TRECVID-2008: Search Task

Alan Smeaton
Dublin City University
&
Paul Over
NIST

Search Task Definition

- Goal: promote progress in content-based retrieval from digital video via open, metrics-based evaluation
- Given a test collection, a topic and a common shot boundary reference, return a ranked list of at most 1,000 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
- Different in 2008:
  - Double the number of topics (48) for automatic runs
  - Evaluate based on a 50% random sample of the judgment pools and use inferred average precision (infAP) in place of average precision

Search Task Definition

- Per-search measures: inferred average precision (infAP), elapsed time
- Per-run measure: mean inferred 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.
  - EXCEPT: experimental **collaborative runs** from FX Palo Alto Labs.

2008 data (same source as 2007)

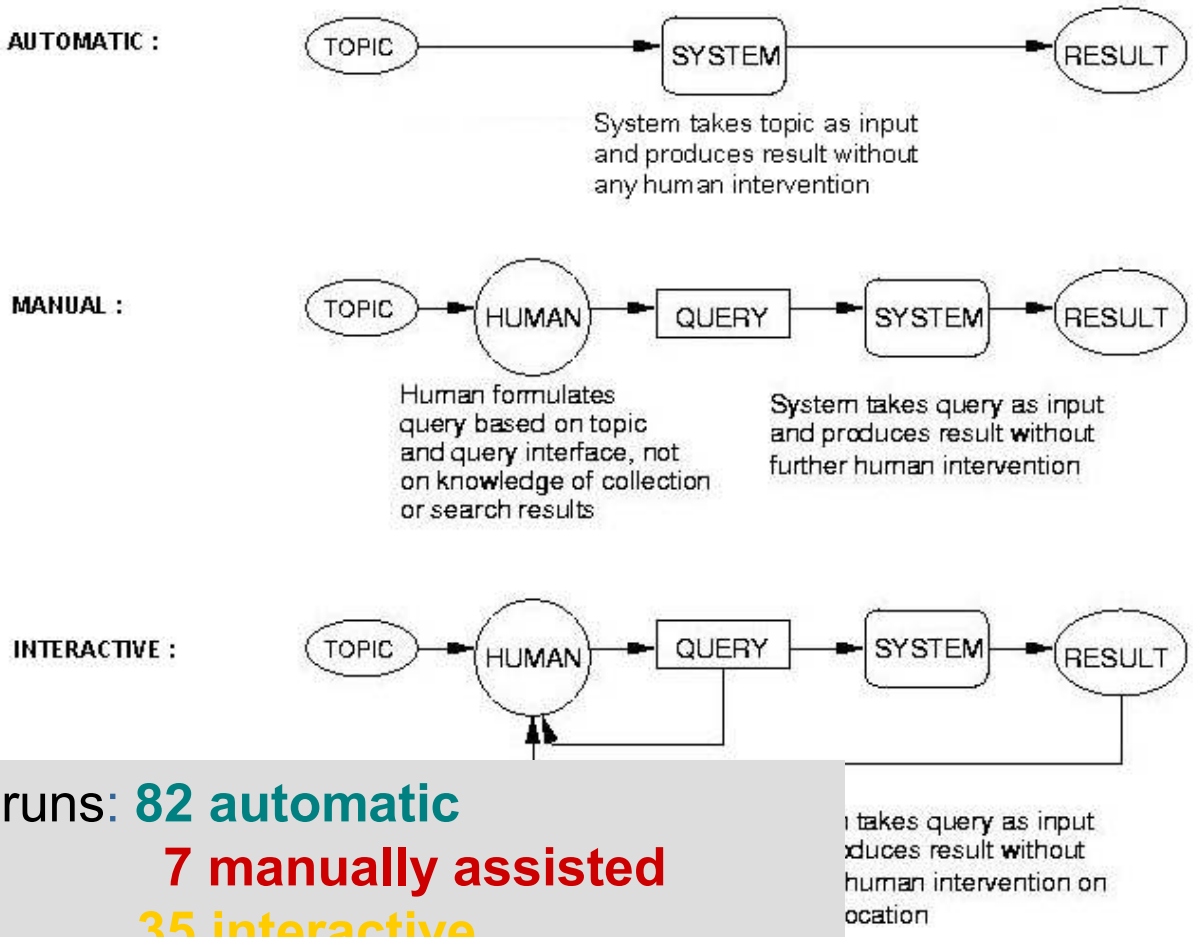
- Educational, cultural, youth-oriented programming, news magazines, historical footage, etc.
- Primarily in Dutch, but also some English, etc.
- Much less repetition
  - No commercials
  - No repeated stock news footage
  - Greater variety of subject matter than in broadcast news

2008: Search task participants

Brno University of Technology	CD ED FE ** SE
Columbia University	CD -- FE -- SE
COST292 Team (Delft Univ.)	CD ** FE RU SE
cs24_kobe (Kobe Univ.)	-- -- ** -- SE
Dublin City University	-- ED -- RU SE
Fudan University	CD ED FE -- SE
FX Palo Alto Laboratory	-- -- -- RU SE
IBM T. J. Watson Research Center	CD ** FE ** SE
KB Video Retrieval	-- -- -- -- SE
K-Space	-- -- ** RU SE
University of Twente and CWI	-- -- FE -- SE
MCG-ICT-CAS	CD ED FE -- SE
Mediamill (Univ. of Amsterdam)	-- ** FE -- SE
MESH	-- -- FE -- SE
MMIS (Open Univ.)	** -- FE -- SE
Microsoft Research Asia	** ** FE ** SE
National Institute of Informatics	CD ** FE RU SE
National University of Singapore	-- -- -- -- SE
National Taiwan University	** ** FE -- SE
Oxford Univ.	** -- FE -- SE
PKU-ICST (Peking Univ.)	** ** FE ** SE
PicSOM (Helsinki University of Technology)	CD -- FE RU SE
REGIM	-- ** FE RU SE
SJTU	-- ED FE -- SE
SP-UC3M (Universidad Carlos III de Madrid)	-- -- FE -- SE
Tsinghua University-Intel China Research Center	CD ** FE RU SE
University of Alabama	-- -- -- -- SE
University of Glasgow	CD -- ** RU SE
VIREO (City University of Hong Kong)	CD ** FE RU SE
VITALAS (CERTH-ITI (GR), CWI(NL), U.Sunderland (UK))	-- -- FE -- SE

