TRECVid Video Evaluation

Alan F. Smeaton
CLARITY: Centre for Sensor Web Technologies
Dublin City University

TRECVid: Video IR Evaluation

- Benchmarking - always historically important in information access … from DBMS, to IR;
- TREC founded in 1991 to redress problems of data collections, common queries, uniform evaluation;
- In 2001, "video retrieval" started as a TREC track;
- Usual TREC mode of operation (data-topics-search submissions-pooling-evaluation by metrics-workshop) but additional, related tasks besides search;
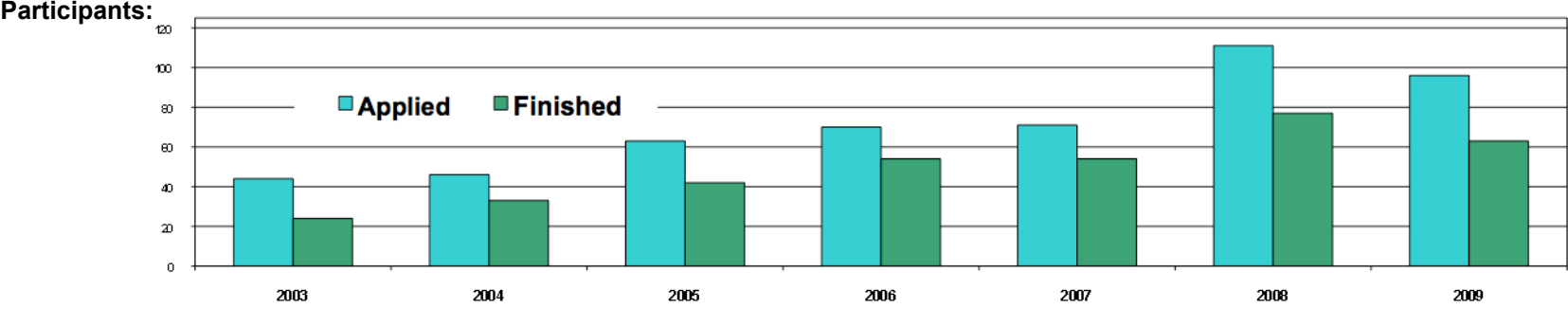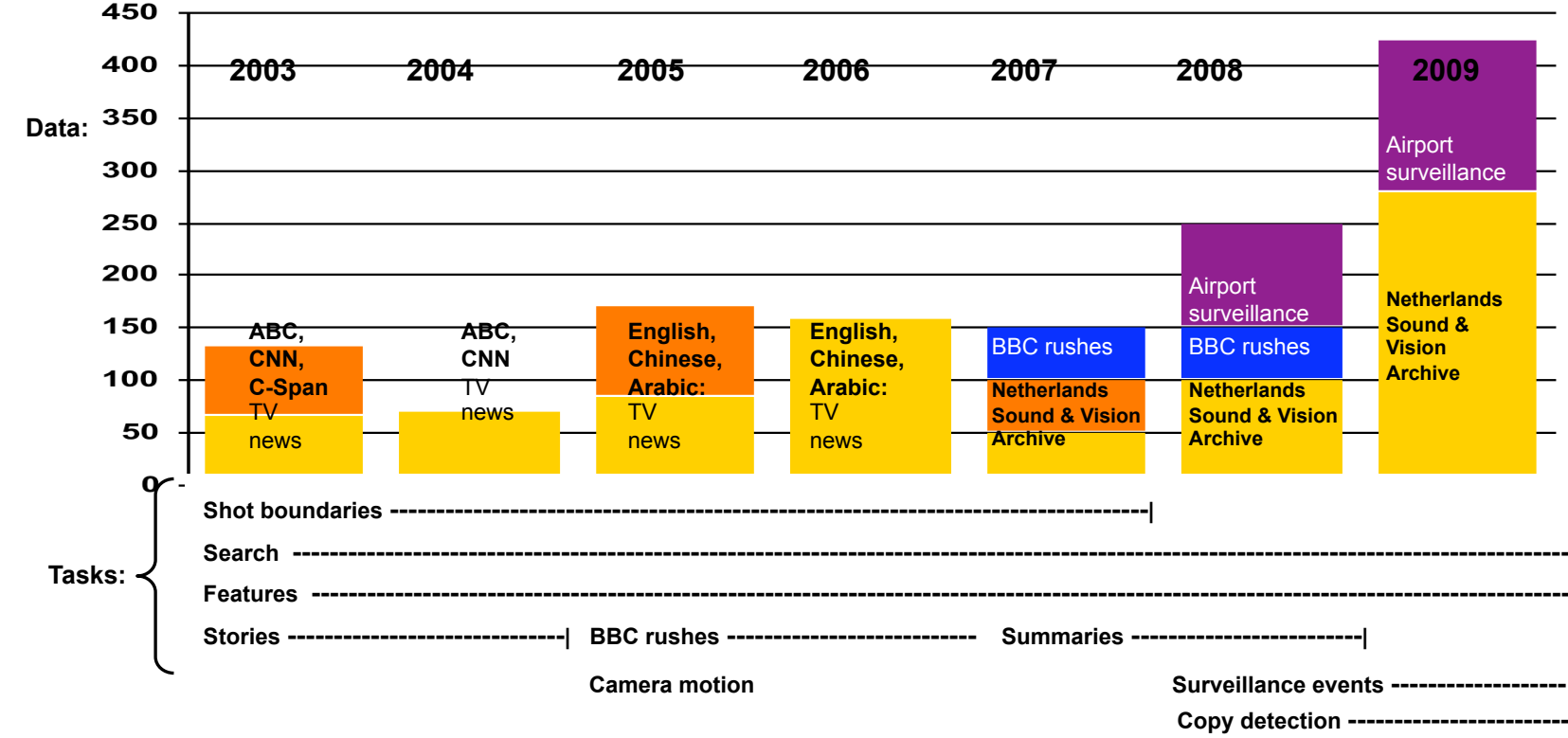- In 2003 TRECVid separated from TREC, started small, grew rapidly;

