_A_YES_ITI_CERTH_1	3.133	0.542	6.000
I_A_YES_ITEC_KLU2_4	2.815	0.542	4.000
I_D_YES_AXES_2_2	3.645	0.500	7.000
I_A_YES_NII2_2	2.949	0.500	5.000
I_A_YES_ITI_CERTH_4	3.666	0.500	5.000
I_A_YES_DCU-iAd-Multi…_1	3.015	0.500	6.000
I_D_YES_AXES_3_3	3.476	0.417	7.000
I_A_YES_ITI_CERTH_2	3.703	0.417	5.000
I_A_YES_DCU-iAD-Single…2	3.752	0.417	6.000
I_D_YES_AXES_4_4	3.626	0.375	7.000





Personal overview of finishers

- All 9 participating groups each described their work in workshop notebook papers
- More detail in their posters and demos
- But here is my take on what each did …

AXES – a European Union FP7 project

- Built on previous participation in 2011
  - On-the-fly, query-time training of concept classifiers using external (Google Images) +ve examples from searchers’ text input
  - Also used text metadata
  - Face processing (2.9M face detections in KIS data)
- Score-based fusion, built on 2011 submission with focus on integrating multiple search services

BUPT-MCPRL

- Two approaches:
  - Traditional text-based, focus on colours, language, places, sound, synonym terms and correlations in an ontology, lead to 2nd highest MIR
  - Bio-inspired method, improves on the TV2011 submission, a bottom-up attention model for salient regions in image

Bio-inspired applied to only 37/361 topics and when used it was great but overall not great

Needs to determine when to use it, automatically

Some submission format issues so some results deflated

DCU-iAD-CLARITY

- Built on previous participation in 2011, 2010
- iPad application in "lean-back" interaction
- Two versions, using one KF representation, and using multiple KFs, per video
- 8 novice users in Latin squares experiment
- Multiple KF out-performs single KF by 1 minute in elapsed time, and also in MIR

ITEC – Klagenfurt Univ.

- Automatic and interactive submissions
- Used concepts from SIN task and heuristic voting
- Relied completely on text-based retrieval
- Rule-based query expansion and query reduction
- Interactive was based on applying filters (e.g. colours, language, music, etc.) to narrow down results of automatic so no relevance feedback or iterations (2 users)

ITI-CERTH

- Focus on was of interface interaction with the VERGE system which integrates
  1. Visual similarity search
  2. Transcription (ASR) search
  3. Metadata search
  4. Aspect models and semantic relatedness of metadata
  5. Semantic concepts (from SIN task)
- More interestingly they compared shot-based and video-based representations of content, finding video-based is substantially better (MIR and time)

KBVR

- Automatic submissions – 3 of them

1. BM25 on ASR and metadata
2. As above but with concept expansion using LSCOM
3. As in 1 but with concept expansion from Wikipedia

Neither 2 or 3 found any improvement because too many concepts drawn in, too much noise, semantic drift.

National Institute of Informatics

- Automatic and Interactive runs submitted
- Automatic used metadata, plus Google Translate (automatic) for language-specific topics
- Results show translation dis-improves but this could be due to the over-aggressive pre-processing
- In interactive, each video is represented as 5 KFs

PicSOM (Aalto University)

- Automatic runs. Baseline was text search of metadata
- Then layered on OCR of all keyframes in collection, giving a small improvement
- They layered on ASR with GNU Aspell spelling correction, not beneficial
- Google Image Search API to locate images visually similar to visual cues from search, reduced performance

Peking University

- Automatic and interactive KIS, top-ranked
- Text is processed by spell correction (Aspell), POS tagging (Stanford parser) to weight POS differently, and OCR on video frames, followed by topic term weighting and inflectional normalisation from a dictionary.
- B&W detection also included, as is detection and filtering of the video language (French, German, etc,)

Questions for participants

- We leave behind a public collection plus nearly 1,200 KIS topics with 117 official submissions

- Did any groups run their 2012 system on earlier test data or earlier systems on later data to separate data effect from system, see system progress?
- Any evidence use of metadata as crucial as in 2010 and 2011?

Questions for participants

Why the large(r) number of topics unanswered by all systems?

- 2010: 67 of 300 (22 %)
- 2011: 139 of 391 (35 %)
- 2012: 106 of 361 (29 %)

Were topics more difficult, what makes a topic more/less difficult ?

Is it the phrasing used in the topic ?

Is it the nature of the topic … object, activity, scene ?

Is it the nature of the target video … what is more memorable ?

Example topics

- 891 1-5 KEY VISUAL CUES: geysers, bus, flags
- 891 QUERY: Find a video of yellow bus driving down winding road in front of building with flags on roof and driving past geysers

- 892 1-5 KEY VISUAL CUES: lake, trees, boats, buildings
- 892 QUERY: Find the video with panned scenes of a lake, tree-lined shoreline and dock with several boats and buildings in the background.

- 893 1-5 KEY VISUAL CUES: man, soccer ball, long hair, green jacket, parking lot, German
- 893 QUERY: Find the video of man speaking German with long hair and green jacket and soccer ball in a parking lot. [**NOT FOUND BY ANY RUN**]

- 894 1-5 KEY VISUAL CUES: Russian jet fighter, red star, white nose cone, sky rolls, burning airship
- 894 QUERY: Find the video of an advance Russian jet fighter with red star on wings and tail and a white nose cone that does rolls in the sky and depicts a burning airship

Topic 892 (2 min 56s)

Thumbnails for Brockville Waterfront 17 June 2007 Return to Program Details

Below are images for every 30 seconds in the program.