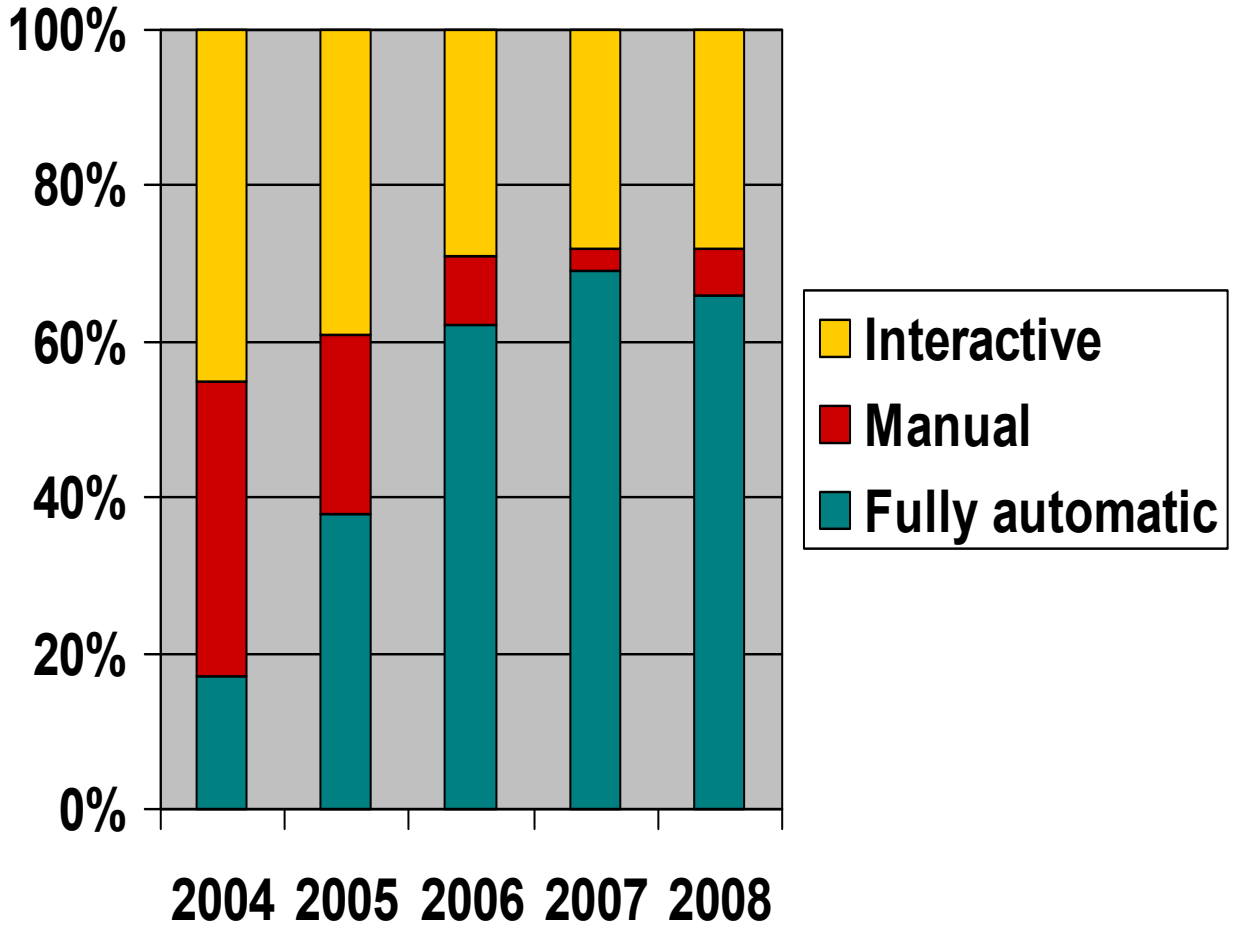
** : group applied but did not submit a run -- : group didn't apply for the task

Search Types: Automatic, Manual and Interactive



Number of runs: **82 automatic**
7 manually assisted
35 interactive

Trends stabil



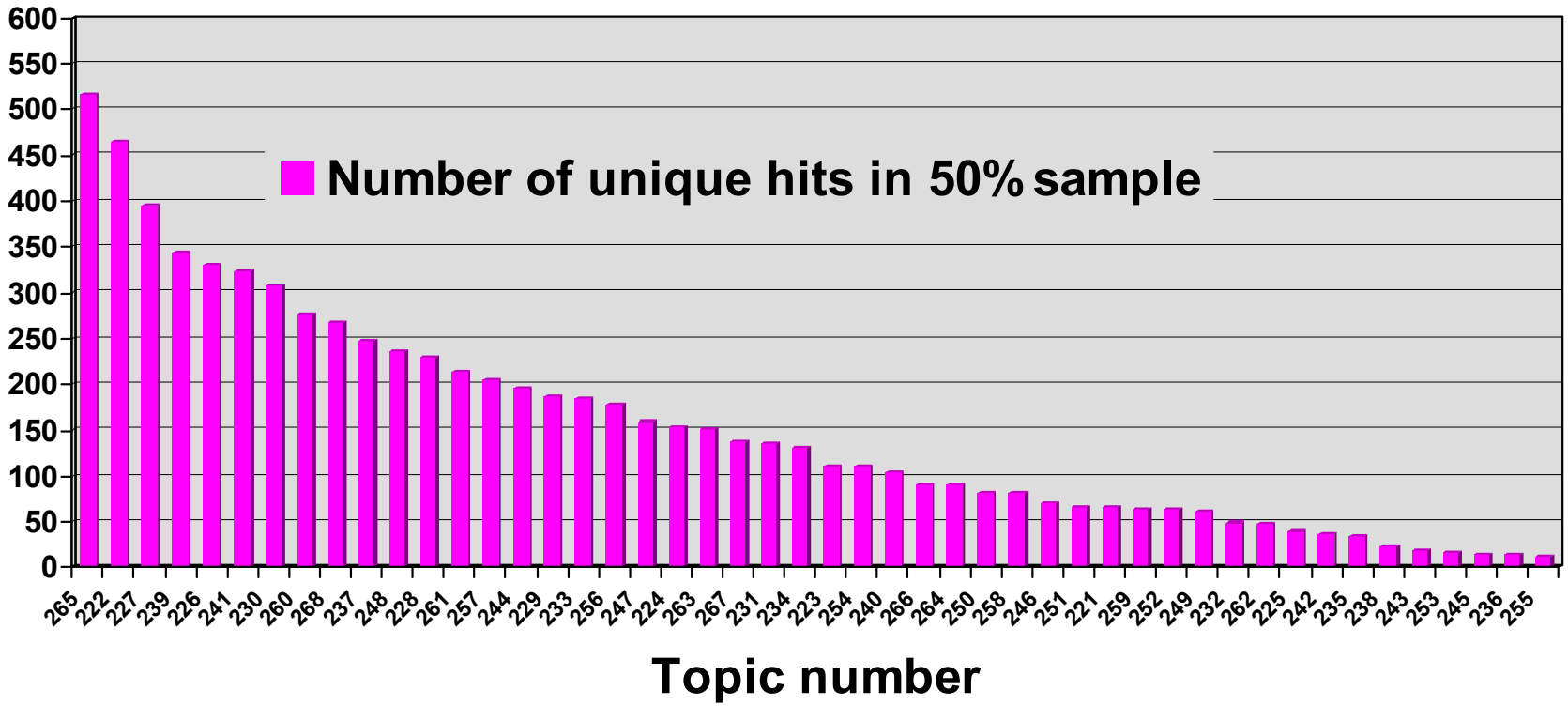
24 Topics (for all systems)

221. Find shots of a person opening a door
222. Find shots of 3 or fewer people sitting at a table
223. Find shots of one or more people with one or more horses
224. Find shots of a road taken from a moving vehicle, looking to the side
225. Find shots of a bridge
226. Find shots of one or more people with mostly trees and plants in the background; no road or building can be seen
227. Find shots of a person's face filling more than half of the frame area
228. Find shots of one or more pieces of paper, each with writing, typing, or printing it, filling more than half of the frame area
229. Find shots of one or more people where a body of water can be seen
230. Find shots of one or more vehicles passing the camera
231. Find shots of a map
232. Find shots of one or more people, each walking into a building
233. Find shots of one or more black and white photographs, filling more than half of the frame area
234. Find shots of a vehicle moving away from the camera
235. Find shots of a person on the street, talking to the camera
236. Find shots of waves breaking onto rocks
237. Find shots of a woman talking to the camera in an interview located indoors - no other people visible
238. Find shots of a person pushing a child in a stroller or baby carriage
239. Find shots of one or more people standing, walking, or playing with one or more children
240. Find shots of one or more people with one or more books
241. Find shots of food and or drinks on a table
242. Find shots of one or more people, each in the process of sitting down in a chair
243. Find shots of one or more people, each looking into a microscope
244. Find shots of a vehicle approaching the camera

24 additional topics (for automatic runs only)

245. Find shots of a person watching a television screen - no keyboard visible
246. Find shots of one or more people in a kitchen
247. Find shots of one or more people with one or more animals
248. Find shots of a crowd of people, outdoors, filling more than half of the frame area
249. Find shots of a classroom scene
250. Find shots of an airplane exterior
251. Find shots of a person talking on a telephone
252. Find shots of one or more people, each riding a bicycle
253. Find shots of one or more people, each walking up one or more steps
254. Find shots of a person talking behind a microphone
255. Find shots of just one person getting out of or getting into a vehicle
256. Find shots of one or more people, singing and/or playing a musical instrument
257. Find shots of a plant that is the main object inside the frame area
258. Find shots of one or more people sitting outdoors
259. Find shots of a street scene at night
260. Find shots of one or more animals - no people visible
261. Find shots of one or more people at a table or desk, with a computer visible
262. Find shots of one or more people in white lab coats
263. Find shots of one or more ships or boats, in the water
264. Find shots of one or more colored photographs, filling more than half of the frame area
265. Find shots of a man talking to the camera in an interview located indoors - no other people visible
266. Find shots of more than 3 people sitting at a table
267. Find shots with the camera zooming in on a person's face
268. Find shots of one or more signs with lettering

