TRECVID 2017 INSTANCE RETRIEVAL

INTRODUCTION AND TASK OVERVIEW

Wessel Kraaij
The Netherlands Organisation for
Applied Scientific Research TNO; Leiden University

George Awad
Dakota Consulting ; National Institute of Standards and Technology

Asad A. Butt
National Institute of Standards and Technology

Disclaimer
The identification of any commercial product or trade name does not imply endorsement or recommendation by the National Institute of Standards and Technology.

Table of contents

- Task Definition
- Data
- Topics (Queries)
- Participating teams
- Evaluation & results
- General observation

Task

From 2013 – 2015

- The task asked systems to find a specific object, person or location in any context using a small set of image and video examples.

In 2016 - 2017

- A new query type was used: *find a specific person in a specific location.*

System task:

- Given a topic with:
  - 4 example images of the target person
  - 4 Region of Interest (ROI)-masked images of the target person
  - 4 shots from which the target person example images came
  - (6 to 12) image and video examples of a known location
- Return a list of up to 1000 shots ranked by likelihood that they contain the topic target person in the target location
- **Automatic** or **interactive** runs are accepted

Data …

- The British Broadcasting Corporation (BBC) and the Access to Audiovisual Archives (AXES) project made **464 h** of the BBC soap opera EastEnders available for research
  - 244 weekly "omnibus" files (MPEG-4) from 5 years of broadcasts
  - 471527 shots
  - Average shot length: 3.5 seconds
  - Transcripts from BBC
  - Per-file metadata
- Represents a "small world" with a slowly changing set of:
  - People (several dozen)
  - Locales: homes, workplaces, pubs, cafes, open-air market, clubs
  - Objects: clothes, cars, household goods, personal possessions, pets, etc
  - Views: various camera positions, times of year, times of day,
- Use of fan community metadata allowed, if documented

Topic creation procedure @ NIST

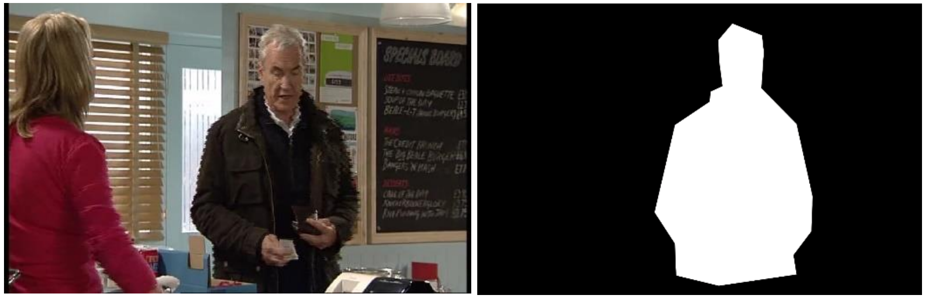
- Viewed several test videos to develop a list of recurring people, locations and their overlapping.
- Chose 10 master locations and identified 6 to 12 image and video examples to each depending on location type (private: kitchen, room, etc; public: pub, café, market, etc)
- Created ≈90 topics targeting recurring specific persons in specific locations.
- Chose representative sample of 30 topics. Each topic includes images for target persons from test videos, many from the sample video (ID 0) and a named location.
- Filtered example shots from the submissions if it satisfies the topic.

Global test condition: type of training data

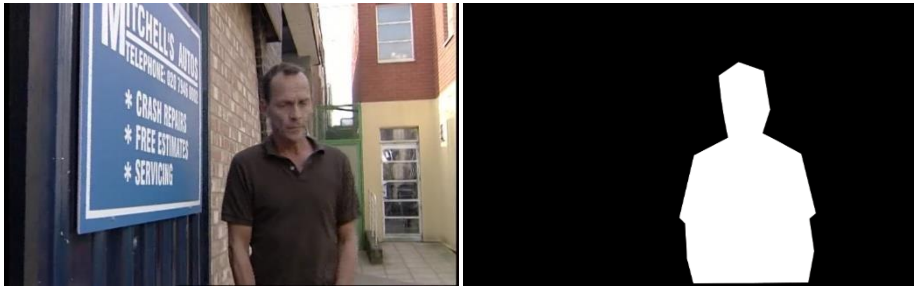
Effect of examples – 2 conditions:

- A – one or more provided images – no video
- E - video examples (+ optionally image examples)

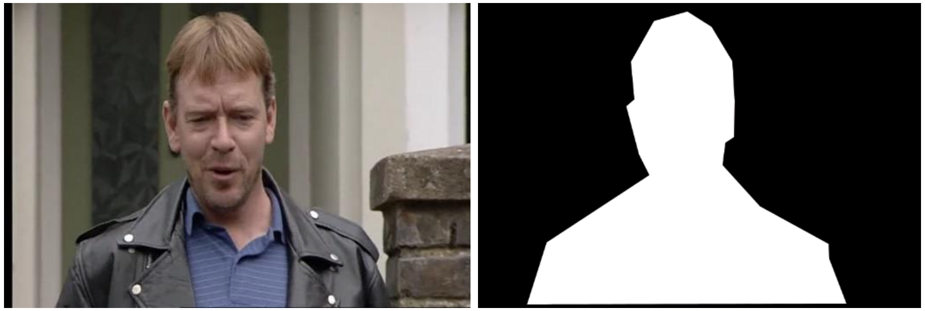
Topics – segmented "person" example images