Major responsibilities

- NIST: Organize data, tasks, and other resources of interest with input from sponsors and participating researchers, define user and system tasks.
- NIST: Define and develop evaluation infrastructure, ground truth, scoring software
- Researchers: create common resources, develop systems, run systems and submit results
- NIST: Evaluate submissions
- NIST & Researchers: Analyze and present results
- NIST: Organize and run annual workshop in mid-November at NIST

Additional contributions (EU)

- NII and the Laboratoire d'Informatique de Grenoble **mirrored** the video data
- Christian Petersohn at the Fraunhofer (Heinrich Hertz) Institute in Berlin provided the **master shot** reference.
- Roeland Ordelman and Marijn Huijbregts at the University of Twente donated the output of their **ASR** system run on the S&V data.
- The LIMSI Spoken Language Processing Group contributed a **speech transcription** of the 2007-2009 S&V data using their Dutch recognizer.
- Christof Monz of Queen Mary, University London contributed **MT** (Dutch to English) for the S&V video.
- For the copy detection task we reused the **video transformation** tools created by Laurent Joyeux for INRIA in 2008.
- Dan Ellis at Columbia U. devised and applied the audio transformations to produce the **audio-only queries** for copy detection.
- The MediaMill team at the University of Amsterdam donated detector scores for **64 concepts** trained over the last 3 years.
- The Columbia U. – City U. Hong Kong team donated **detector scores** for the TRECVID 2009 test data.