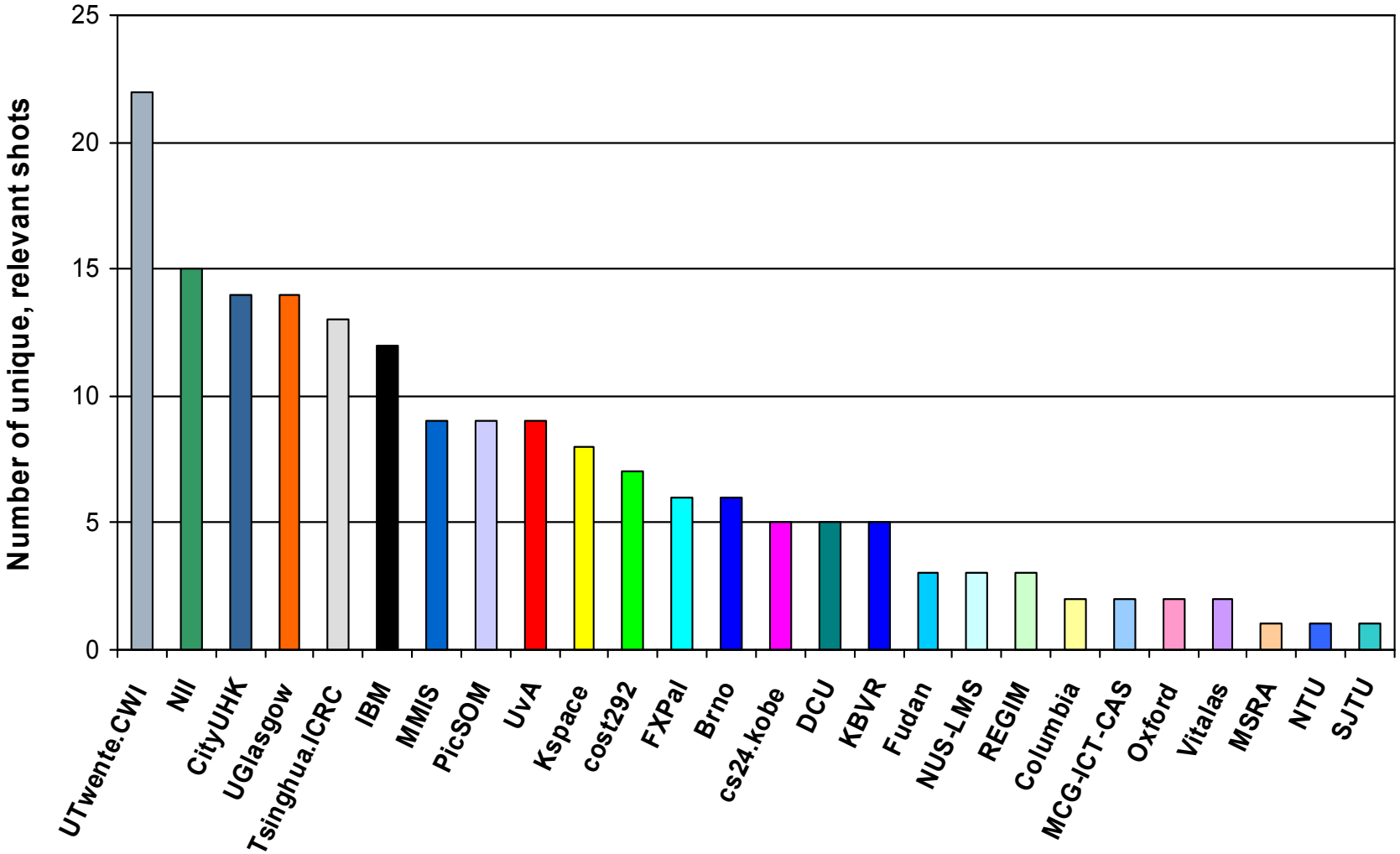
Distribution of hits for each topic



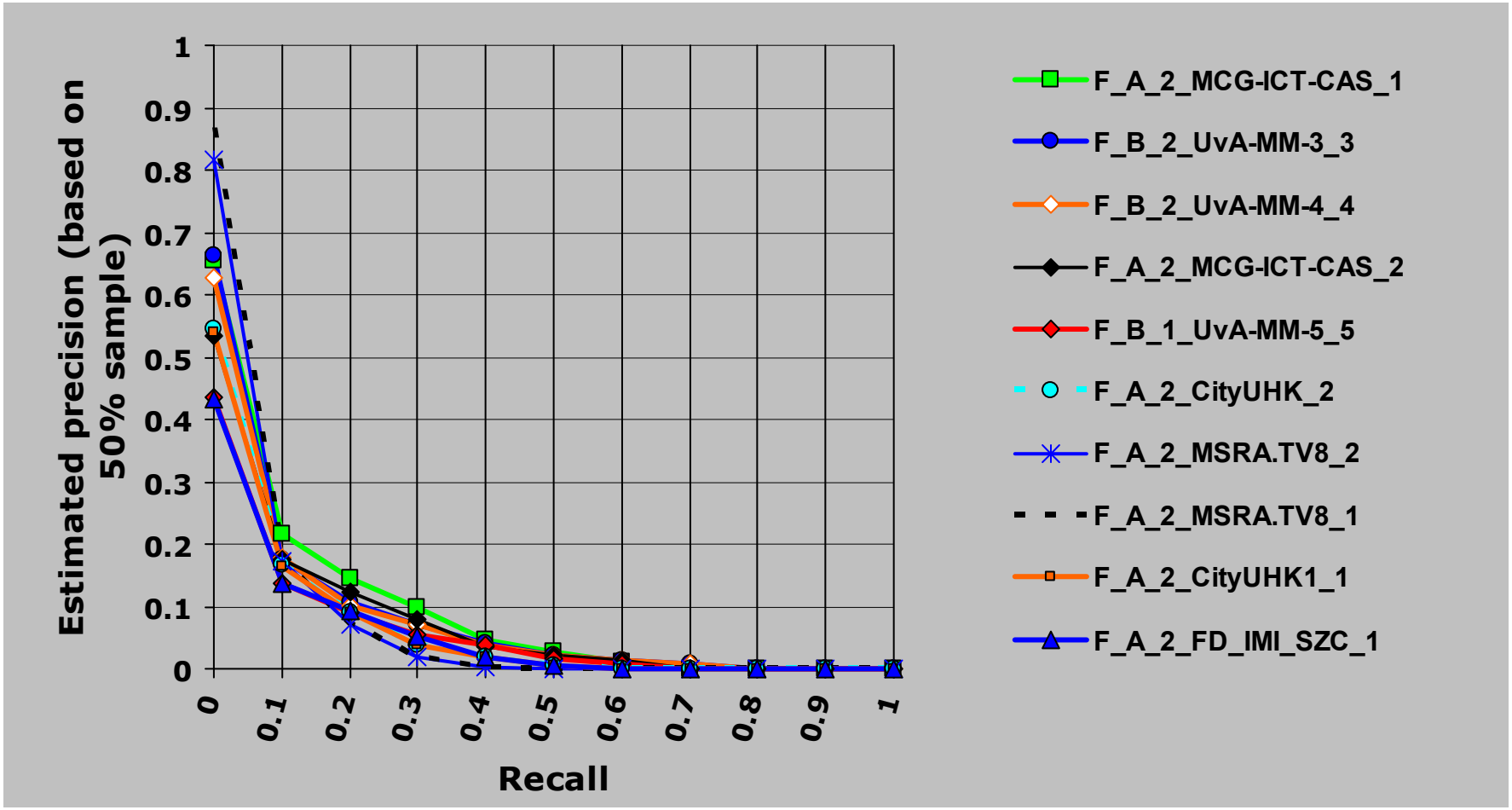
Frequency of target topic-shots

- **2008**
  - **Test shots * 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%**
- 2003
  - Test shots * topics: **775,632**
  - Relevant topic-shots: **2,114 0.3%**