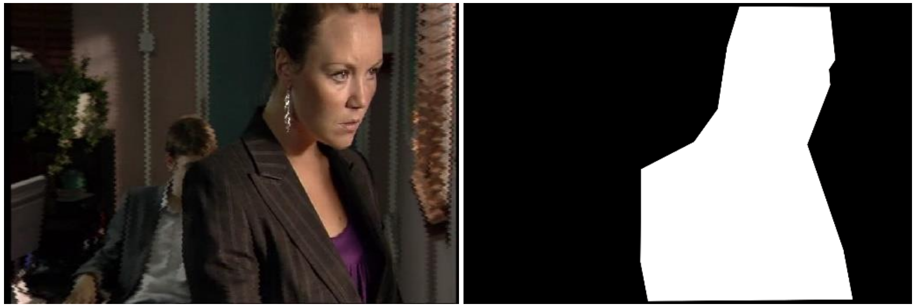
Archie



Billy

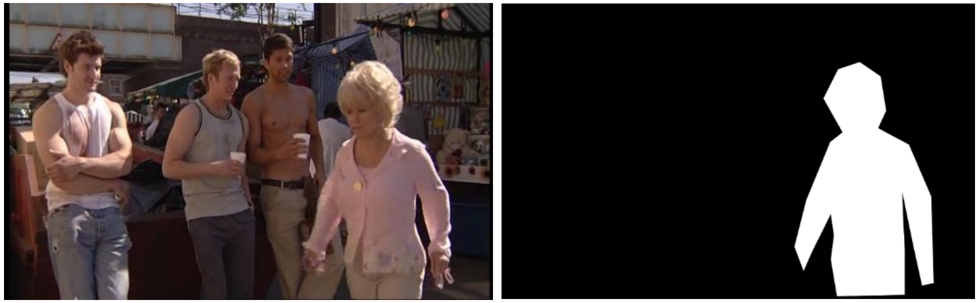


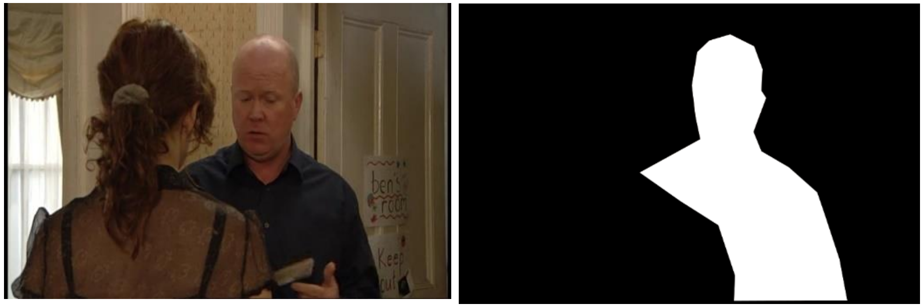
Ian

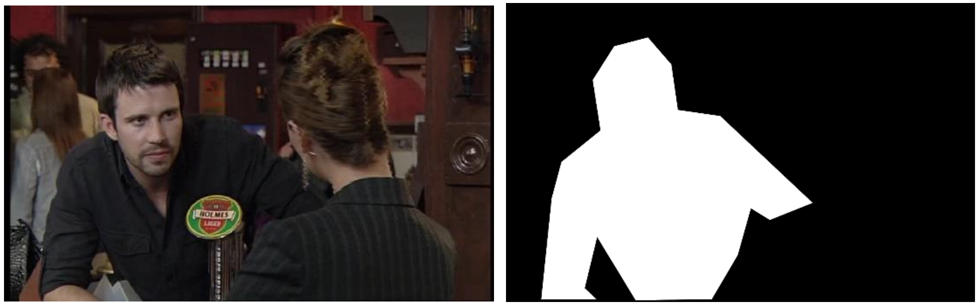


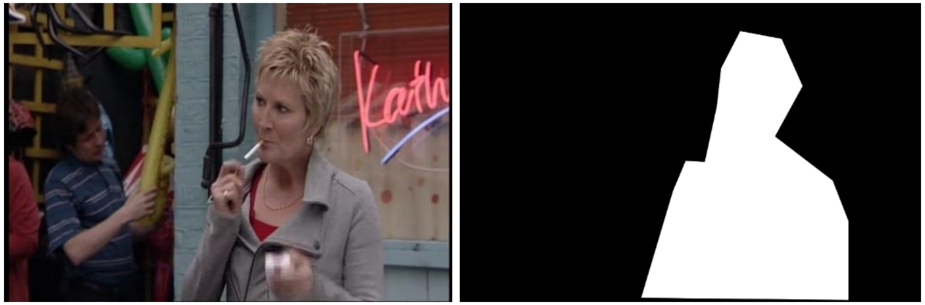
Janine

Topics – segmented example images









Topics – 10 Master locations

Foyer

Kitchen1

Kitchen2

LR1

LR2

Cafe1

Cafe2

Laundrette

market

Pub

NIST
National Institute of Standards and Technology

Topics – 2017

	Peggy	Billy	Ian	Janine	Archie	Ryan	Shirley	Phil
Cafe1	x	x	x	x		x	x	x
Market			x	x	x		x	x
LR2	x	x			x		x	x
Kitchen2	x	x		x		x	x	x
Launderette	x	x	x	x	x	x	x	

30 x topics : find {Peggy, Billy, Ian, Janine, Archie, Ryan, Shirley, Phil} in {Cafe1,Market,LR2,Kitchen2,Launderette}

INS 2017: 8 Finishers (out of 19)

