TRECVID-2009: Search Task

Alan Smeaton
CLARITY, Dublin City University
&
Paul Over
NIST

Search Task Definition

- Task: given a video test collection, a topic, and a common shot boundary reference
  - Normal: return a ranked list of at most 1,000 shots which best satisfy the need
  - High-Precision: return a ranked list of at most 10 shots which best satisfy the need
- Test and training videos were viewed by NIST personnel, notes taken on content, topic candidates chosen, examples added from development set and Web … same as has been done in previous years

Search Task Measures

- Per-search measures: average precision (AP), elapsed time
- Per-run measure: mean average precision
- Interactive search participants were asked to have their subjects complete pre, post-topic and post-search questionnaires;
- Each result for a topic can come from only 1 user search; same searcher does not need to be used for all topics.

2009 data (same source as 2007, 08)

- Educational, cultural, youth-oriented programming, news magazines, historical footage, etc.
- Primarily in Dutch, but also some English, etc.
- Much less repetition
  - No commercials
  - No repeated stock TV news footage
  - Greater variety of subject matter than in broadcast TV news
- Greater volume of data

2009: Search task finishers

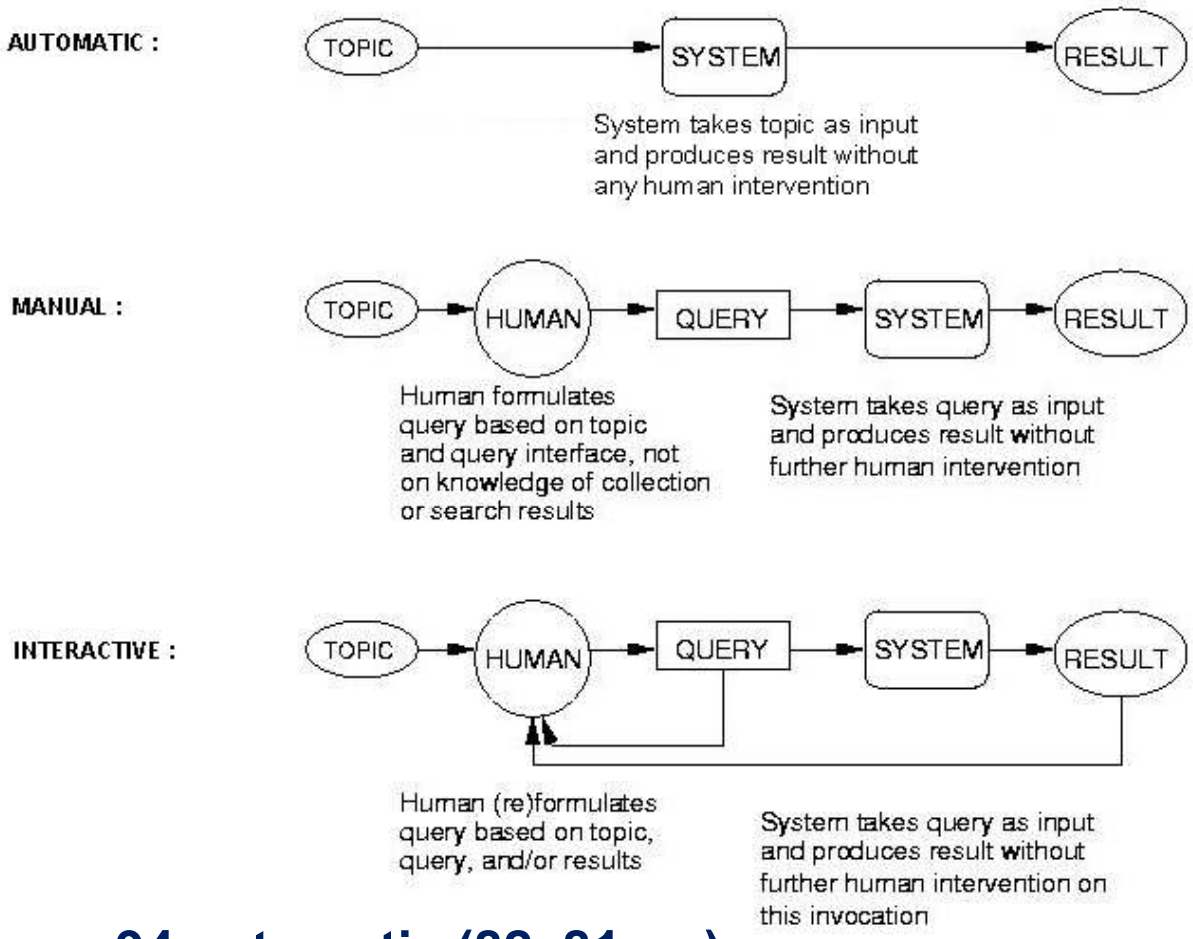
Aristotle University of Thessaloniki	--	FE	SE	-
Beijing University of Posts and Telecom.-BUPT-MCPRL	ED	FE	SE	CD
Beijing University of Posts and Telecom.-PRIS	ED	**	SE	-
Brno University of Technology	ED	FE	SE	**
Budapest Academy of Sciences	--	**	SE	**
Centre for Research and Technology Hellas	--	FE	SE	-
Chinese Academy of Sciences-MCG-ICT-CAS	--	--	SE	CD
City University of Hong Kong	ED	FE	SE	CD
Helsinki University of Technology TKK	--	FE	SE	-
KB Video Retrieval	--	--	SE	-
Kobe University (*)	--	**	SE	-
Laboratoire REGIM	ED	FE	SE	-
National Institute of Informatics	ED	FE	SE	CD
Peking University-PKU-ICST	ED	FE	SE	**
The Open University	--	**	SE	-
University of Amsterdam (*)	ED	FE	SE	-
University of Glasgow	--	**	SE	**
University of Surrey	--	--	SE	-
Zhejiang University	--	FE	SE	--

- 19 participants from 48 who applied, and most are renewals on 2008
- 30 finished in 2008, 24 in 2007
- What does this say ? Teaming rules ?