Relatively few unique, relevant shots by group



Automatic runs - top 10 mean infAP (of 82)



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

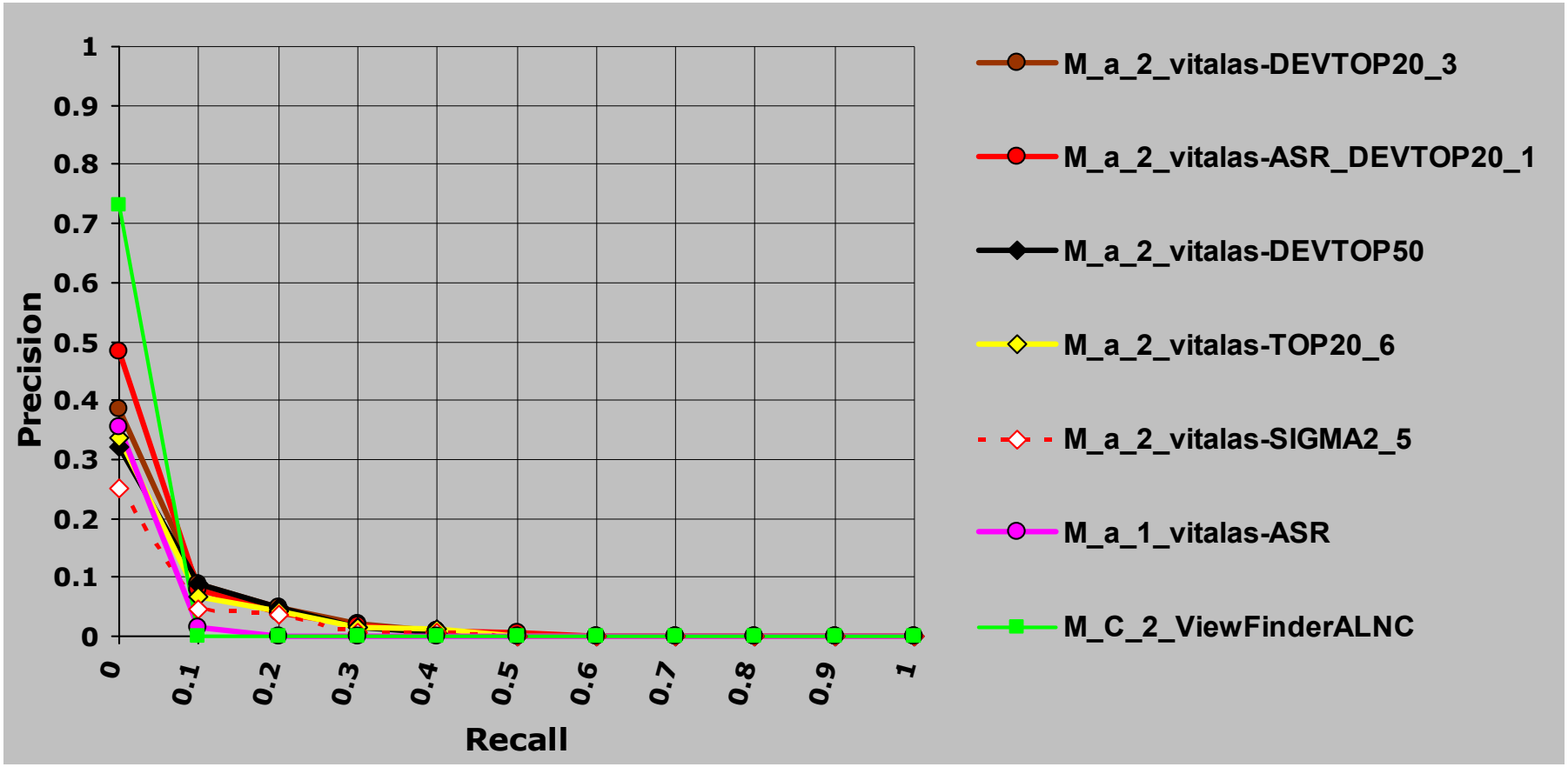
Significant differences among top 10 automatic runs (using randomization test, 10**4 iterations, p < 0.05)

Run name	(mean infAP)
A_2 _MCG-ICT-CAS_1	0.067
B_2 _UvA-MM-3_3	0.054
B_2 _UvA-MM-4_4	0.053
A_2 _MCG-ICT-CAS_2	0.053
B_1_UvA-MM-5_5	0.044
A_2_CityUHK2_2	0.042
A_2_MSRA.TV8_2	0.041
A_2_MSRA.TV8_1	0.041
A_2_CityUHK1_1	0.041
A_2_FD_IMI_SZC_1	0.040

A_2_MCG-ICT-CAS_1

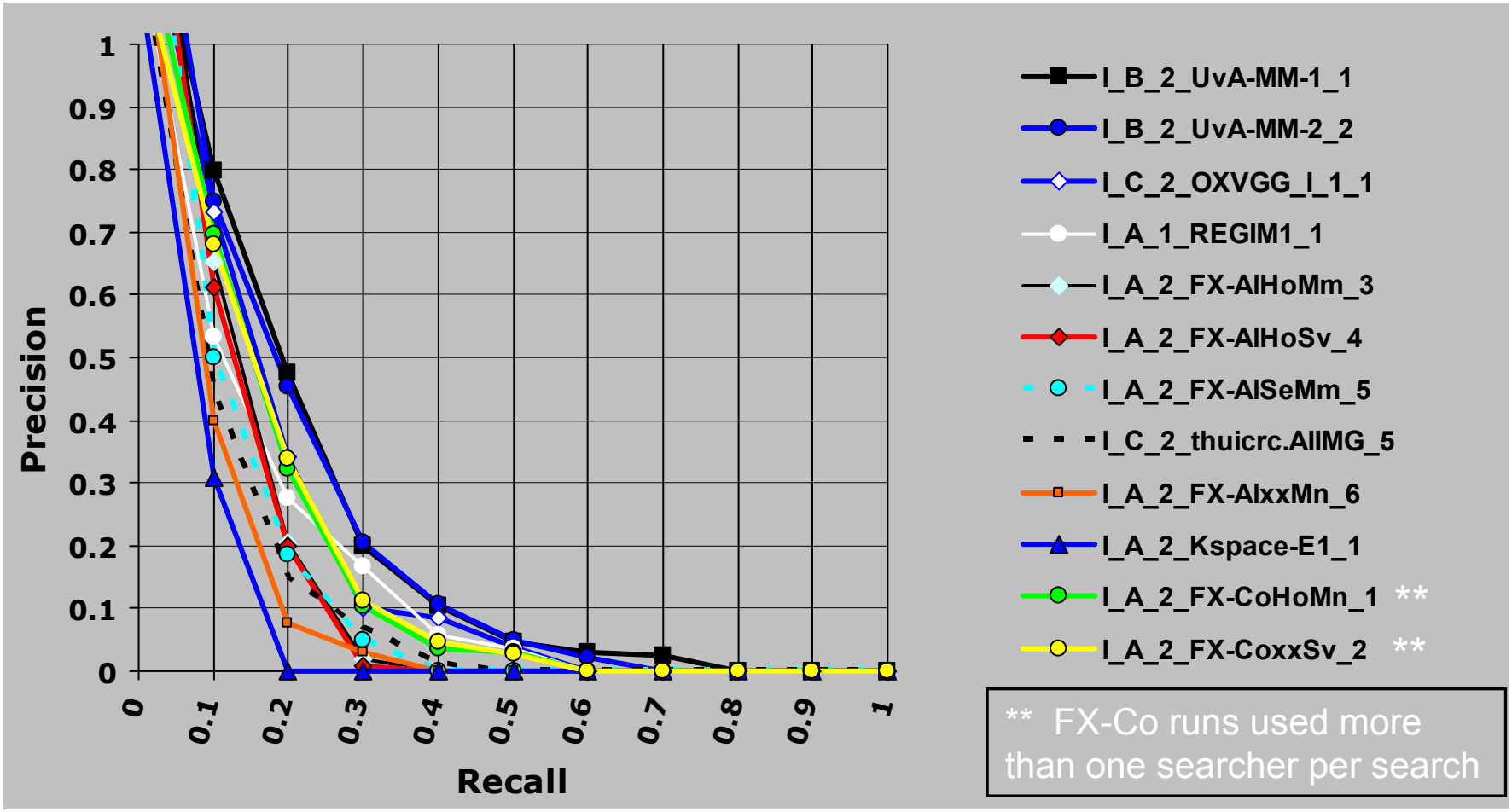
- ↘ B_2 _UvA-MM-3_3
- ↘ B_2 _UvA-MM-4_
- ↘ A_2 _MCG-ICT-CAS_2
- ↘ B_1_UvA-MM-5_5
- ↘ A_2_CityUHK2_2
- ↘ A_2_MSRA.TV8_2
- ↘ A_2_MSRA.TV8_1
- ↘ A_2_CityUHK1_1
- ↘ A_2_FD_IMI_SZC_1