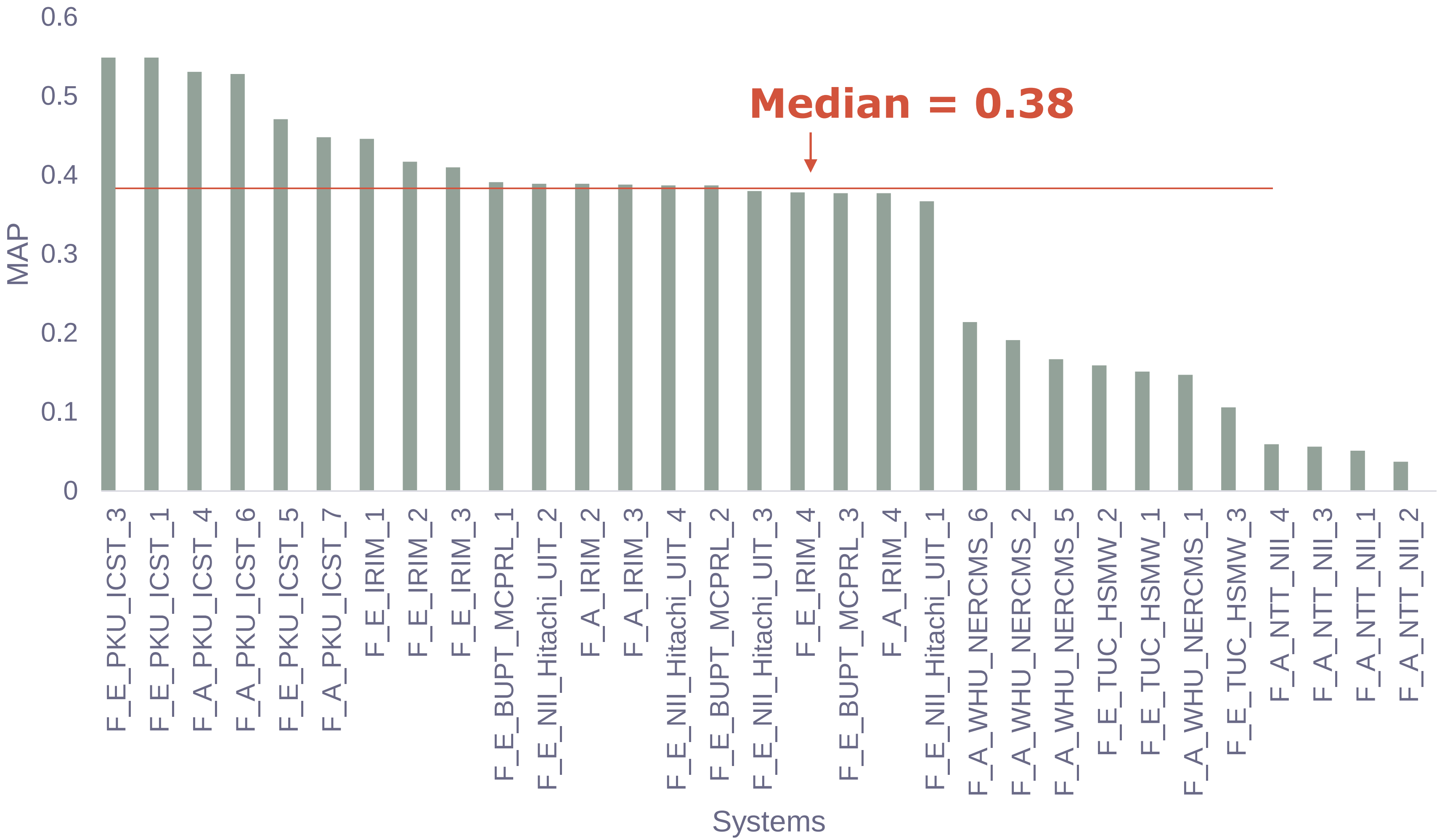
Team	Organization	Run Types Submitted F: automatic, I: Interactive
BUPT_MCPRL	Beijing University of Posts and Telecommunications	F_E (3), I_E (1)
TUC_HSMW	Chemnitz University of Technology, University of Applied Sciences Mittweida	F_E (3), I_E (1)
ITI_CERTH	Information Technologies Institute, Centre for Research and Technology Hellas	I_A (1)
IRIM	EURECOM; LABRI ; LIG ; LIMSI; LISTIC	F_A (3), F_E (4)
NII_Hitachi_UIT	National Institute of Informatics, Japan (NII); Hitachi, Ltd; University of Information Technology, VNU-HCM, Vietnam (HCM-UIT)	F_E (4)
WHU_NERCMS	National Engineering Research Center for Multimedia Software, Wuhan University	F_A (4) , I_A (4)
NTT_NII	NTT Communication Science Laboratories, National Institute of Informatics	F_A (4)
PKU_ICST	Peking University	F_A (3), F_E (3), I_E (1)

Evaluation

For each topic the submissions were pooled and judged down to at least rank 100 (on average to rank 247, max 520), resulting in 75 165 judged shots (≈ 370 person-h).