** : group applied, but did not submit a run

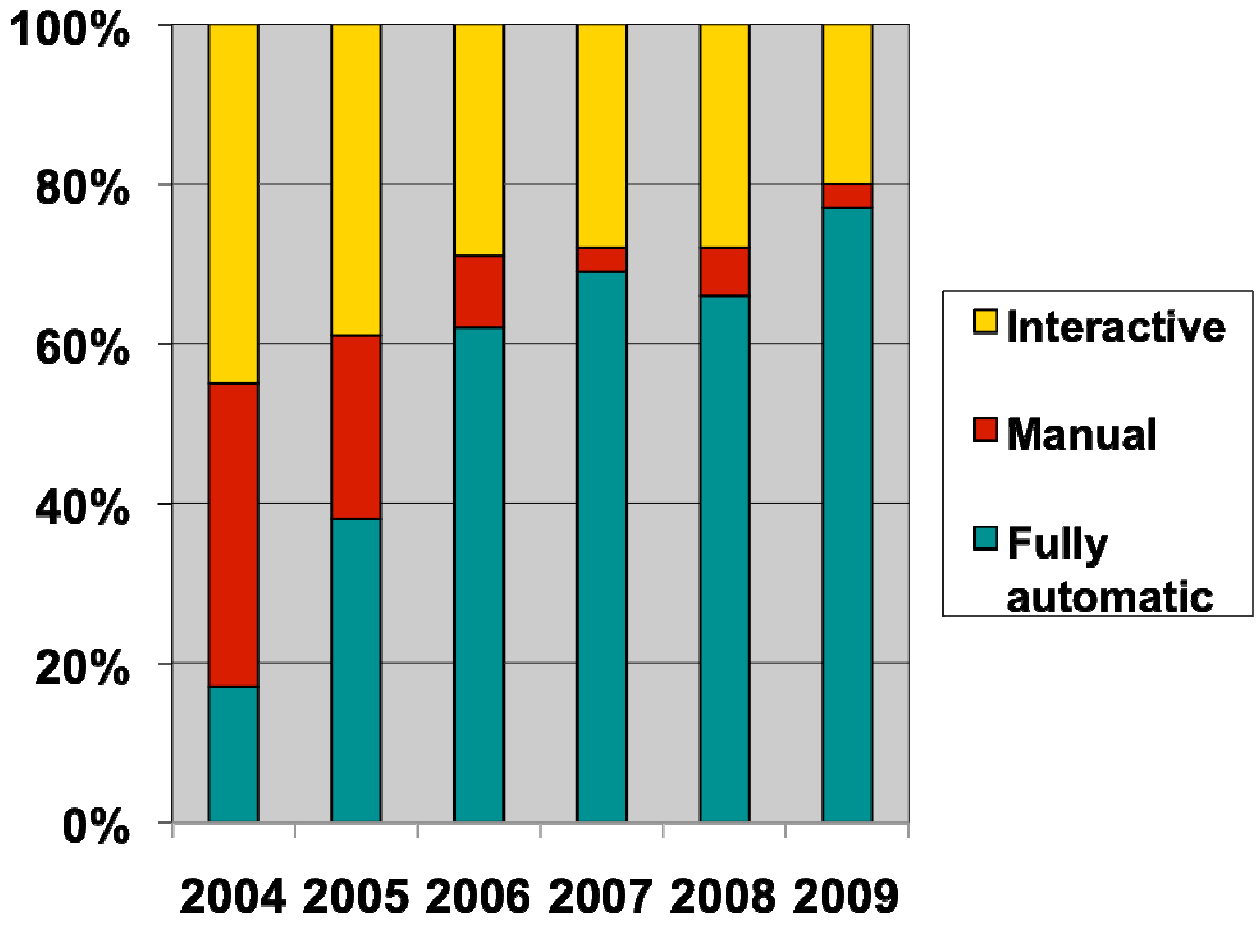
-- : group didn't apply for the task

Search Types: Automatic, Manual and Interactive



Number of runs: **94 automatic (82, 81, …)**
3 manually assisted (8, 4, ..)
24 interactive (34, 33, …)

Automatic growing; interactive shrinking some



24 Topics

269) Find shots of a road taken from a moving vehicle through the front window.

270) Find shots of a crowd of people, outdoors, filling more than half of the frame area.

271) Find shots with a view of one or more tall buildings (more than 4 stories) and the top story visible.

272) Find shots of a person talking on a telephone.

273) Find shots of a close-up of a hand, writing, drawing, coloring, or painting.

274) Find shots of exactly two people sitting at a table.

275) Find shots of one or more people, each walking up one or more steps.

276) Find shots of one or more dogs, walking, running, or jumping.

277) Find shots of a person talking behind a microphone.

278) Find shots of a building entrance.

279) Find shots of people shaking hands.

280) Find shots of a microscope.

281) Find shots of two more people, each singing and/or playing a musical instrument.

282) Find shots of a person pointing.

283) Find shots of a person playing a piano.

284) Find shots of a street scene at night.

285) Find shots of printed, typed, or handwritten text, filling more than half of the frame area.

286) Find shots of something burning with flames visible.

287) Find shots of one or more people, each at a table or desk with a computer visible.

288) Find shots of an airplane or helicopter on the ground, seen from outside.

289) Find shots of one or more people, each sitting in a chair, talking.

290) Find shots of one or more ships or boats, in the water.

291) Find shots of a train in motion, seen from outside.

292) Find shots with the camera zooming in on a person's face.

Frequency of target topic-shots

- **2009**
  - **Test shots * 24 topics: 2,253,648**
  - **Relevant topic-shots: 10,619 0.47%**
- 2008
  - Test shots * 48 topics: 1,618,848
  - Relevant topic-shots: 7,333 0.45%
- 2007
  - Test shots * topics : 435 408
  - Relevant topic-shots: 4,704 1.1%
- 2006:
  - Test shots * topics: 1,907,616
  - Relevant topic-shots: 7,225 0.4%
- 2005
  - Test shots * topics: 1,098,360
  - Relevant topic-shots: 8,395 0.8%
- 2004
  - Test shots * topics: 800,808
  - Relevant topic-shots: 1,800 0.2%