Example topics

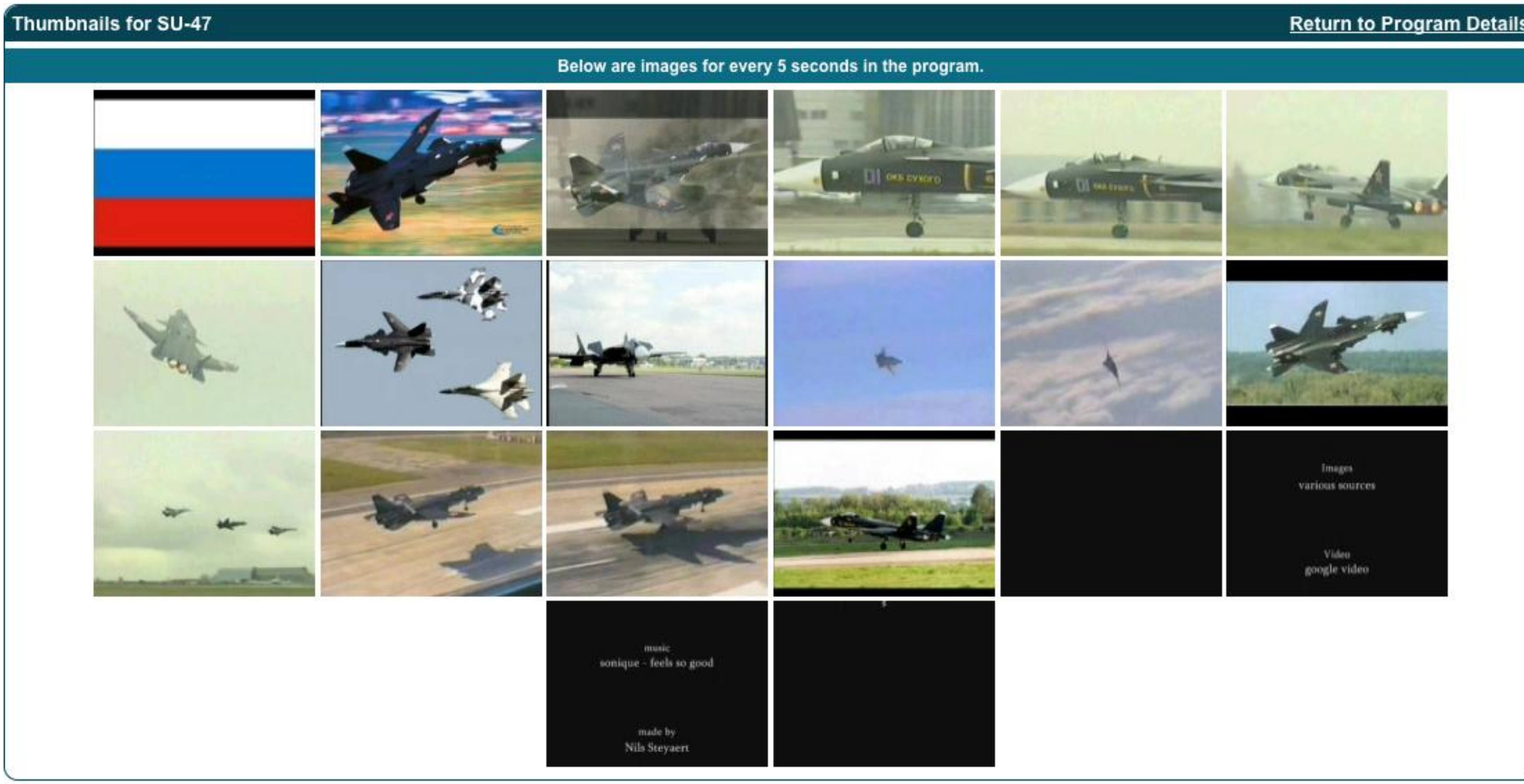
- 891 1-5 KEY VISUAL CUES: geysers, bus, flags
- 891 QUERY: Find a video of yellow bus driving down winding road in front of building with flags on roof and driving past geysers

- 892 1-5 KEY VISUAL CUES: lake, trees, boats, buildings
- 892 QUERY: Find the video with panned scenes of a lake, tree-lined shoreline and dock with several boats and buildings in the background.

- 893 1-5 KEY VISUAL CUES: man, soccer ball, long hair, green jacket, parking lot, German
- 893 QUERY: Find the video of man speaking German with long hair and green jacket and soccer ball in a parking lot. [**NOT FOUND BY ANY RUN**]

- 894 1-5 KEY VISUAL CUES: Russian jet fighter, red star, white nose cone, sky rolls, burning airship
- 894 QUERY: Find the video of an advance Russian jet fighter with red star on wings and tail and a white nose cone that does rolls in the sky and depicts a burning airship

Topic 894 (1 min 35s)



Questions for participants

Why the large(r) number of topics unanswered by all systems?

- 2010: 67 of 300 (22 %)
- 2011: 139 of 391 (35 %)
- 2012: 106 of 361 (29 %)

Were topics more difficult, what makes a topic more/less difficult ?

Is it the phrasing used in the topic ?

Is it the nature of the topic … object, activity, scene ?

Is it the nature of the target video … what is more memorable ?

Questions – Comments – Discussion