Manual runs – All 7



Another view: in highest scoring run, on average 1 or 2 of the top 10 shots returned contained the desired video

Interactive runs - top 10 mean infAP (of 35) + 2



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

Significant differences among top 10+2 interactive runs (using randomization test, 10**4 iterations, p < 0.05)

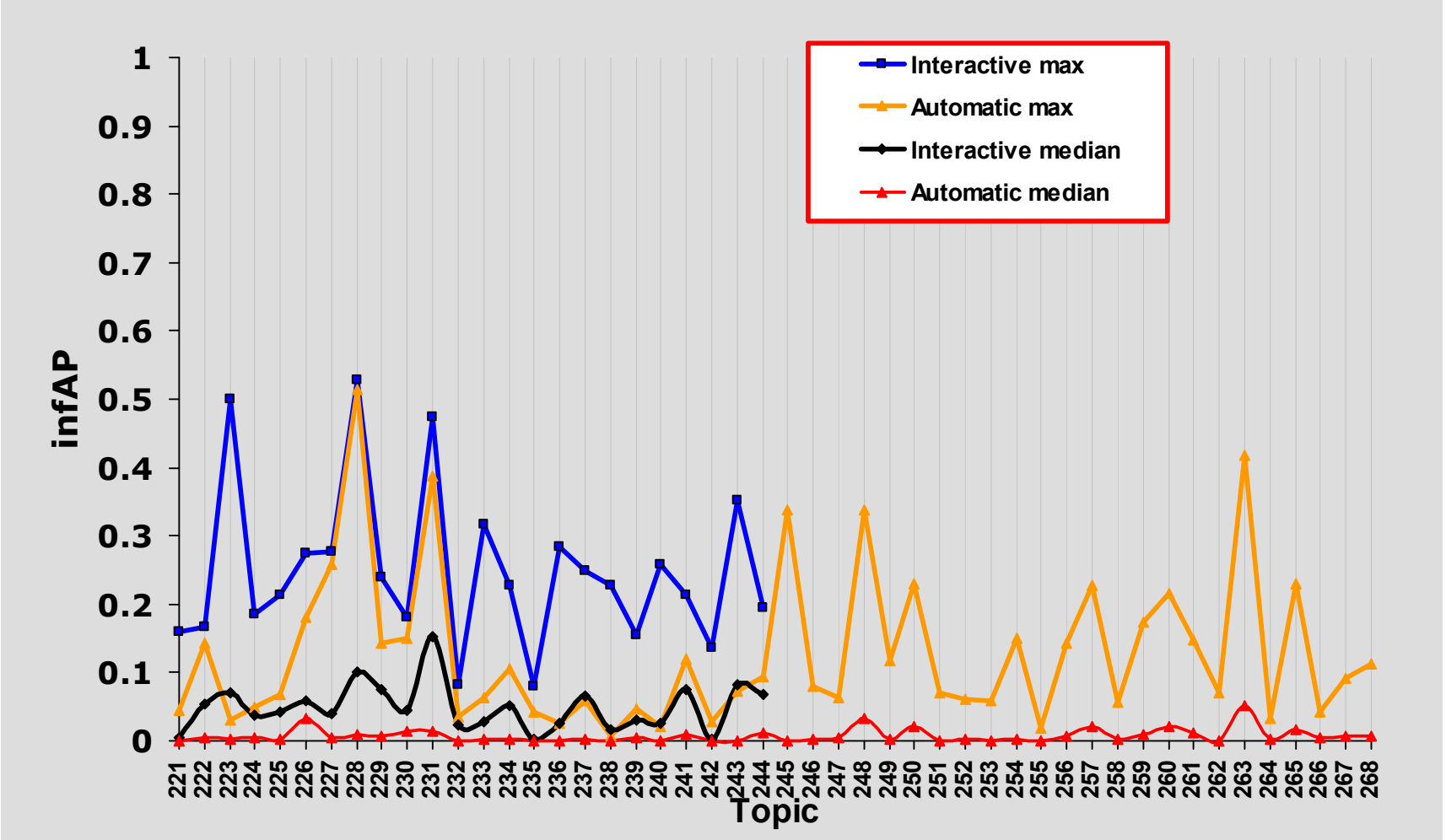
Run name	(mean infAP)
B 2 UvA-MM-1 1	0.194
B 2 UvA-MM-2 2	0.181
C 2 OXVGG_I_1 1	0.158
A 2 FX-CoHoMm 1	0.148
A 2 FX-CoxxSv 2	0.147
A 1 REGIM1 1	0.125
A 2 FX-AlHoMm 3	0.112
A 2 FX-AlHoSv 4	0.109
A 2 FX-AlSeMm 5	0.100
C 2 thuicrc.AllMG 5	0.099
A 2 FX-AlxxMm 6	0.076
A 2 KSpace-E1 1	0.068

B_2_UvA-MM-1
- C_2_OXVGG_I_1_1
- A_2_FX-CoHoMm_1
- A_2_FX-CoxxSv_2
- A_1_REGIM1_1
- A_2_FX-AlHoMm_3
- A_2_FX-AlHoSv_4
- A_2_FX-AlSeMm_5
- C_2_thuicrc.AllMG_5
- A_2_FX-AlxxMm_6
- A_2_KSpace-E1_1