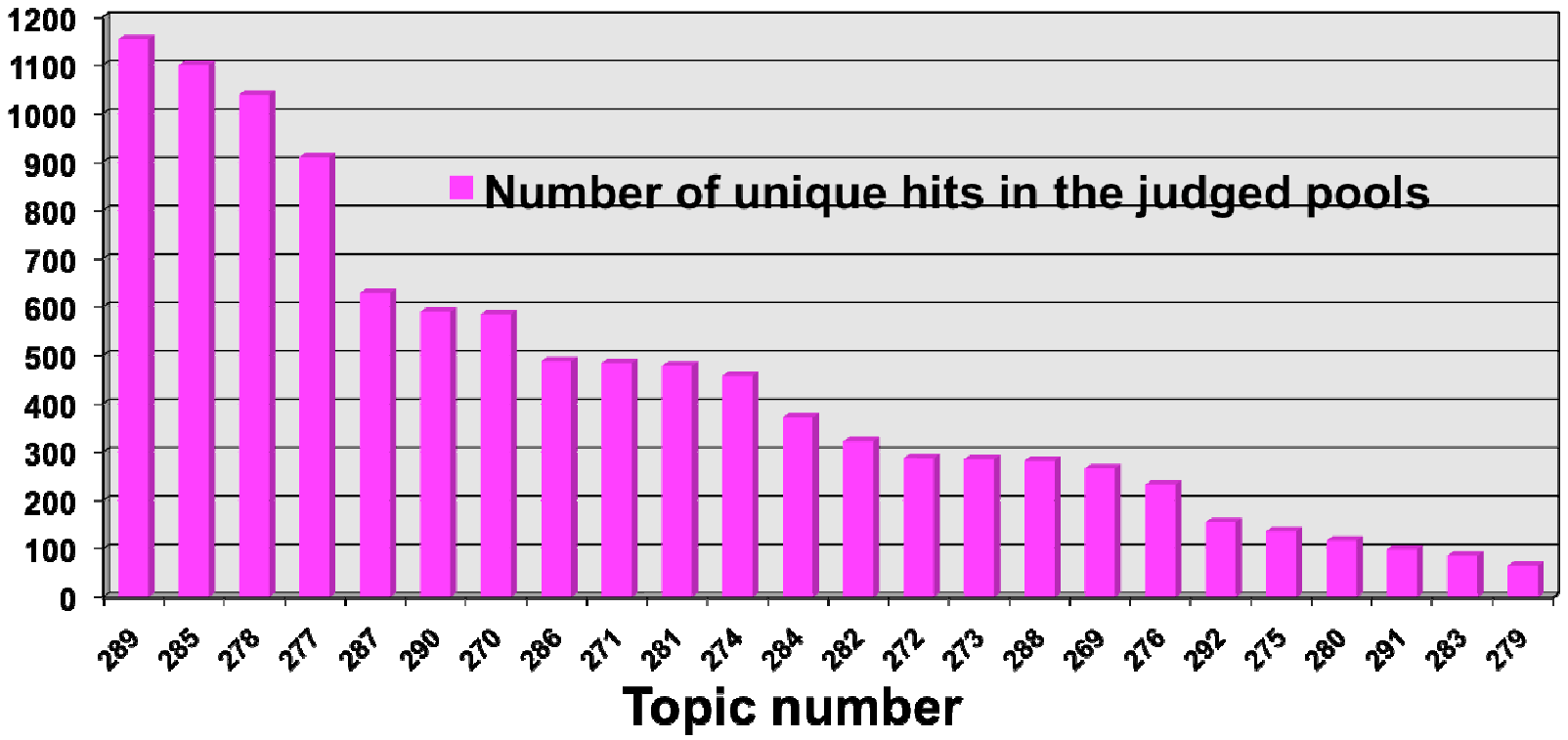
Topics similar in nature

Sound and Vision cycle

Growth in collection size

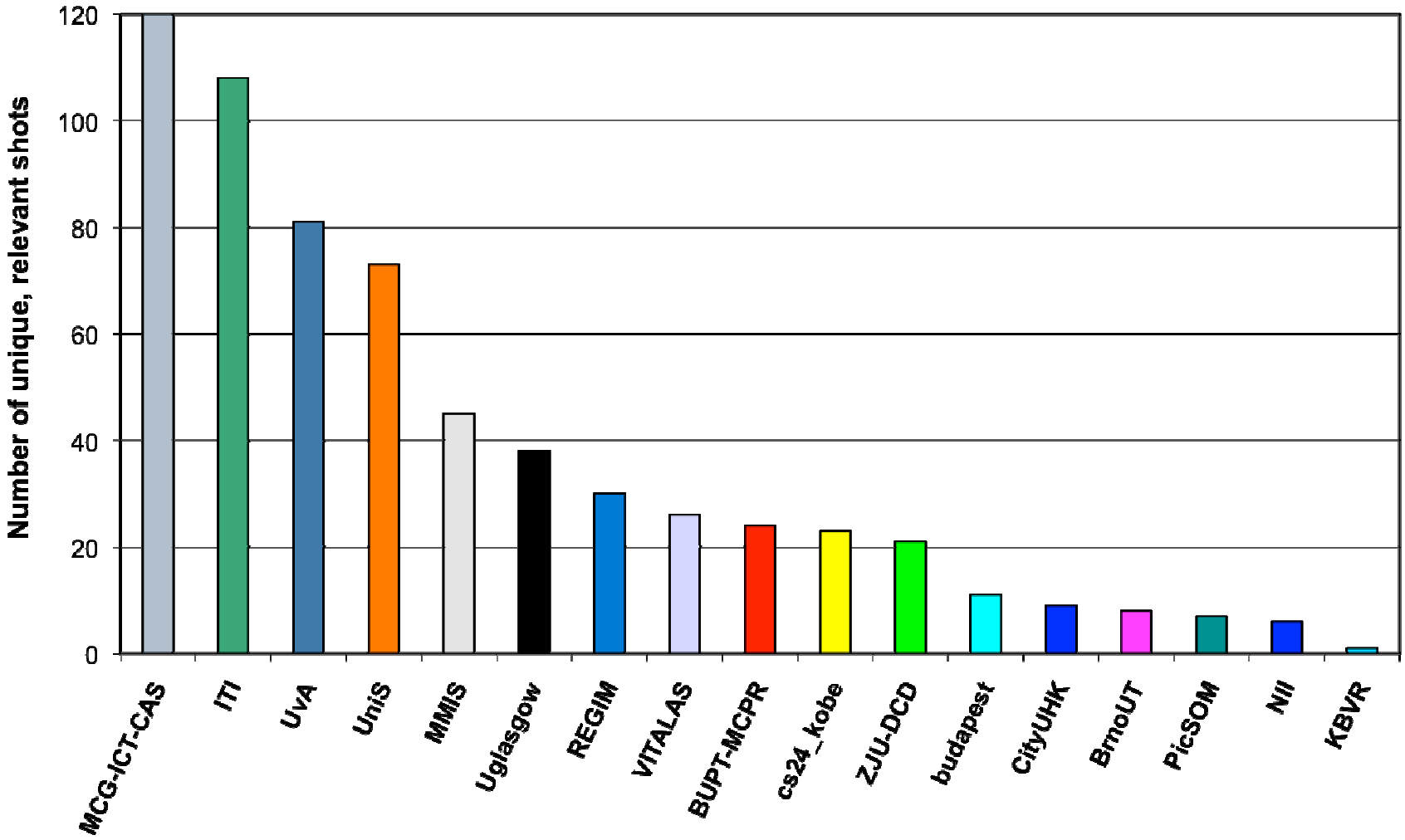
Broadcast news cycle

Distribution of (relevant) hits for each topic



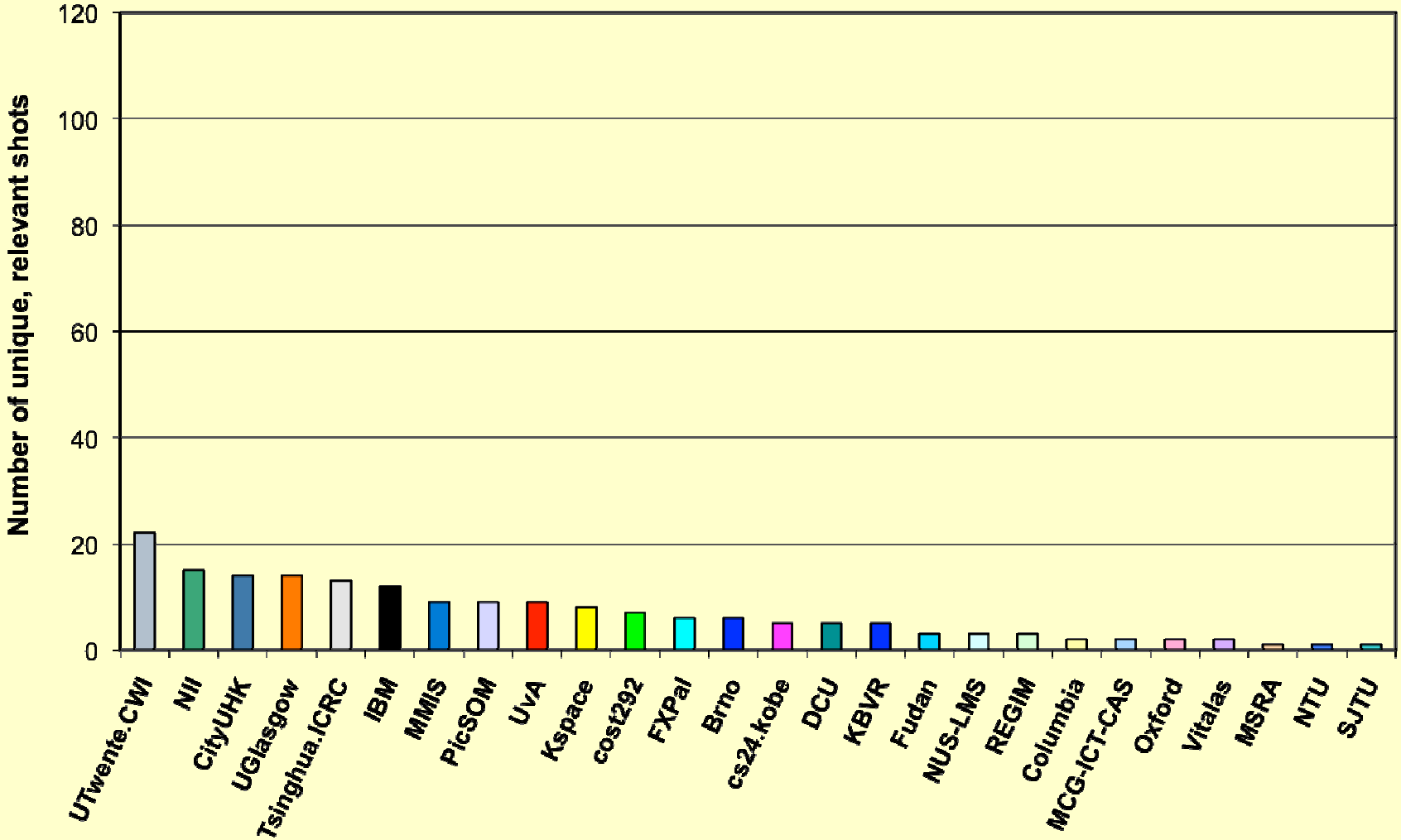
Much more than previous years, collection size ?