TRECVID Evolution

Data:

2003	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**

450
400
350
300 — Airport surveillance (2009)
250 — Airport surveillance (2008)
200
150
100
50
0

- **2003:** ABC, CNN, C-Span TV news
- **2004:** ABC, CNN TV news
- **2005:** English, Chinese, Arabic: TV news
- **2006:** English, Chinese, Arabic: TV news
- **2007:** BBC rushes / Netherlands Sound & Vision Archive
- **2008:** Airport surveillance / BBC rushes / Netherlands Sound & Vision Archive
- **2009:** Airport surveillance / Netherlands Sound & Vision Archive

Tasks:

Shot boundaries --|

Search ---

Features --

Stories -------------------------------| BBC rushes --------------------------- Summaries -------------------------|

Camera motion

Surveillance events --------------------

Copy detection -------------------------

Participants:

2009: Details

- Data:
  - 100+280 hrs - Netherlands Institute for Sound and Vision (S&V)
  - 100+50 hrs of airport surveillance data - UK Home Office

- 4 evaluated tasks
  - S&V news magazine, cultural, educational/entertainment
    1. Content-based copy detection - 1407 video queries, 1407 audio queries, 9849 audio+video queries
    2. High-level feature extraction - 20 features
    3. Search (automatic, manually-assisted, interactive) - 24 multimedia topics
  - Airport surveillance video (5 cameras * 2 hours * 10 days)
    Surveillance event detection

TV2009 Finishers

Aristotle University of Thessaloniki	National Institute of Informatics
Asahikasei Co.	NEC Laboratories America, Inc. and UIUC
AT&T Labs - Research	
Beijing Univ of Posts and Telecom.-MCPRL	NHK Science and Technical Research Laboratories
Beijing Univ of Posts and Telecom.-PRIS	**Oxford/IIIT**
Brno University of Technology	Peking University-PKU-ICST
Budapest Academy of Sciences	Peking University-PKU-IDM
Carnegie Mellon University	Shanghai Jiao Tong University-IICIP
Centre for Research and Technology Hellas	Shanghai Jiao Tong University-IS
Chinese Academy of Sciences-IVA_NLPR_IA_CAS	Simon Fraser University
Chinese Academy of Sciences-MCG-ICT-CAS	Telefonica I+D
	The Institute of Statistical Mathematics
City University of Hong Kong	**The Open University**
Computer Research Institute of Montreal	**TNO**
	Tokyo Institute of Technology
Florida International University	Toshiba Corporation
France Telecom Research & Development-Beijing	Tsinghua University-IMG
Fudan University	Tsinghua University-MPAM
Fuzhou University	TUBITAK UZAY
GDR ISIS - IRIM consortium	**Universidad Autónoma de Madrid**
Helsinki University of Technology TKK	**Universidad Carlos III de Madrid**
IBM Watson Research Center	**University of Amsterdam**
Institut EURECOM	**University of Brescia**
Istanbul Technical University	University of Central Florida
IUPR - DFKI	University of Electro-Communications
JOANNEUM RESEARCH Forschungsgesellschaft mbH-JRS	**University of Glasgow**
KB Video Retrieval	**University of Karlsruhe (TH)**
Kobe University	**University of Marburg**
Lab d'Informatique de Grenoble	University of Ottawa
Lab d'Informatique Fondamentale de Marseille	**University of Surrey**
Laboratoire REGIM	**UPS - IRIT - SAMoVA**
LSIS, Université Sud Toulon Var	Xi'an Jiaotong University
Nanjing University	Zhejiang University

We produce …

Category A results (Top 50)

We produce …

InfAP by feature (top 10 runs)