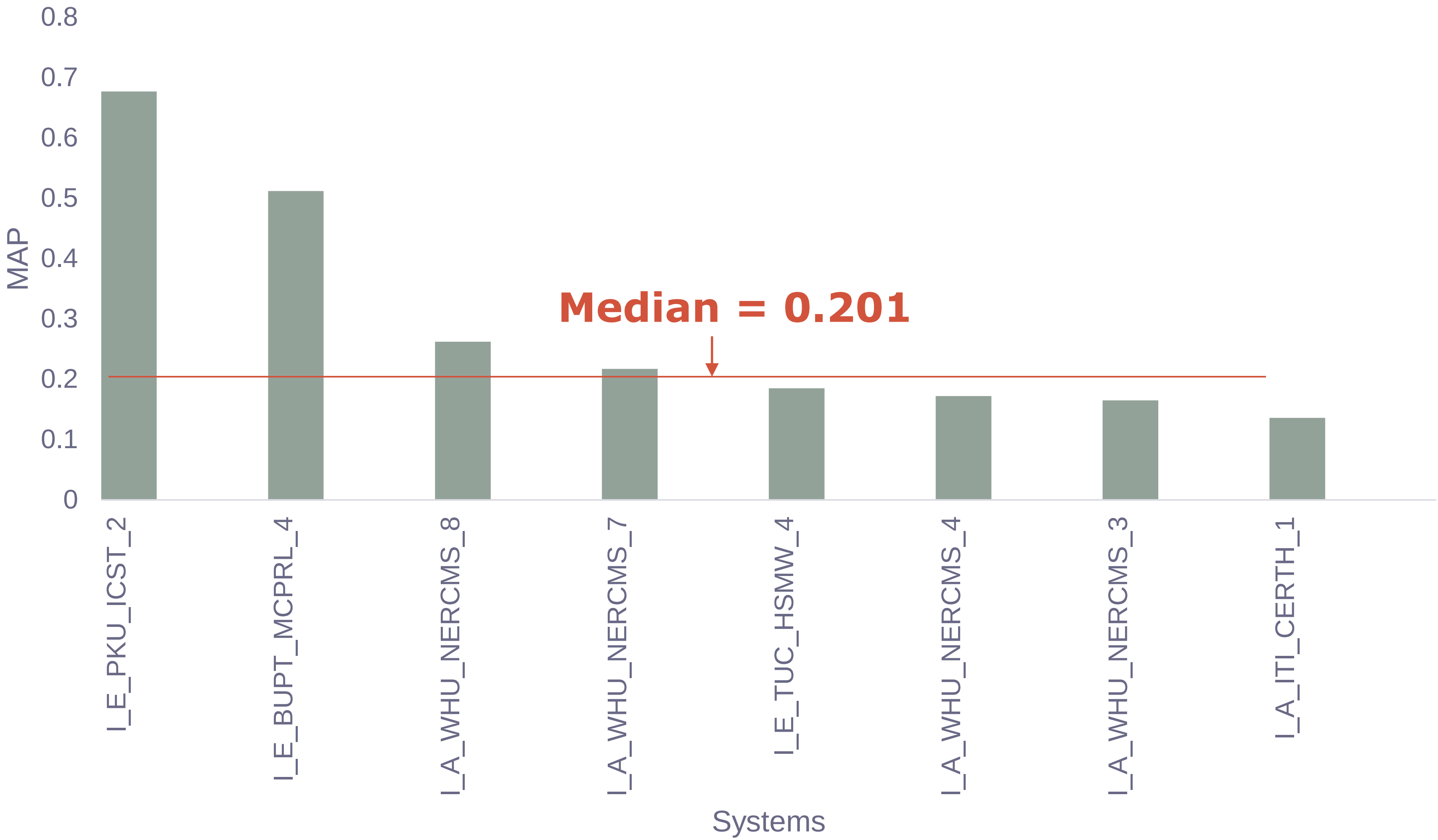
- 10 NIST assessors played the clips and determined if they contained the topic target or not.
- 10 604 clips (avg. 353 / topic) contained the topic target (14 %)
- True positives per topic: min 15 med 179 max 1771
- The task is treated as a form of ranking and thus the trec_eval_video tool was used to calculate average precision, recall, precision, etc.
- To measure efficiency, speed was also measured.

Results by team (Automatic)