More unique, relevant shots found by some groups

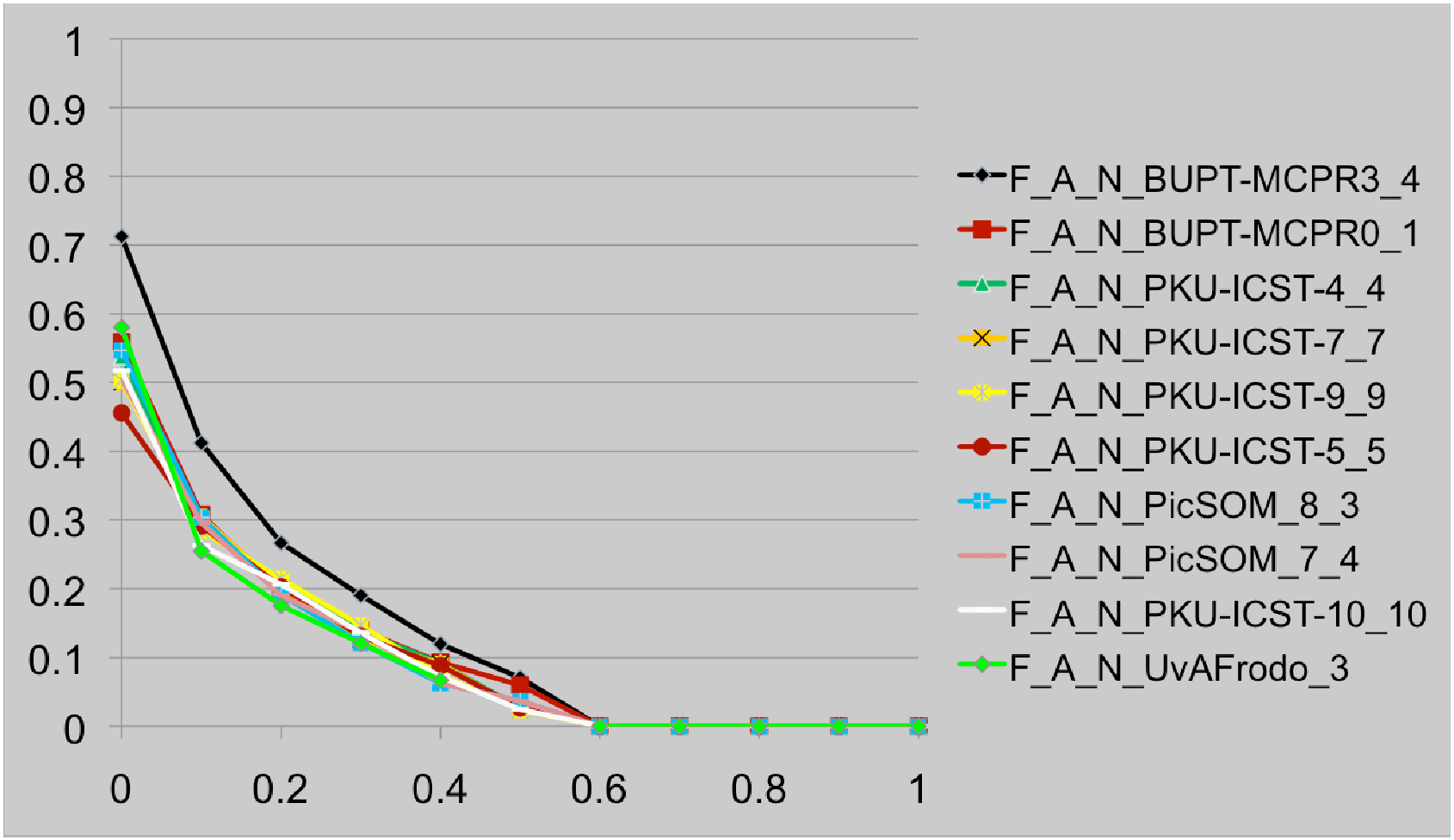


Can't be fewer runs … 122 vs.124, must be collection size

2008 Relatively few unique, relevant shots by group



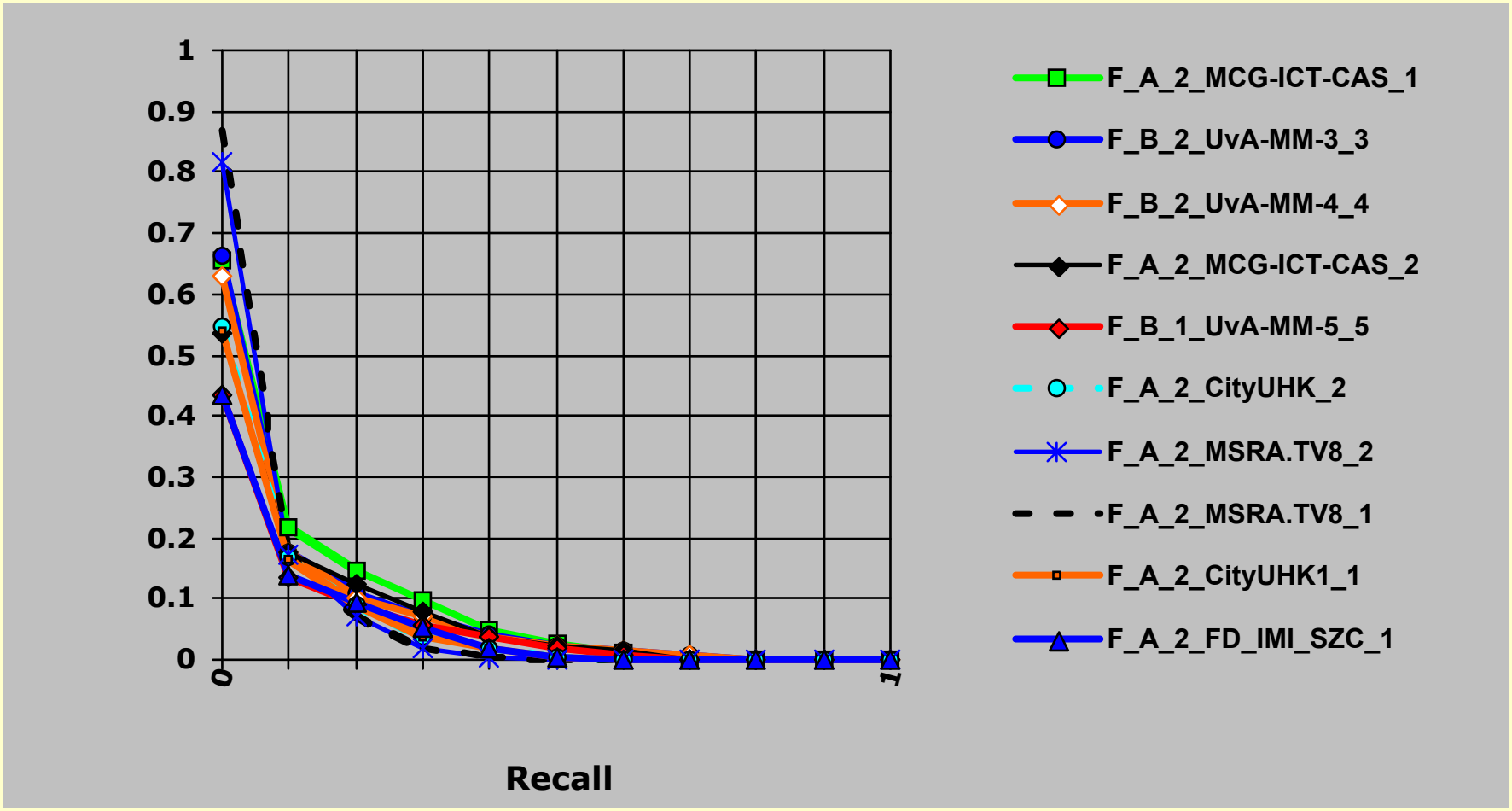
Normal **automatic** runs - top 10 mean AP (of 88 runs)



Another view: in highest scoring run, on average almost 5 of the top 10 shots returned contain the desired video