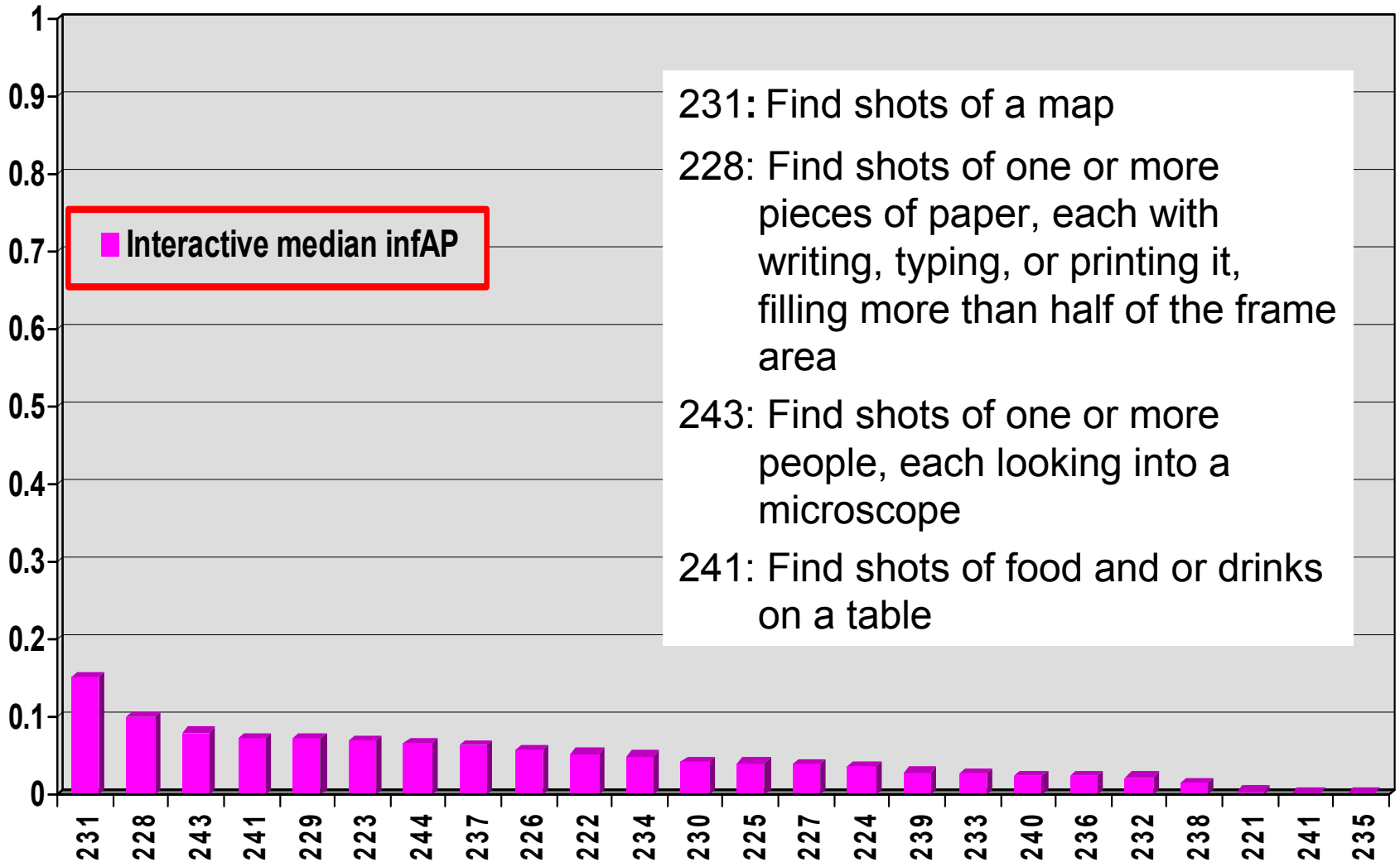
B_2_UvA-MM-2_2; C_2_OXVGG_I_1_1
- A_1_REGIM1_1
- A_2_FX-AlHoMm_3
- A_2_FX-AlHoSv_4
- A_2_FX-AlSeMm_5
- C_2_thuicrc.AllMG_5
- A_2_FX-AlxxMm_6
- A_2_KSpace-E1_1