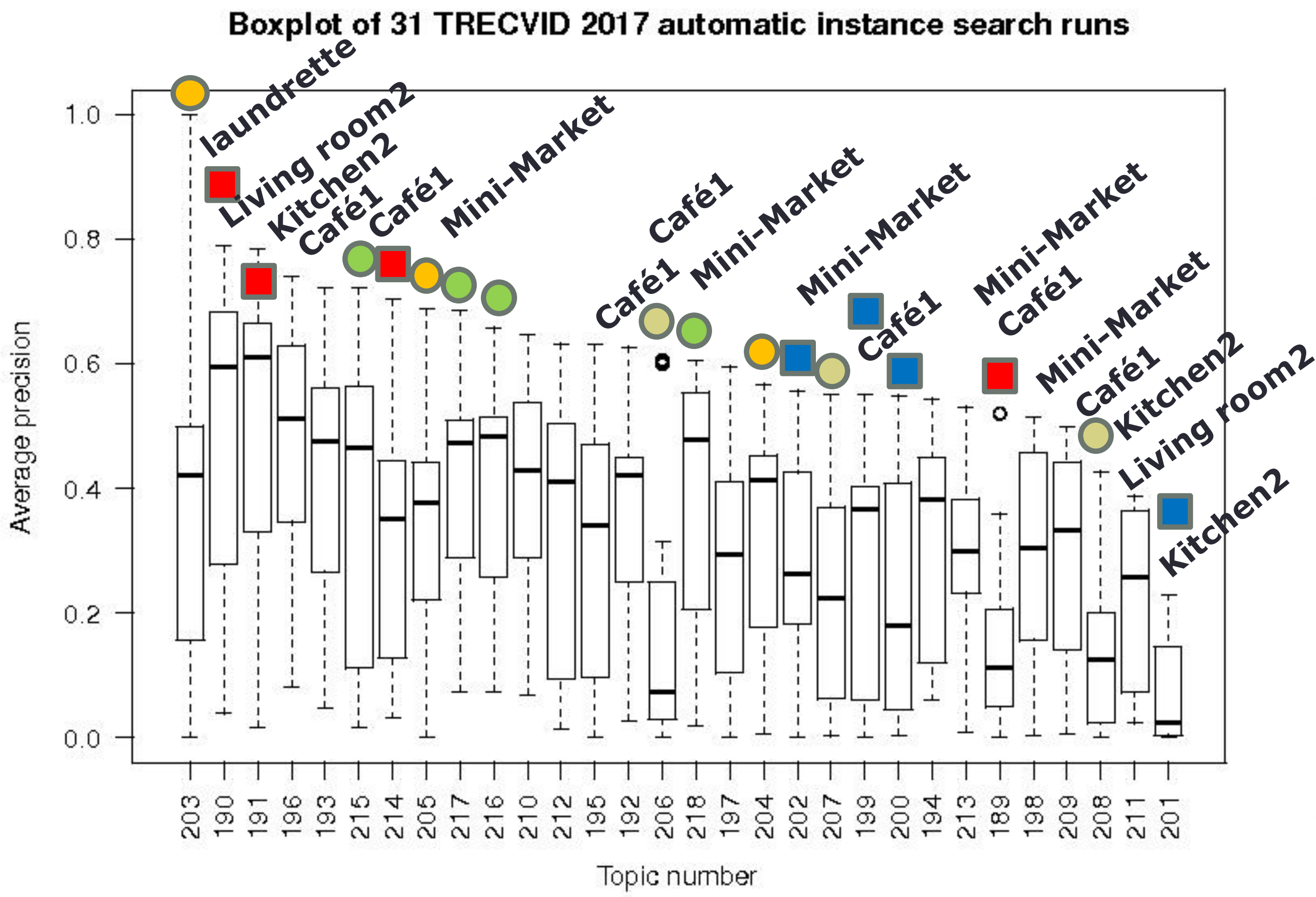


National Institute of Standards and Technology

Results by team (Interactive)



Results by topic - automatic



#	Query
203	Find Archie in this Laundrette
190	Find Peggy in this LivingRoom 2
191	Find Peggy in this Kitchen 2
196	Find Ian at this Cafe 1
193	Find Billy in this Laundrette
215	Find Phil in this Cafe 1
214	Find Peggy in this Laundrette
205	Find Archie in this Mini-Market
217	Find Phil at this Kitchen 2
216	Find Phil in this Living Room 2
210	Find Shirley in this Laundrette
212	Find Shirley in this Kitchen 2
195	Find Billy in this Kitchen 2
192	Find Billy in this Cafe1
206	Find Ryan in this Cafe 1
218	Find Phil in this Mini-Market
197	Find Ian in this Laundrette
204	Find Archie in this Living Room 2
202	Find Janine in this Mini-Market
207	Find Ryan in this Laundrette
199	Find Janine in this Cafe 1
200	Find Janine in this Laundrette
194	Find Billy in this Living Room 2
213	Find Shirley in this Mini-Market
189	Find Peggy in this Cafe1
198	Find Ian in this Mini-Market
209	Find Shirley in this Cafe 1
208	Find Ryan in this Kitchen 2
211	Find Shirley in this Living Room 2
201	Find Janine in this Kitchen 2

What is the effect of person vs location on the performance ?

- **Mini-Market is hard**
- **Archie🟠, Peggy🟥, and phil🟢 are easy**
- **Janine🟦 and Ryan● are hard**

Automatic Run results + Randomization testing

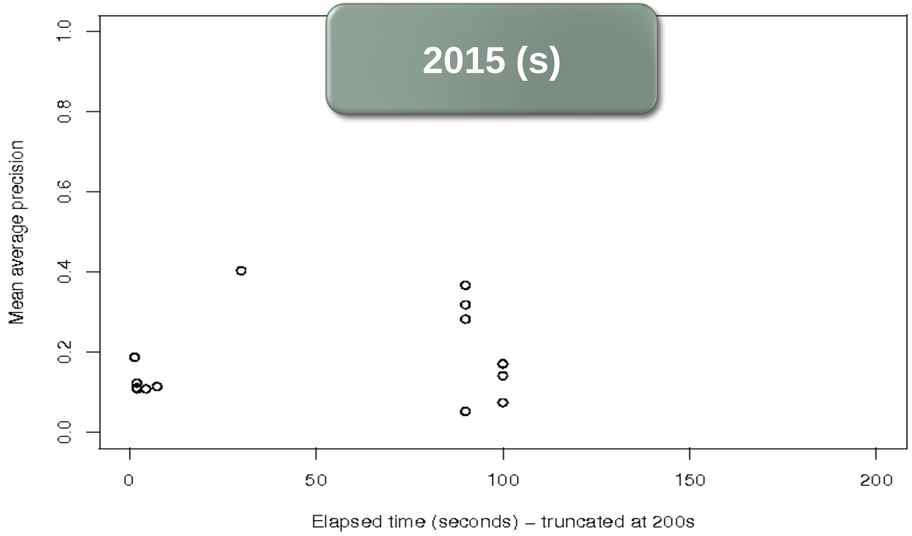
MAP **Top 10 runs across all teams (automatic)**