2008 Automatic runs - top 10 mean infAP

(mean elapsed time (mins) / topic)



Another view: in highest scoring run, on average between 2 and 3 of the top 10 shots returned are estimated to contain the desired video

Significant differences in top 10 automatic runs

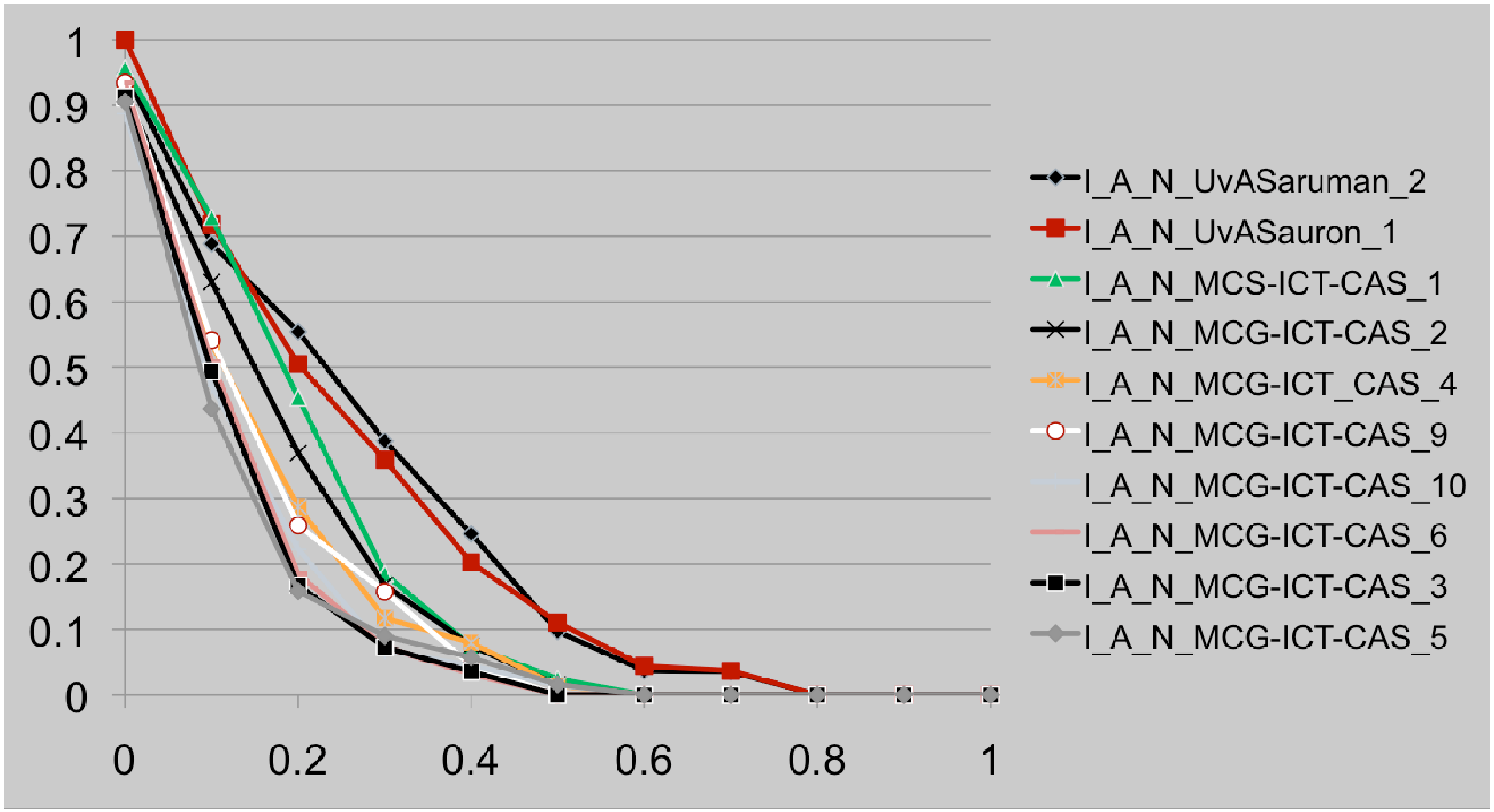
(using randomization test, 10**4 iterations, p < 0.05)