FX-Co* runs used more than one searcher per search

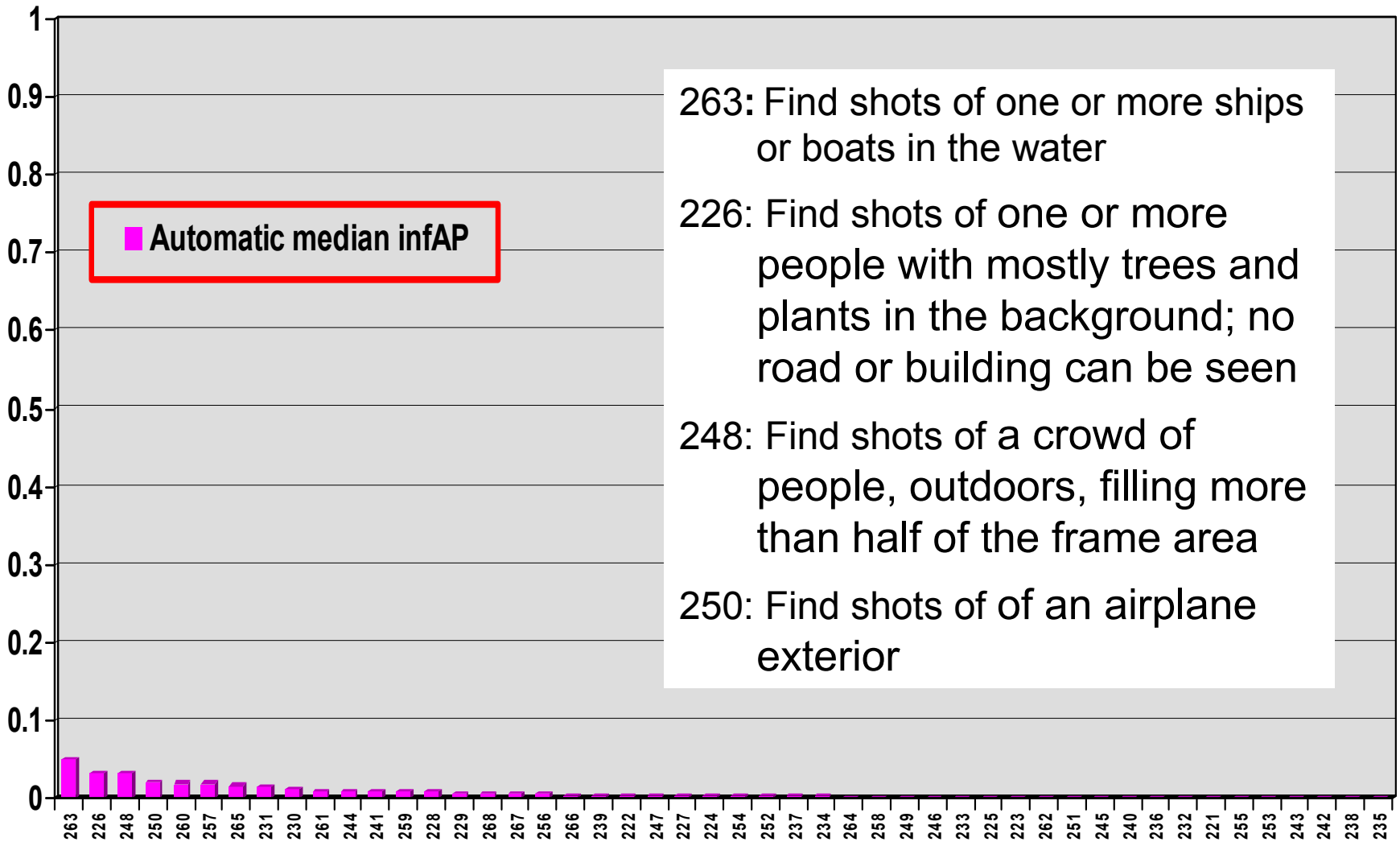
Inferred average precision by topic



Interactive runs' median infAP by topic



Automatic runs' median infAP by topic



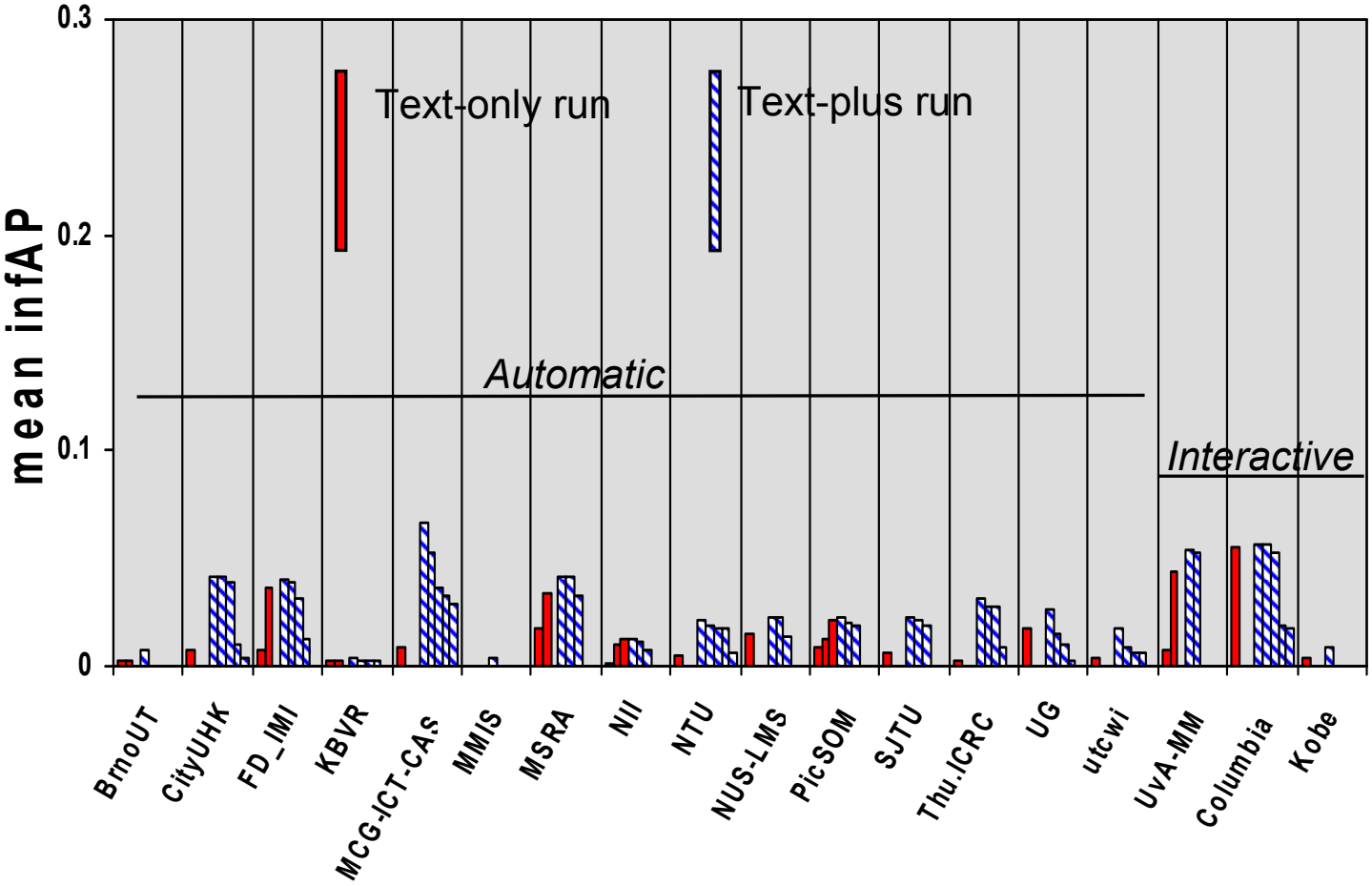
263: Find shots of one or more ships or boats in the water

226: Find shots of one or more people with mostly trees and plants in the background; no road or building can be seen

248: Find shots of a crowd of people, outdoors, filling more than half of the frame area

250: Find shots of of an airplane exterior

Text–only versus Text-plus



Speakers to follow

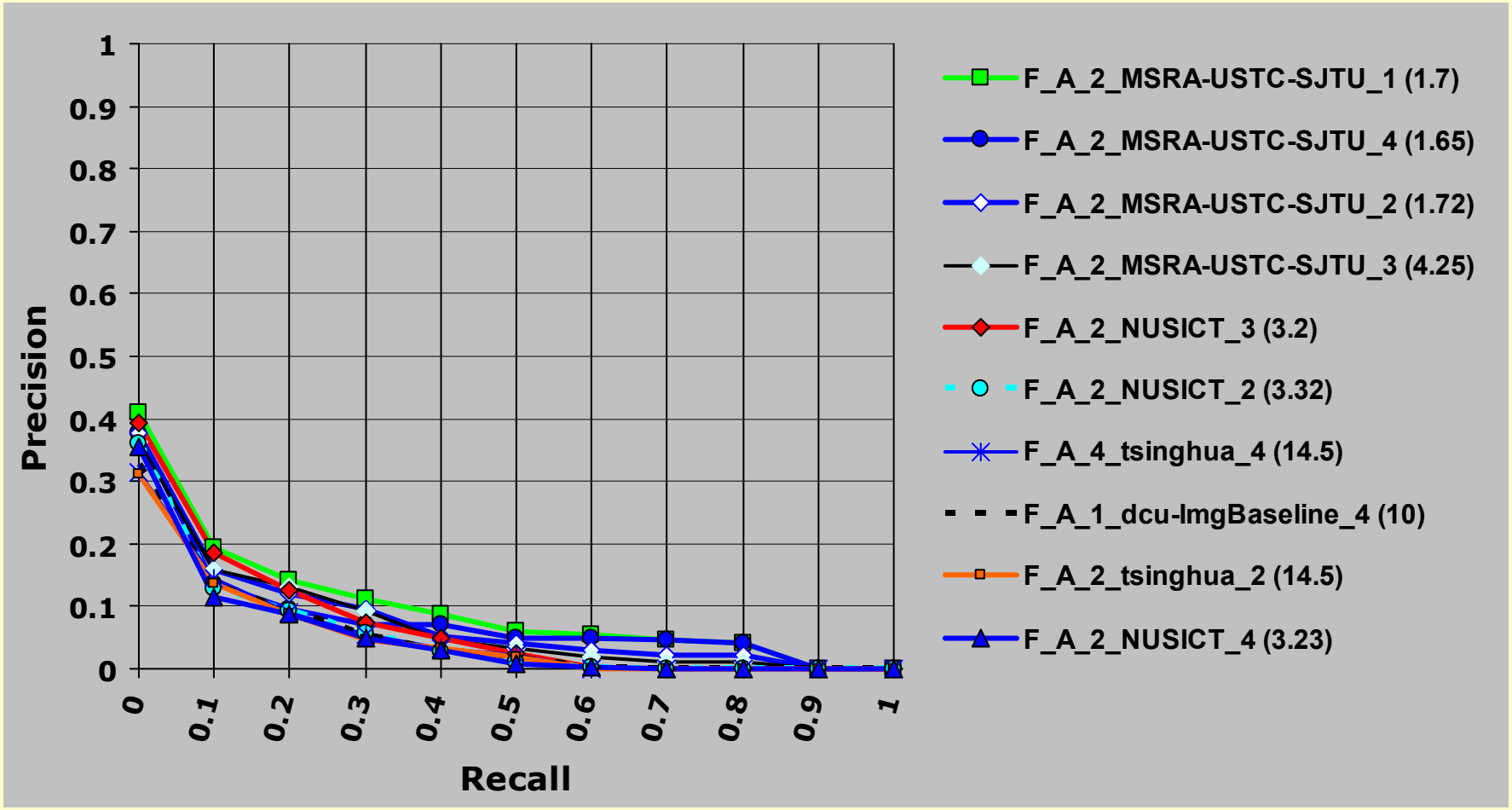
- Automatic
  - University of Amsterdam (MediaMill)
    - Optimal query mode (speech, detector, or example-based search) prediction by topic
  - Chinese Academy of Sciences (MCG-ICT-CAS)
    - Distribution based concept selection method
    - SIFT visual-keywords feature in low dimensional LDA semantic space
    - Re-ranking based on the motion and face
    - Dynamic fusion based on the Smoothed Similarity Cluster
- Interactive
  - K-Space
    - Large multi-site interactive search experiment
  - FX Palo Alto
    - Using program-based clustering to enhance search
    - Collaborative search

Some questions…

- 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?
- What method 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?