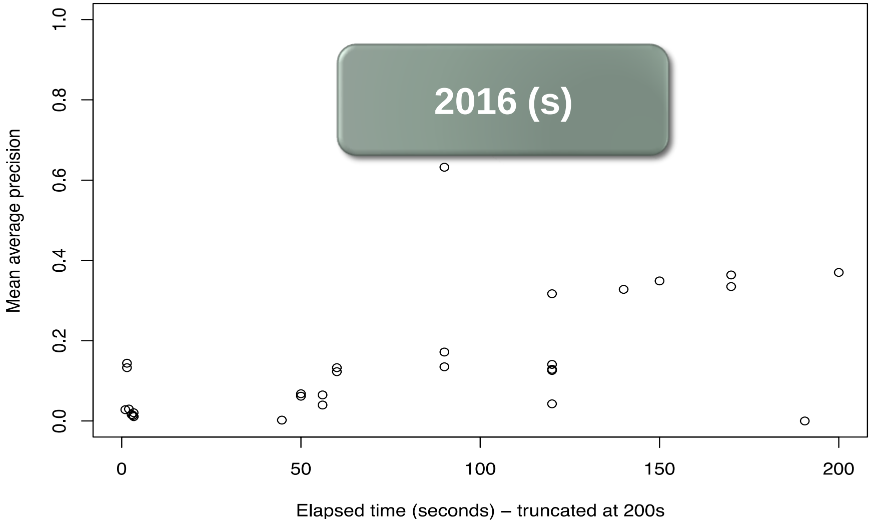
0.549	F_E_PKU_ICST_3	=		>	>	>	>	>	>	>	>
0.549	F_E_PKU_ICST_1		=	>	>	>	>	>	>	>	>
0.531	F_A_PKU_ICST_4			=	>	>	>	>	>	>	>
0.528	F_A_PKU_ICST_6				=	>	>	>	>	>	>
0.471	F_E_PKU_ICST_5					=			>	>	>
0.448	F_A_PKU_ICST_7						=				>
0.446	F_E_IRIM_1							=	>	>	>
0.417	F_E_IRIM_2								=	>	>
0.410	F_E_IRIM_3									=	
0.391	F_E_BUPT_MCPRL_1										=
		1	2	3	4	5	6	7	8	9	10

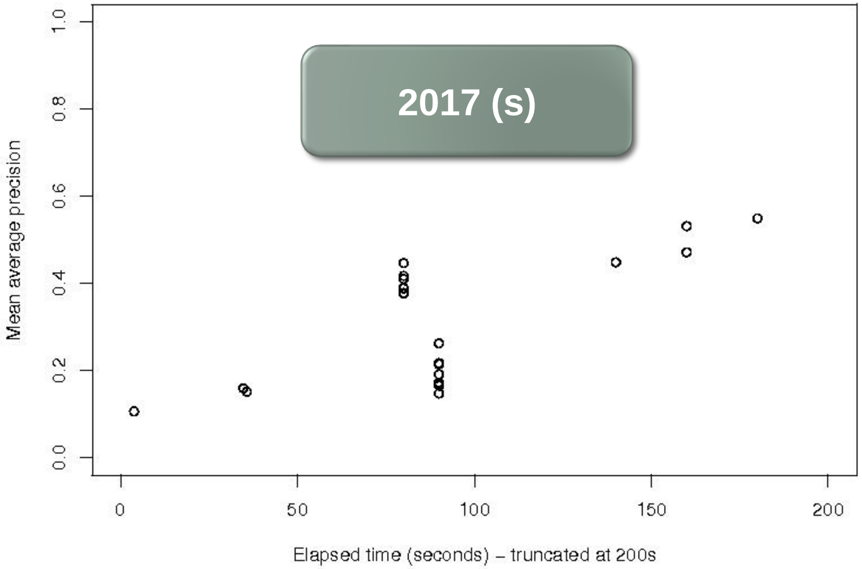
p = probability the row run scored better than the column run due to chance

> $p < 0.05$

Mean Average Precision vs. per query clock processing time (automatic)