Run name	(mean AP)
F_A_N_BUPT-MCPR3_4	0.131
F_A_N_BUPT-MCPR0_1	0.104
F_A_N_PKU-ICST-4_4	0.098
F_A_N_PKU-ICST-7_7	0.096
F_A_N_PKU-ICST-9_9	0.095
F_A_N_PKU-ICST-5_5	0.095
F_A_N_PicSOM_8_3	0.091
F_A_N_PicSOM_7_4	0.091
F_A_N_PKU-ICST-10_10	0.090
F_A_N_UvaFrodo_3	0.089

BUPT-MCPR3_4

- ↘ BUPT-MCPR0_1
- ↘ PKU-ICST-10_10
- ↘ PicSOM_7_4
- ↘ PicSOM_8_3

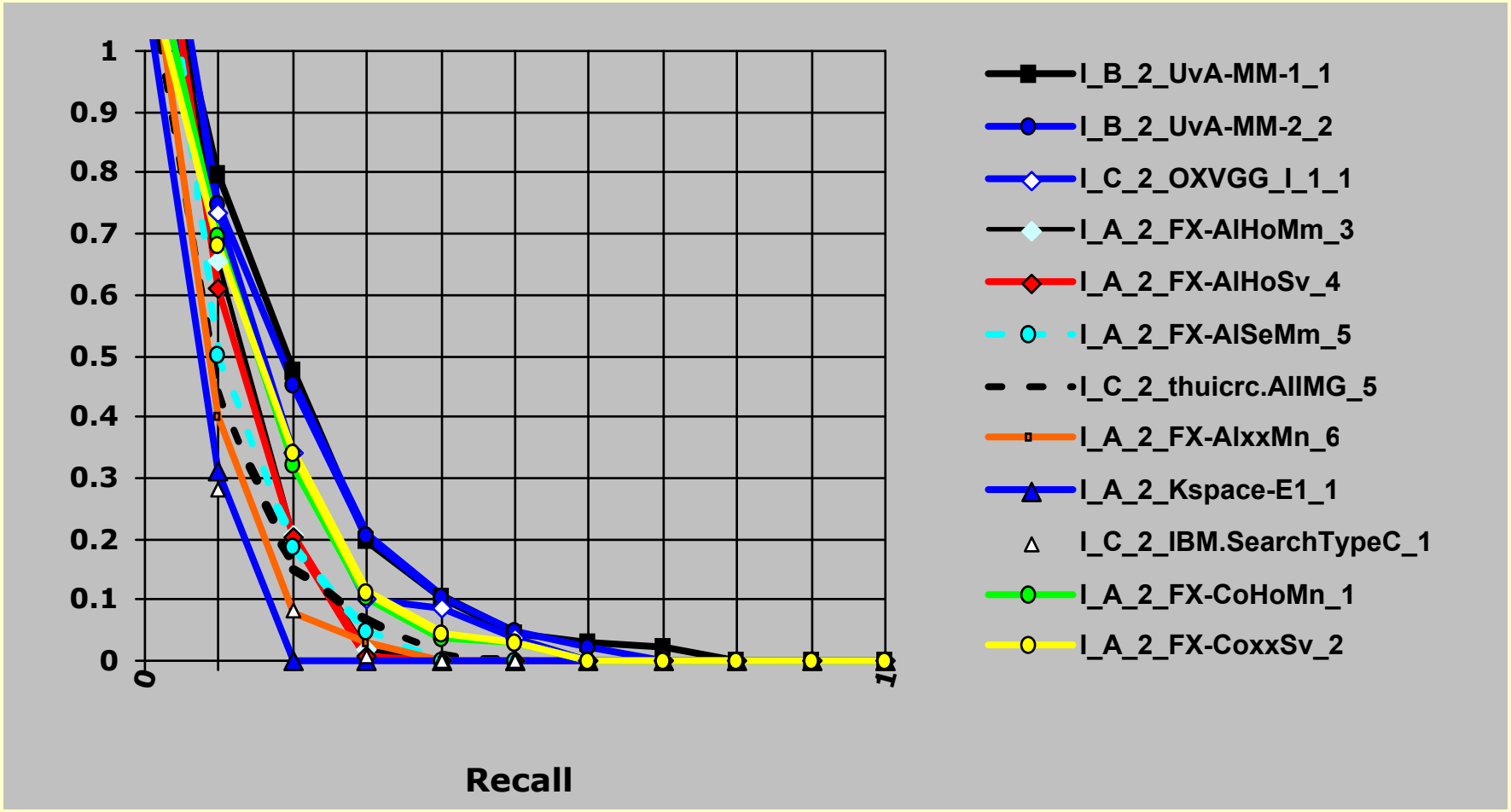
Normal **interactive** runs - top 10 mean AP (of 24)



Another view: in highest scoring run, on average 8 of the top 10 shots returned contained the desired video

2008 Interactive runs - top 10 mean infAP

(mean elapsed time (mins) / topic)



Another view: in highest scoring run, on average an estimated 7 of the top 10 shots returned contained the desired video