Legend: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 (Max), Median

1 Classroom   2 Bridge   3 Emergency_Vehicle   4 Dog   5 Kitchen   6 Airplane_flying   7 Two people   8 Bus

9 Driver   10 Cityscape   11 Harbor   12 Telephone   13 Street   14 Demonstration_Or_Protest   15 Hand

16 Mountain   17 Nighttime   18 Boat_ship   19 Flower   20 Singing
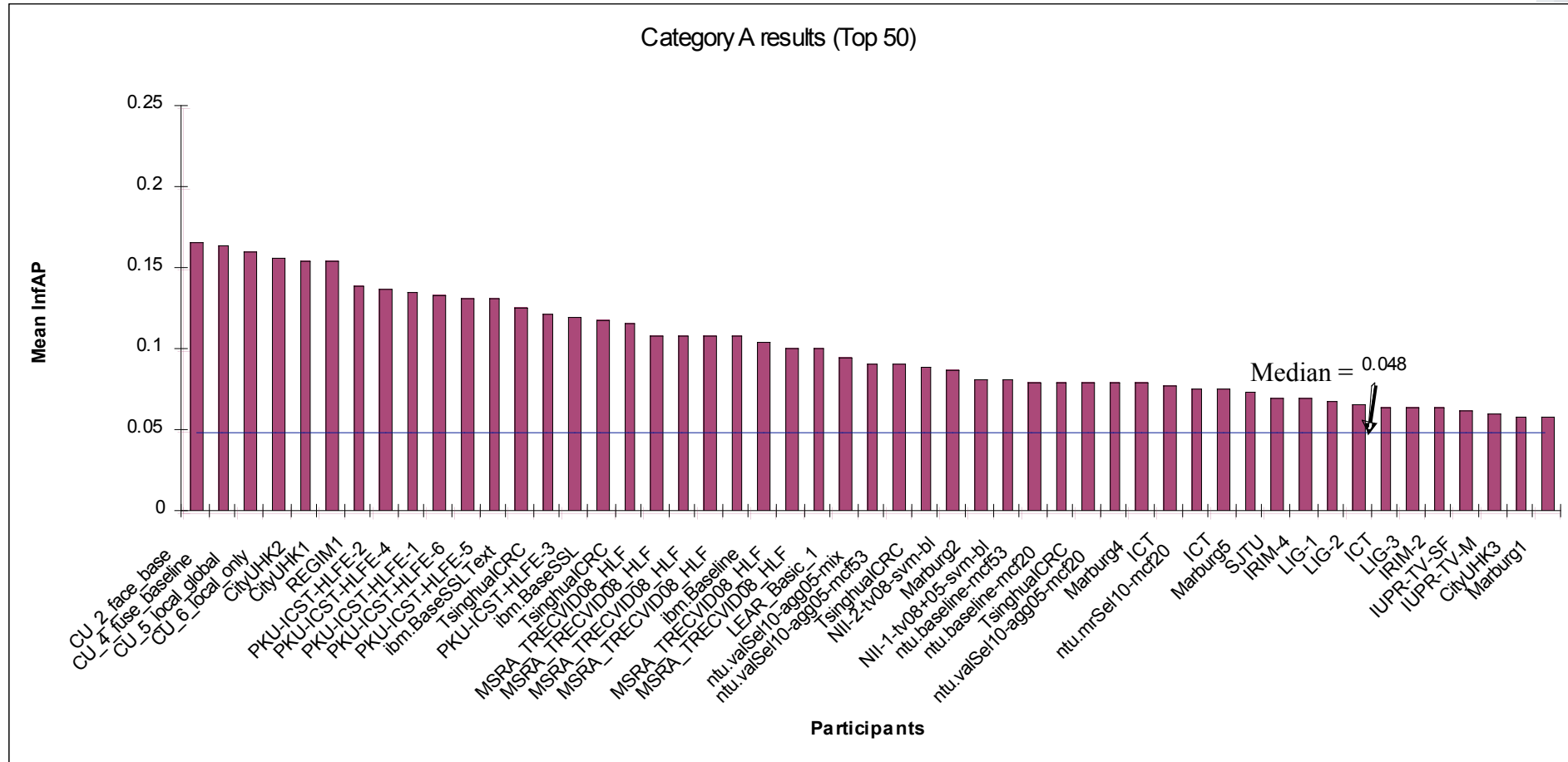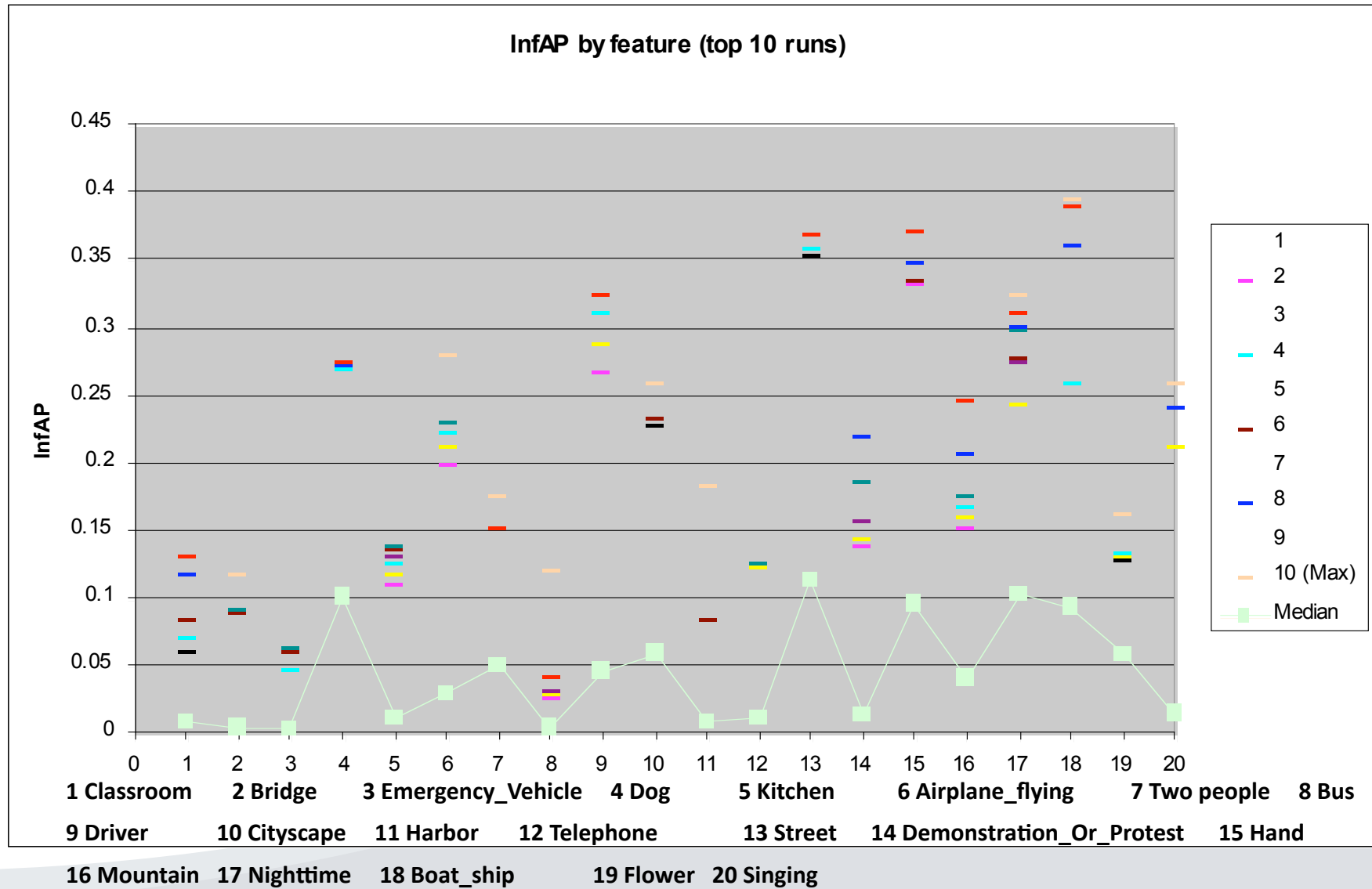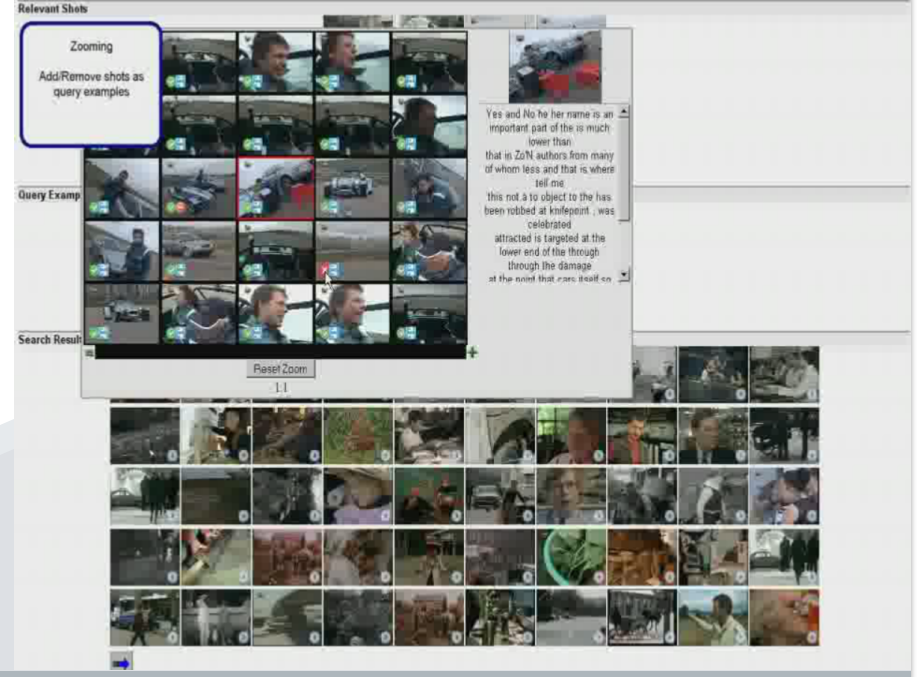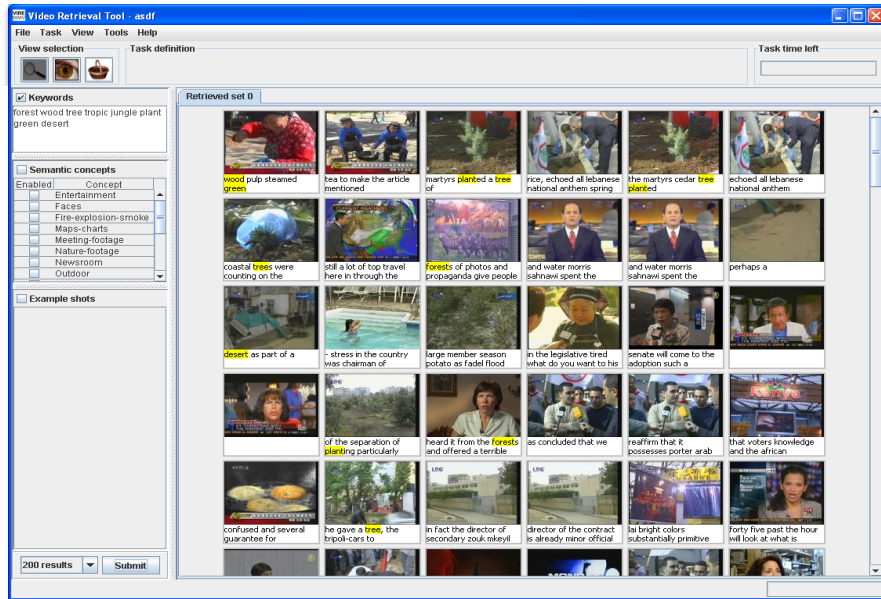
What's TREC/TRECVid impact ?