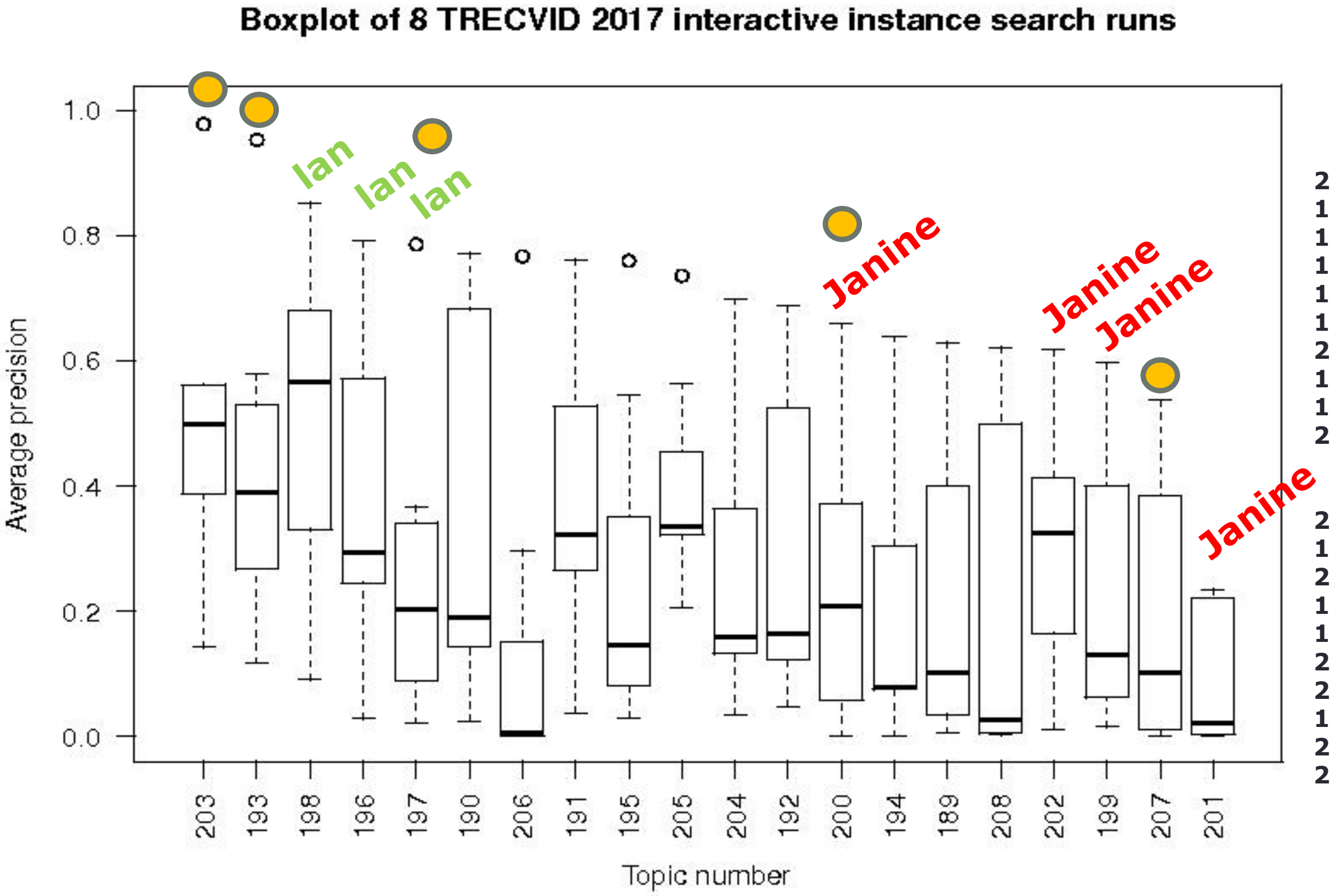






NIST
National Institute of Standards and Technology

Results by topic - interactive



#	Query
203	Find Archie in this Laundrette
193	Find Billy in this Laundrette
198	Find Ian in this Mini-Market
196	Find Ian at this Cafe 1
197	Find Ian in this Laundrette
190	Find Peggy in this LivingRoom 2
206	Find Ryan in this Cafe 1
191	Find Peggy in this Kitchen 2
195	Find Billy in this Kitchen 2
205	Find Archie in this Mini-Market
204	Find Archie in this Living Room 2
192	Find Billy in this Cafe1
200	Find Janine in this Laundrette
194	Find Billy in this Living Room 2
189	Find Peggy in this Cafe1
208	Find Ryan in this Kitchen 2
202	Find Janine in this Mini-Market
199	Find Janine in this Cafe 1
207	Find Ryan in this Laundrette
201	Find Janine in this Kitchen 2