Automatic runs - top 10 MAP (of 81)

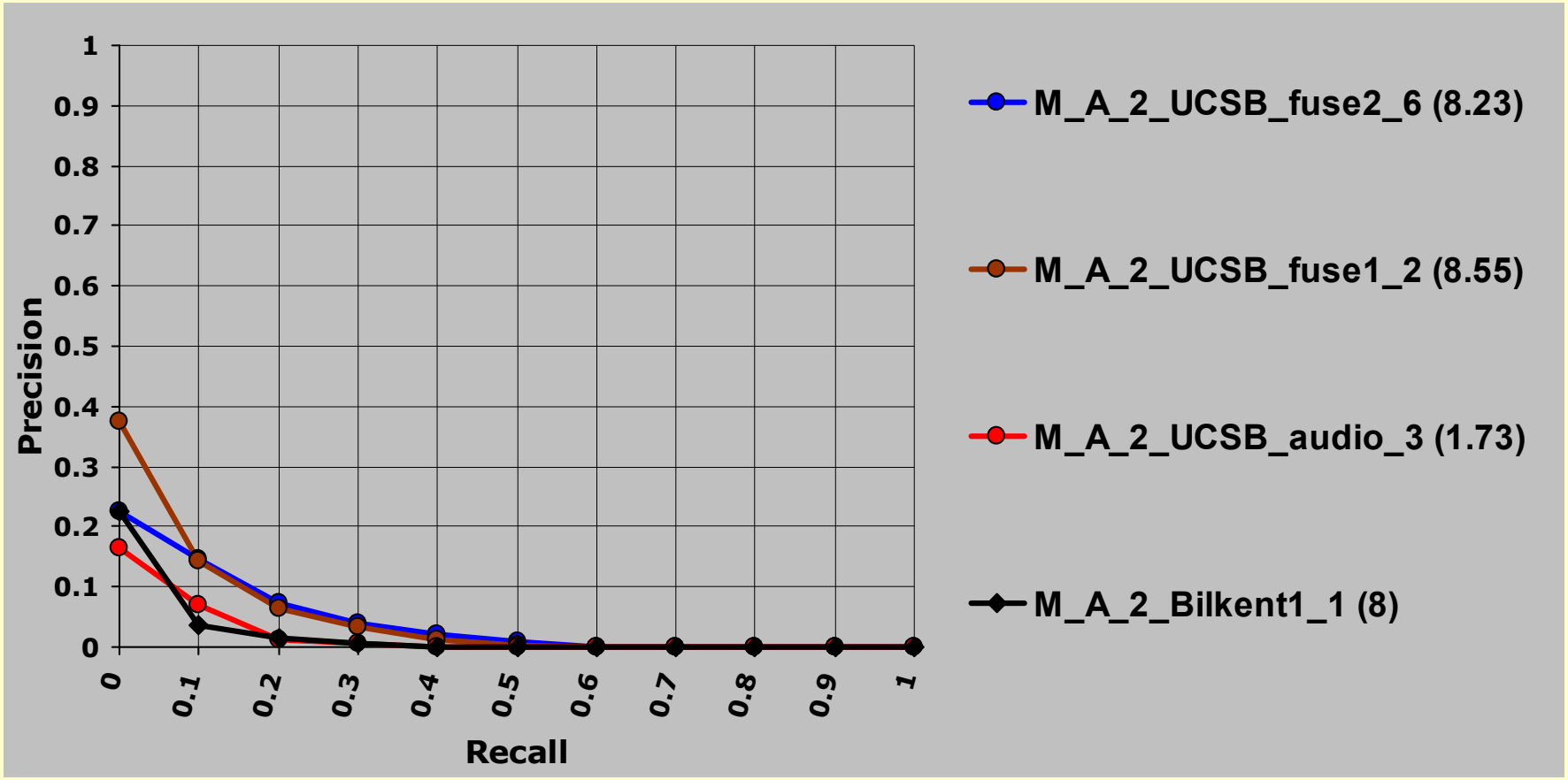
(mean elapsed time (mins) / topic)



Another view: in highest scoring run, on average a little more than 2 of the top 10 shots returned contained the desired video

Manual runs – All 4

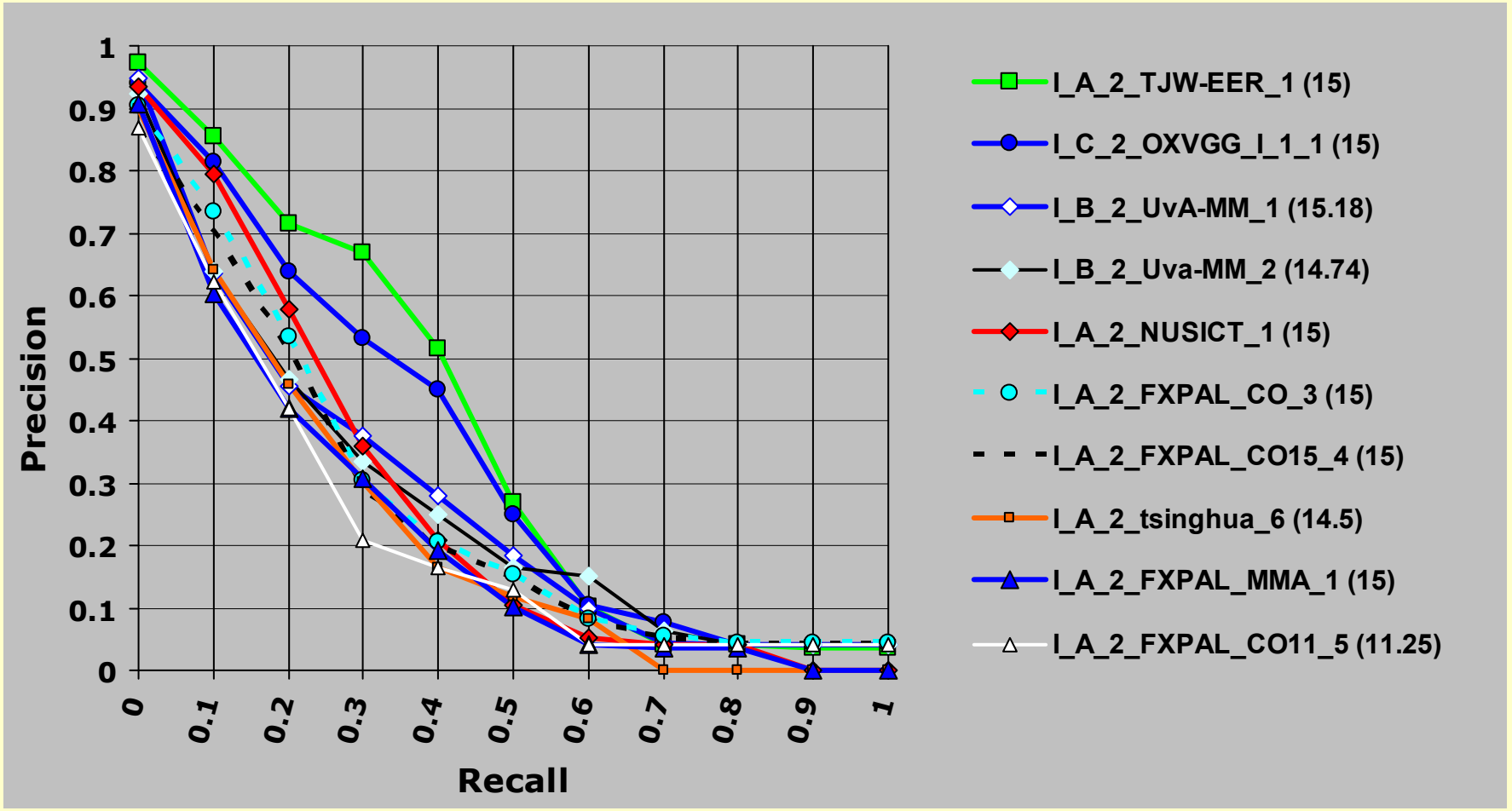
(mean human effort (mins) / topic)



Another view: in highest scoring run, on average not quite 1 of the top 10 shots returned contained the desired video

Interactive runs - top 10 MAP (of 33)

(mean elapsed time (mins) / topic)



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

Average precision by topic