- Three aspects
  - Human capital
  - Economic impact
  - Scientific impact

- Human impact … since 1991, 200 separate organisations have taken part in 1 or more TREC evaluations.
- In 2008 alone, 400 researchers contributed to TRECVid.
- There are more European participants in TRECVid than from any other region.
- There are literally hundreds of PhD dissertations built around TREC, TRECVid and its derivatives.

Economic Impact

- In 2008 NIST commissioned a retrospective economic impact assessment of TREC, 1991-2009, to quantify direct economic impact.
- Benefits …
  - technical achievements are known .. collections, methodology, scoring, etc.
  - improved IR performance documented elsewhere in scientific literature
  - spawned new IR evaluations like CLEF, TRECVid, NCTIR, INEX, FIRE, Mirex,Amaryllis
  - technology transfer ?

Technology transfer

- From research labs into
  - industry search products and services
  - end users better search experience
- TREC played significant role in improving web search by Google, Yahoo! Microsoft. Fact.
- Search is 215% quicker in 2009 than 1999, of which 32% is enabled by TREC activities
- US users spent 8.6 billion hours searching during 1999-2009 … that means US internet users would have had to spend 3.15B additional hours searching. Try measuring that !

Technology transfer

- But what can we measure ?
- Impact study also quantified private sector and academic/government lab benefits, as perceived by these stakeholders (actual search time saved, re-using TREC data and results) … estimated at US$122M (extrapolated)

- Given TREC cost over 19 years, was US$29M, that's a great benefit-to-cost ratio.

TRECVid's Scientific Impact

- There are a lot of publications from TRECVid ... workshop and external … nobody knows how many, or what they are about;
- Given a definition that it uses TRECVid data, or is based on a TRECVid technique, or compares against TRECVid results …

Statistics

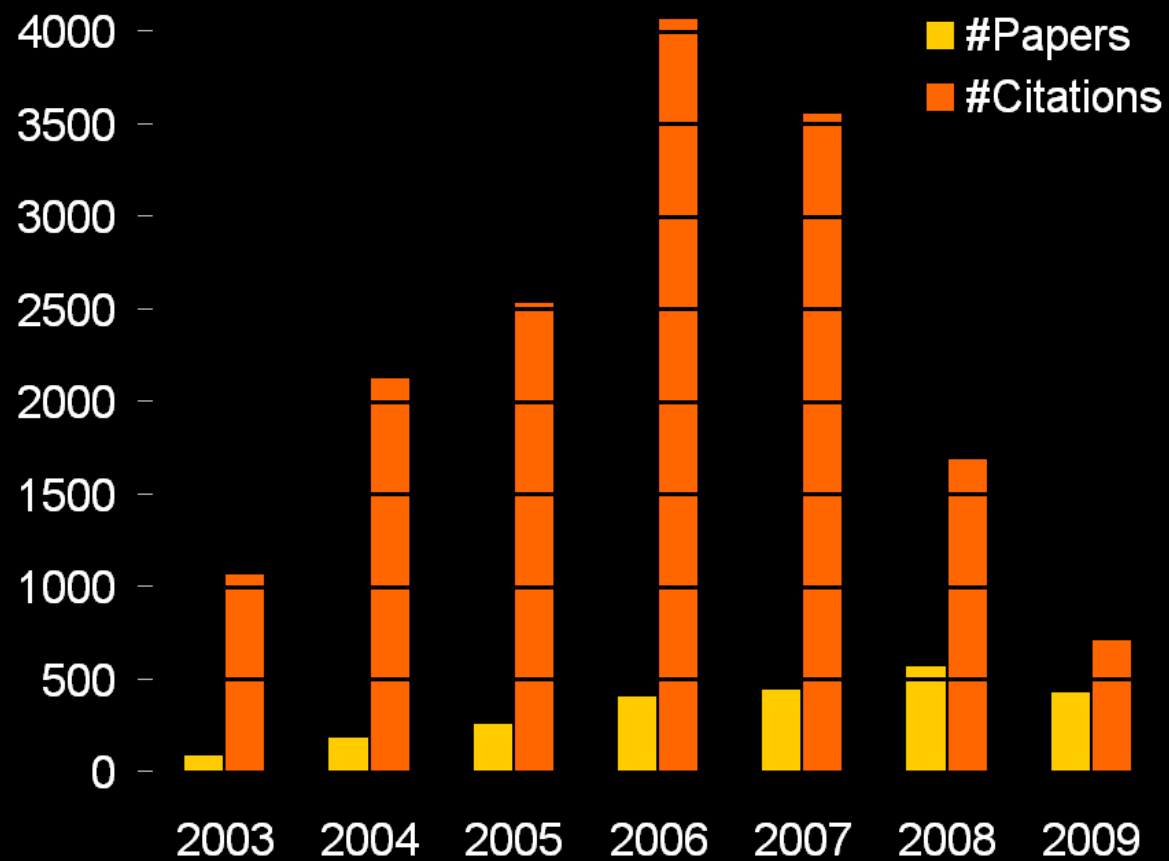|	TVid papers	Non-TVid papers	Citations
2003	28	64	1,066
2004	30	158	2,124
2005	37	225	2,537
2006	50	361	4,068
2007	48	397	3,562
2008	64	509	1,691
2009	53	375	712
Total	**310**	**2,089**	**15,760**