● Laundrette

NIST
National Institute of Standards and Technology

Interactive Run Results, Randomization testing

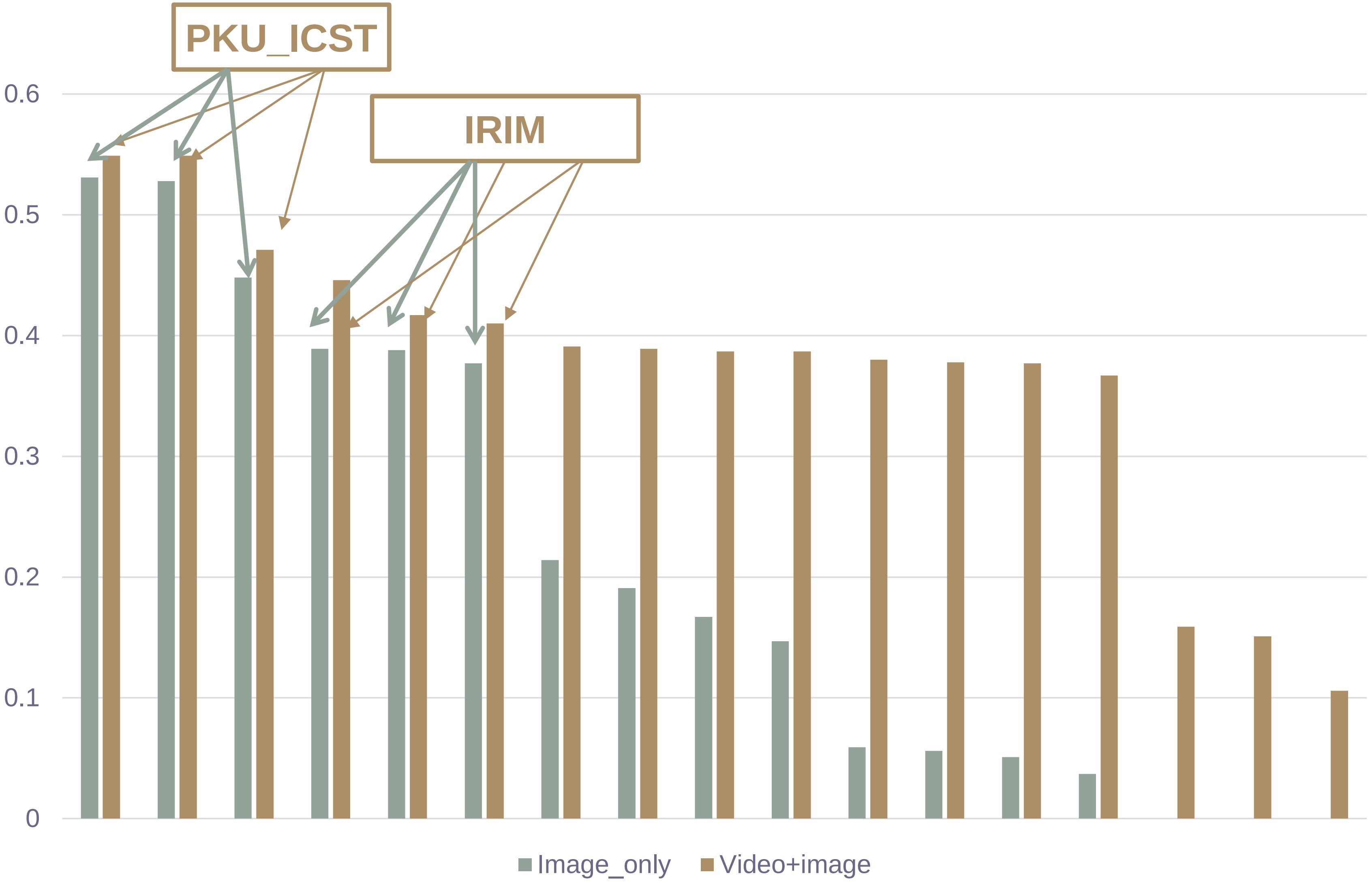
ALL 8 runs by all teams (interactive)

MAP		1	2	3	4	5	6	7	8
0.677	I_E_PKU_ICST_2	=	>	>	>	>	>	>	>
0.512	I_E_BUPT_MCPRL_4		=	>	>	>	>	>	>
0.262	I_A_WHU_NERCMS_8			=	>		>	>	>
0.217	I_A_WHU_NERCMS_7				=		>	>	>
0.185	I_E_TUC_HSMW_4					=			
0.172	I_A_WHU_NERCMS_4						=		
0.165	I_A_WHU_NERCMS_3							=	
0.136	I_A_ITI_CERTH_1								=

p = probability the row run scored better than the column run due to chance

> $p < 0.05$

Results by example set (A/E) - automatic



NIST
National Institute of Standards and Technology

Some general observations about the task

- Decrease in number of participants and stable % of finishers
  - BBC worked on fixing data permissions issues ☺.
- Task guidelines were updated to become more clear about what is allowed for task categories
- More teams are using E condition - training with video examples – (e.g tracking characters)
- Interactive search task:
  - Limited participation
- Second year: Performance is better than 1st year

NII Hitachi UIT

- Challenge 1: improve precision of face recognition:
  - Choose second highest face score in top ranked key frames as hard negative
  - RBF kernel instead of linear kernel for SVM
- Challenge 2: improve recall with scene tracking:
  - For each shot in top 100
    - Scan back and forward to track and re-identify the person

- Submitted 4 runs
- Experiment with name mention in transcript (no gain)

ITI CERTH

- Focus on interactive task
- VERGE system includes several modes for navigation:
  - Visual similarity (DCNN)
  - 346 visual concepts (SIN)
  - Face detection
  - Scene similarity
- Late fusion of DCNN face descriptors and scene descriptors
- Submitted 1 interactive run
- Hypothesis: performance is limited by sub-optimal face detector

NTT

- Location search based on Aggregated Selective Match Kernel [Tolias et al 2013]
- Person search based on OpenFace (limited to frontal faces)
- Fusion based on ranks or scores
- Submitted 4 automatic runs. Submission type ‘A’
- Results were influenced by limitations of OpenFace

WHU-NERCMS

- Components
  1. Filter to delete irrelevant shots
  2. Person search based on face recognition and speaker identification
  3. Scene retrieval based on landmarks and CNN features
  4. Fusion based on multiplying scores
- New for TV17: scene retrieval and Gaussian shape expansion module
- Submitted 4 automatic and 4 interactive runs
- Analysis:
  - scene retrieval is limited by pre-trained CNN
  - Gaussian Shape Expansion methods is successful

Overview of submissions (1)

- 8 out of 8 teams described Instance Search runs for the TV notebook
- 4 teams will present their INS experiments

9:20 - 9:40, BUPT-MCPRL@TRECVID 2017: Instance Search (**BUPT_MCPRL - Beijing University of Posts and Telecommunications**)
9:40 - 10:00, PKU_ICST at TRECVID 2017: Instance Search Task (**PKU_ICST - Peking University**)
10:00 - 10:20, TUC+HSMW at TRECVID Instance Search 2017 (**TUC_HSMW - Chemnitz University of Technology University of Applied Sciences Mittweida**)

10:20 - 10:40, **Break** with refreshments

10:40 - 11:00, IRIM at TRECVID 2017: Instance Search (**IRIM - EURECOM; LABRI; LIG; LIMSI;LISTIC**)
11:00 - 11:20, INS Discussion