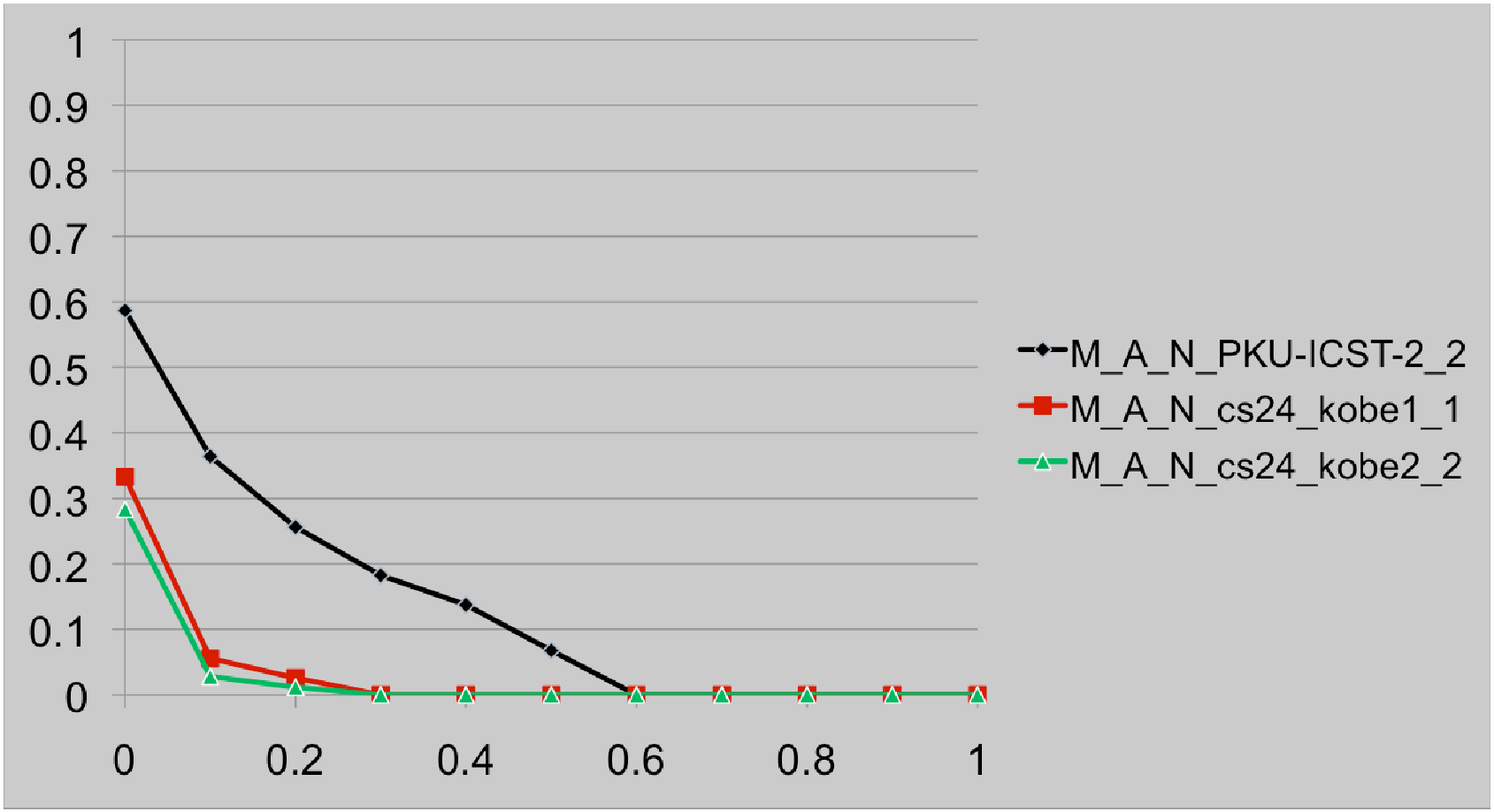
Significant differences in top 10 interactive runs

(using randomization test, 10**4 iterations, p < 0.05)

Run name		(mean AP)
I A N	UvASaruman _2	0.246
I A N	UvASauron_1	0.241
I A N	MCS-ICT-CAS_1	0.186
I A N	MCS-ICT-CAS_2	0.169
I A N	MCS-ICT-CAS_4	0.149
I A N	MCS-ICT-CAS_9	0.139
I A N	MCS-ICT-CAS_10	0.118
I A N	MCS-ICT-CAS_6	0.117
I A N	MCS-ICT-CAS_3	0.112
I A N	MCS-ICT-CAS_5	0.109

UvASaruman _2
UvASauron_1
- ↘ **MCG-ICT-CAS_1**
  - ↘ **MCG-ICT-CAS_4**
    MCG-ICT-CAS_9
    - ↘ **MCG-ICT-CAS_10**
    - ↘ **MCG-ICT-CAS_6**
    - ↘ **MCG-ICT-CAS_3**
    - ↘ **MCG-ICT-CAS_5**
- ↘ **MCG-ICT-CAS_2**
  - ↘ **MCG-ICT-CAS_9**
  - ↘ **MCG-ICT-CAS_10**
  - ↘ **MCG-ICT-CAS_6**
  - ↘ **MCG-ICT-CAS_3**
  - ↘ **MCG-ICT-CAS_5**

Normal Manual runs – All 3



Another view: in highest scoring run, on average about 4 of the top 10 shots returned contained the desired video

High-precision runs (mean AP)

Interactive:

- I_C_P_UniS_1 0.712

Manual:

- M_A_P_PKU-ICST-1_1 0.354

Automatic:

- F_A_P_PKU-ICST-6_6 0.263
- F_A_P_PKU-ICST-3_3 0.236
- F_A_P_NII.SEVIS_7 0.215
- F_A_P_NII.SEVIS_9 0.159
- F_A_P_NII.SEVIS_10 0.142
- F_A_P_NII.SEVIS_8 0.126

Significant differences:

PKU-ICST-6_6

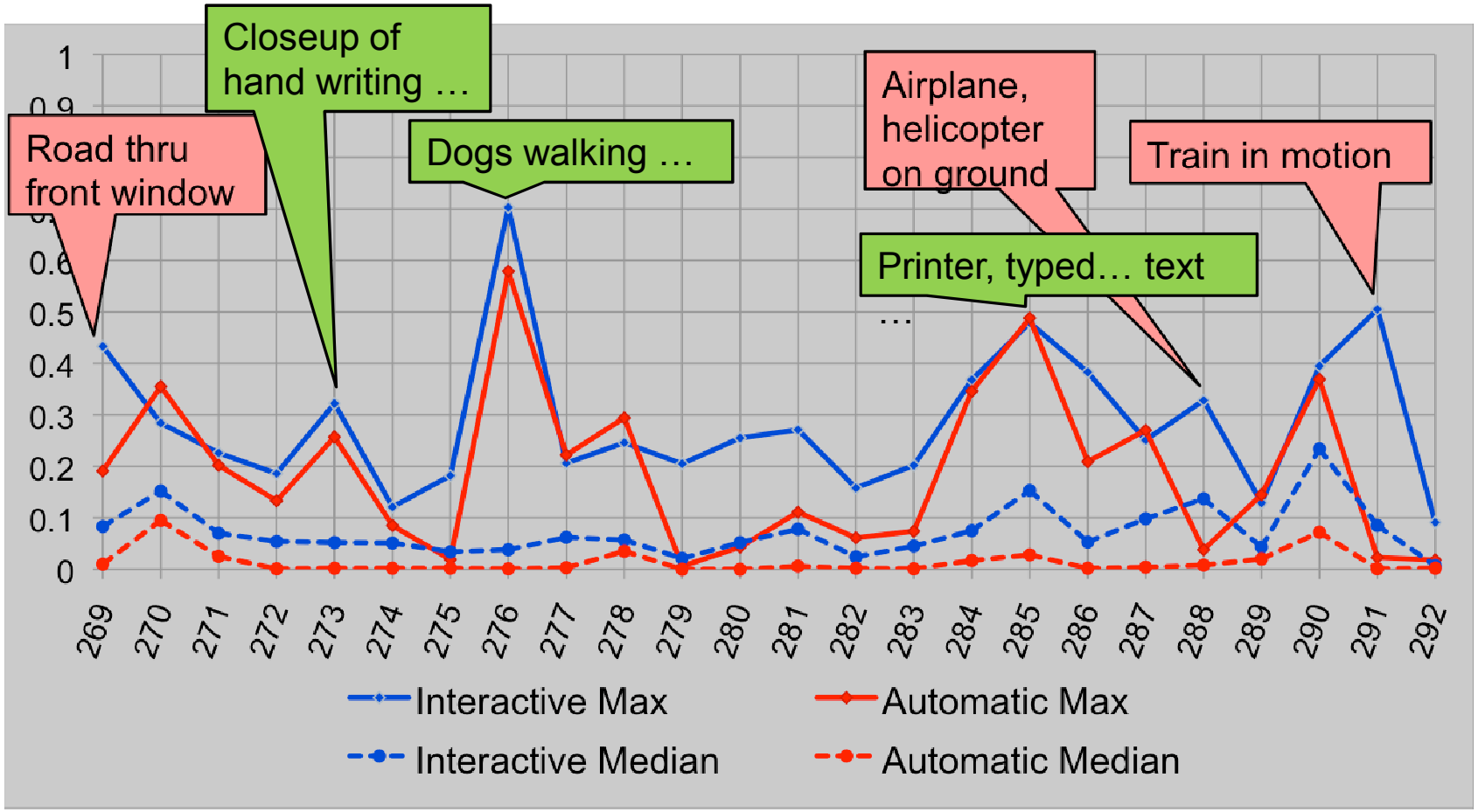
PKU-ICST-3_3

- ↘ **NII.SEVIS_8**

NII.SEVIS_7

- ↘ **NII.SEVIS_8**
- ↘ **NII.SEVIS_10**

Variation in AP by topic (normal search runs)



Crowds of people (270), Building entrance (278), People at desk with computer (287) each had automatic max better then interactive max

Speakers to follow …

- University of Amsterdam (MediaMill)
  - Helping searchers find good strategies
    - Active zooming
    - Relevance feedback using passive sampling of browsing
- VITALAS – CWI Amsterdam
  - Detailed study of some novice vs. professional searchers, interactive search
  - 29-author EU research project Aristotle U. Thessaloniki
- Kobe University
  - Making the most of positive and negative examples

Approaches

- **Beijing University of Posts and Telecom.-BUPT-MCPRL**
  - Automatic, using HLFs/concepts, and visual example-based retrieval, then weighting the combination as multimodal fusion, then including face scores.
  - 10 runs are variation combinations of the above, use Weight Distribution based on Semantic Similarity (WDSS) yielding top performing automatic run
- **Brno University of Technology**
  - Automatic runs based on transformed local image features (points, edges, homogeneous regions), i.e. SIFT
  - Used with face detection and global features, and then color layout and texture features. Similar to previous submissions.

Approaches

- **Budapest Academy of Sciences**
  - Hungarian Academy of Sciences - linear combinations of
    - ASR text
    - image similarity of representative frames
    - face detector output for topics involving people
    - weight of high level feature classifiers considered relevant by text based similarity to the topic
    - motion information extracted from videos where relevant to topic,
    - … plus some shot contexts (neighbor shots).

Approaches

- **Centre for Research and Technology Hellas**
  - ITI/CERTH Thessaloniki in interactive search, combining retrieval functionalities in various modalities (i.e. textual, visual and concept search) with a user interface supporting interactive search over all queries submitted.
- **Chinese Academy of Sciences-MCG-ICT-CAS**
  - Interactive search using "VideoMap" system with a map based display interface, giving a global view of similarity relationships throughout the whole video collection
  - Multiple modality feedback strategies, including the visual-based feedback, concept-based feedback and community-based feedback

Approaches

- **City University of Hong Kong w/ Columbia U**
  - Automatic search - previous years focus on concept-based search, using various techniques to determine which concepts to use, include Flickr usage
  - Now also factor in visual query examples and address combination of multiple search modalities
  - Multimodal search fusion - yielded 10% improvement
- **Helsinki University of Technology TKK**
  - Automatic runs combined ASR/MT text search and concept-based retrieval.
  - If none of the concept models could be matched with the query, used content-based retrieval based on the video and image examples instead.
  - Portfolio of 10 runs with text, visual similarity, own concepts, and donated (MediaMill and CU-VIREO374) concepts individually, and in combinations

Approaches

- **KB Video Retrieval** (David Etter)
  - Automatic search, focus on query expansion by adding terms (texts) and images, using Wikipedia titles and images as a source
- **Laboratoire REGIM**
  - Combine text search (against ASR transcript) and visual (colour, texture, shape) from keyframes
- **National Institute of Informatics**
  - Automatic runs only
  - Trained an SVM concept detector for each query, also used kNN matching on visual, concept selection using visual features, concept selection using text descriptions

Approaches

- **Peking University-PKU-ICST**
  - Automatic, and manual search
  - 10 search runs with list of in-house variations
  - multi-modal including weighted combination of visual-based, concept-based, audio features, and faces for some topics
  - Two retrieval approaches - pairwise similarity and learning-based ranking - excellent performance
- **The Open University**
  - 8 automatic search submissions based on determining the distance from a query image to a pre-indexed collection of images to build a list of results ordered by visual similarity.
  - Used four metric measures (Euclidian, Manhattan, Canberra and Squared Chord) and two data normalisations

Approaches

- **University of Glasgow**
  - Automatic runs based on MPEG7 features, concepts, and BoW derived from SIFT features
  - Investigation into estimating topic distribution using the Latent Dirichlet Allocation (LDA) with run variants to explore this
  - Median performance
- **Beijing University of Posts and Telecom.-PRIS**
- **University of Surrey**
- **Zhejiang University**

Questions 2008…

- Did systems adapt to new data/topic characteristics?
  - What old approaches stopped/continued working?
  - What new approaches were tried with(out) success?
- Did systems do anything special to support search for events?
- How did systems handle search for grayscale video?
- What is collaborative search all about?
- What experimental designs are being used to isolate the system effect from the search effect in interactive searches?

Some questions for 2009 …

- What old approaches stopped/continued working?
- What new approaches were tried with(out) success?
- What method/test was used to distinguish real differences between runs from chance differences?
- What experimental designs were used to isolate the system effect from the searcher and topic effects in interactive searches?
- What sort of search tasks make sense against some subset of the Internet Archive video?
- Please cite the TRECVid reference, even in TRECVid workshop papers as this does help us make the case

VideOlympics 2009

- Following CIVR in Amsterdam and Niagara, 7 systems took part at CIVR in 2009 Santorini
- DCU, NUS, CAS (previously part of NUS), MediaMill/UvA, Grenoble/ Marseilles, Tsinghua, and ITI Greece, the home team
- Organisation was impeccable and Cees and Marcel did a great job.
- Guest searchers introduced halfway through the 7 topics mixed things up a bit as they did really well (the topics also got easier too)
- A couple of searchers found 100+ relevant shots in 5 minutes for a couple of the topics.
- Guest searchers included Tat-Seng's wife, Nicu's wife, Rita's husband, Yannis' girlfriend, somebody else's partner, the guy from the conference venue who does the AV, and the conference venue manager
- A small, but successful activity