Statistics

- Total 2,399 TRECVid-related papers
  - 310 TVid workshop papers
  - 2,089 non TVid papers
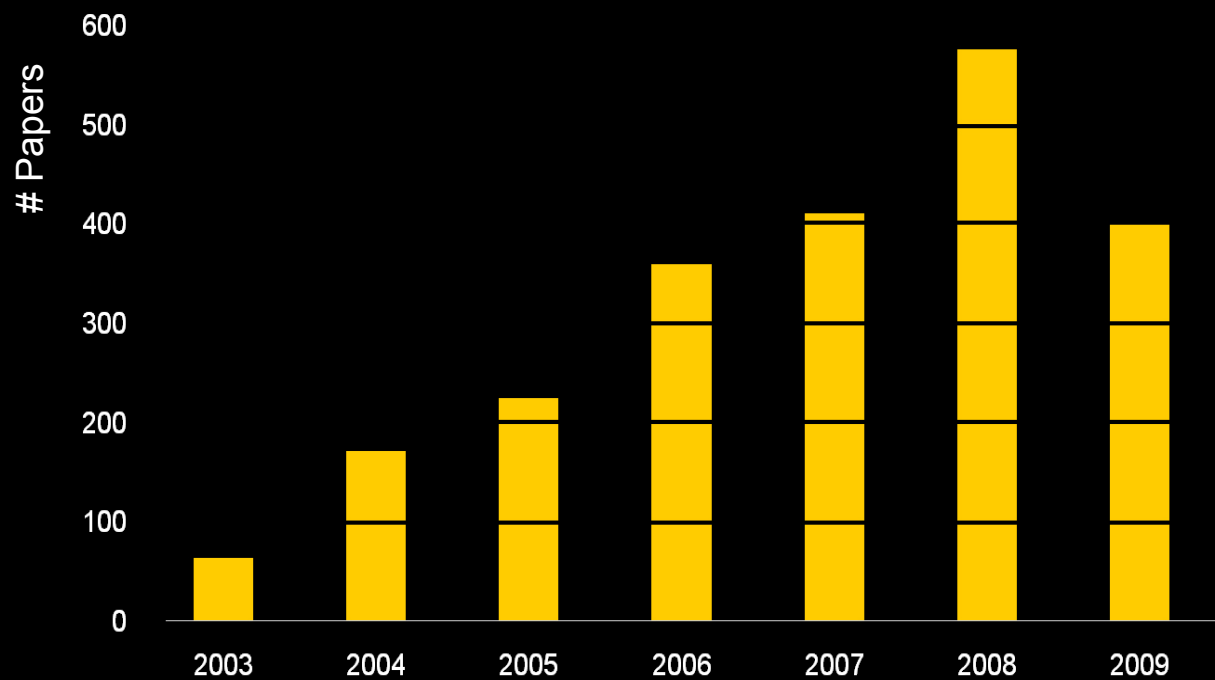
Statistics

- Total 15,760 citations

TREC/TRECVid Future

- Future of TREC is about user behaviour data like social networks, Twitter, geo-based ... and trying to increase industry involvement via data provision, etc.
- TRECVid multimedia search is strong
- Both these are US-funded but EU is strong
- CLEF is Europe's largest IR benchmarking evaluation but its history .. DELOS, DELOS-2, TrebleCLEF, Promise … is of persistent uncertainty, reactive as it adapts to funding instruments
- Little scope for long-term planning, sustainability
- Successful despite the conditions