



LIG at NTCIR-12 Lifelog Semantic Access Task (LSAT)

Bahjat Safadi

Philippe Mulhem

Georges Quénot

Jean-Pierre Chevallet

MRIM group – LIG laboratory - Grenoble



AUVERGNE – Rhône-Alpes

Outline

1. Overview
2. Index generation
 - Visual Indexing
 - Temporal Indexing
 - Log Indexing
3. Query processing
4. Official results
5. Conclusion
6. Perspectives

1. Overview (1/2)

- To retrieve specific moments in lifeblogger's life based on

- Time/Space/Activity search criteria

- Difficulty

Availability



Accessibility

- Data

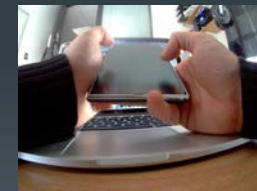
- Images

- Visual (ImageNet) concepts

- Semantic content

(location/activity/time) from sensors

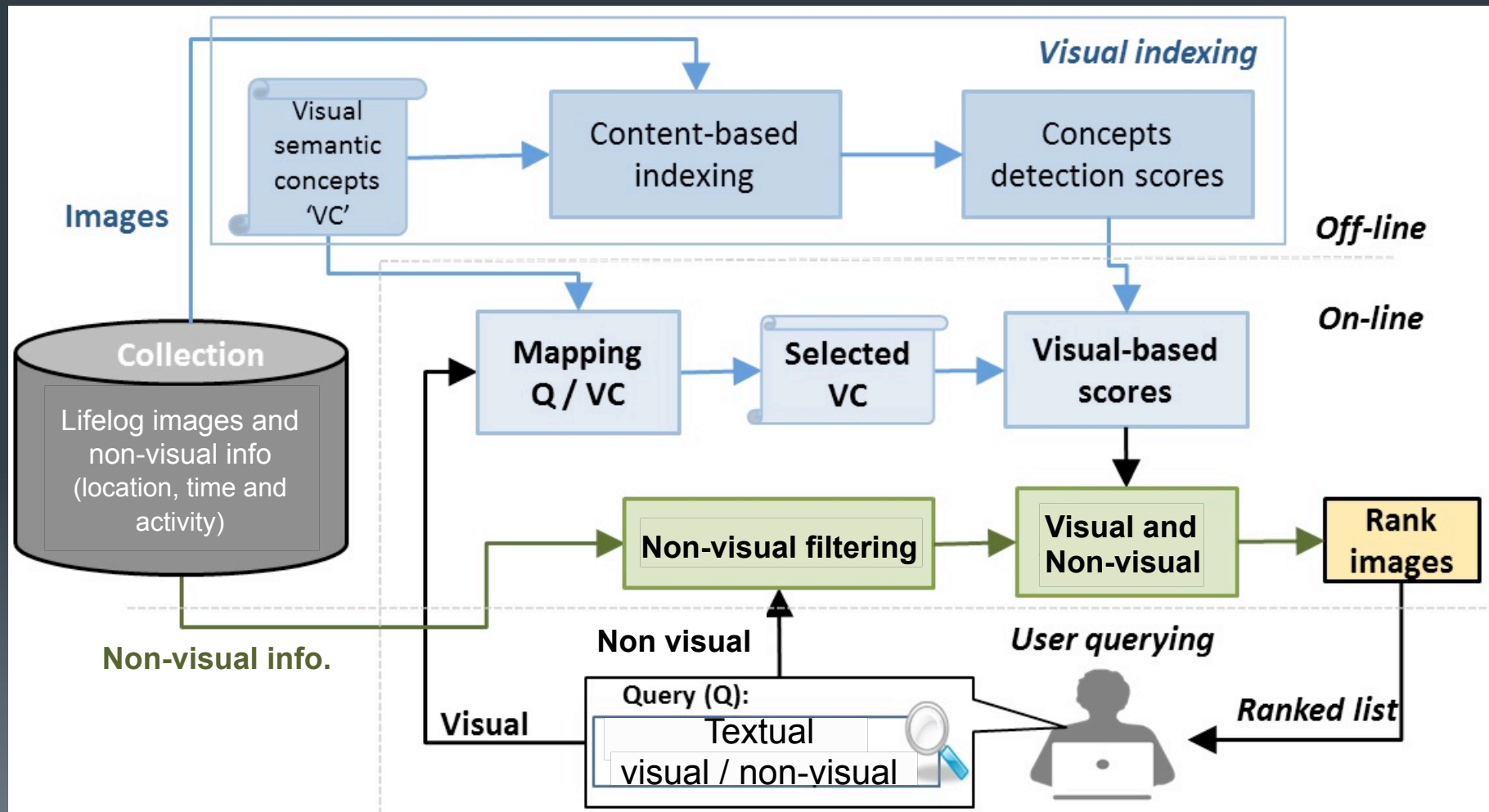
- No training set



Day :
2015-02-18
Time : 03:19 PM
Location : Work

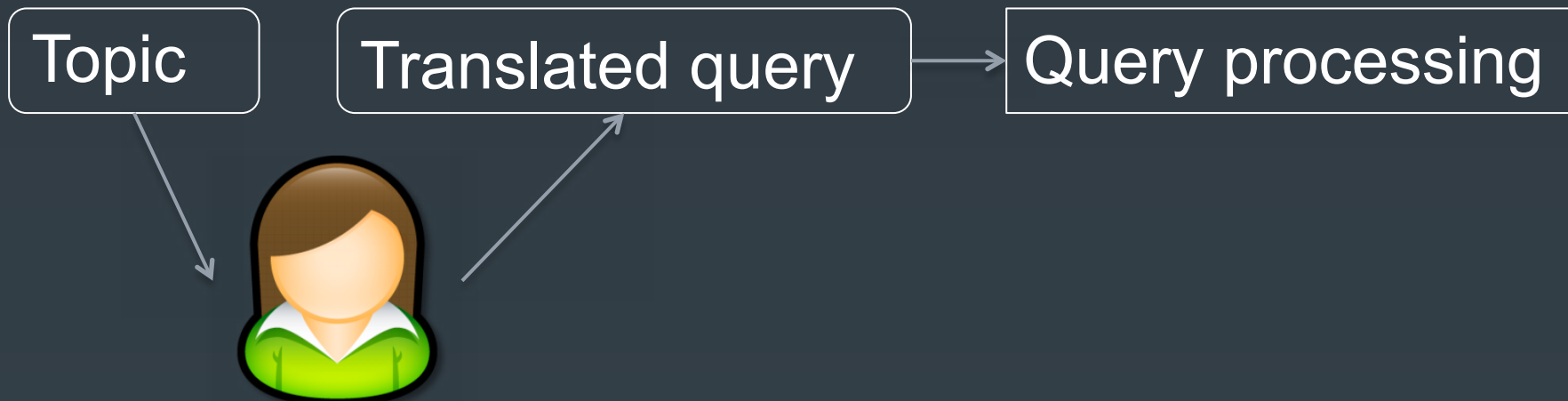
1. Overview (2/2)

- 2 facets processed : Visual and Non-visual



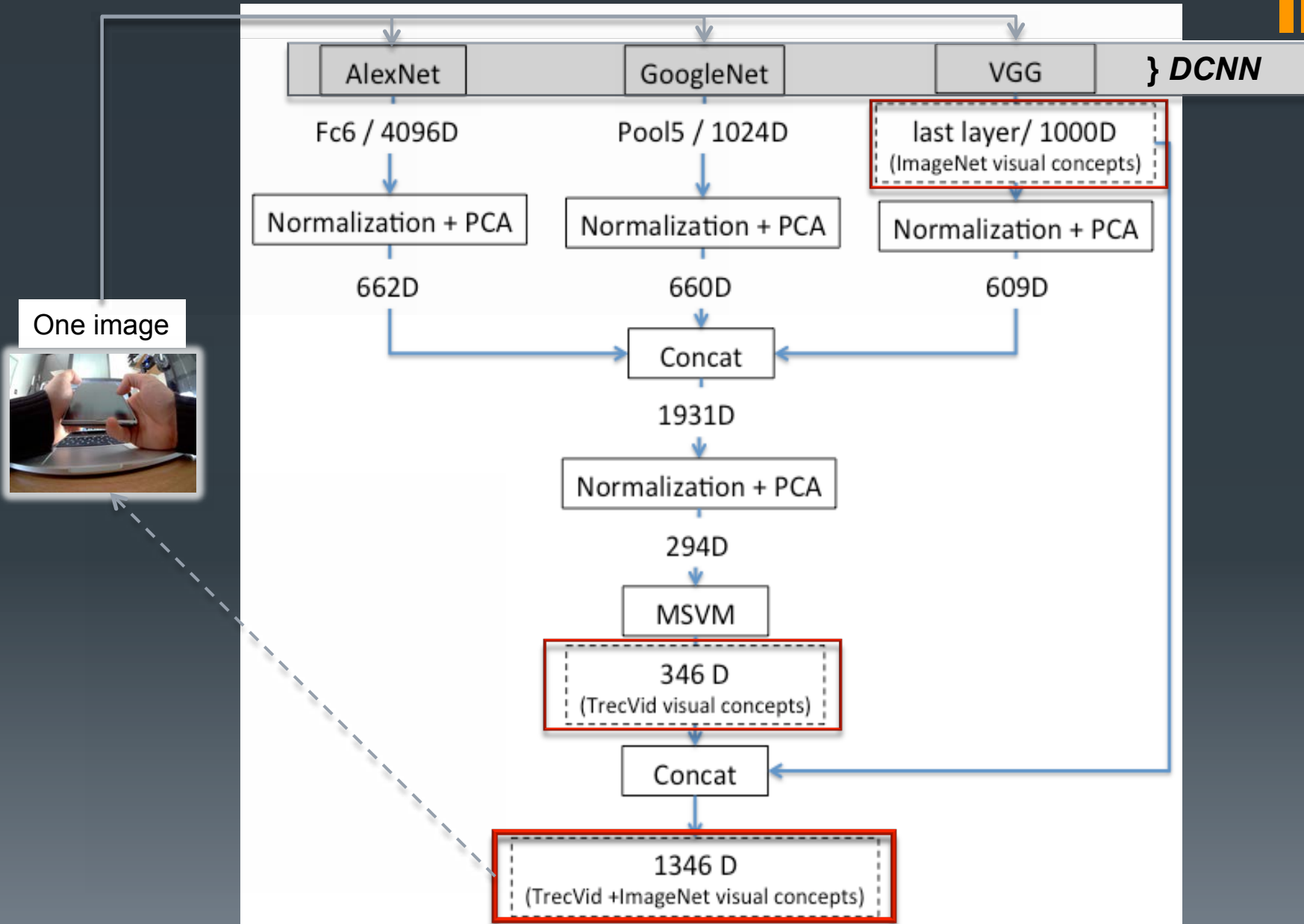
2. Index generation

- LSAT topic to query mapping



- Metadata (non-visual)
 - Temporal terms
 - Activity
 - Location
- Content-Based data
 - Visual concepts (ImageNet, TRECVID)

2.1. Visual indexing



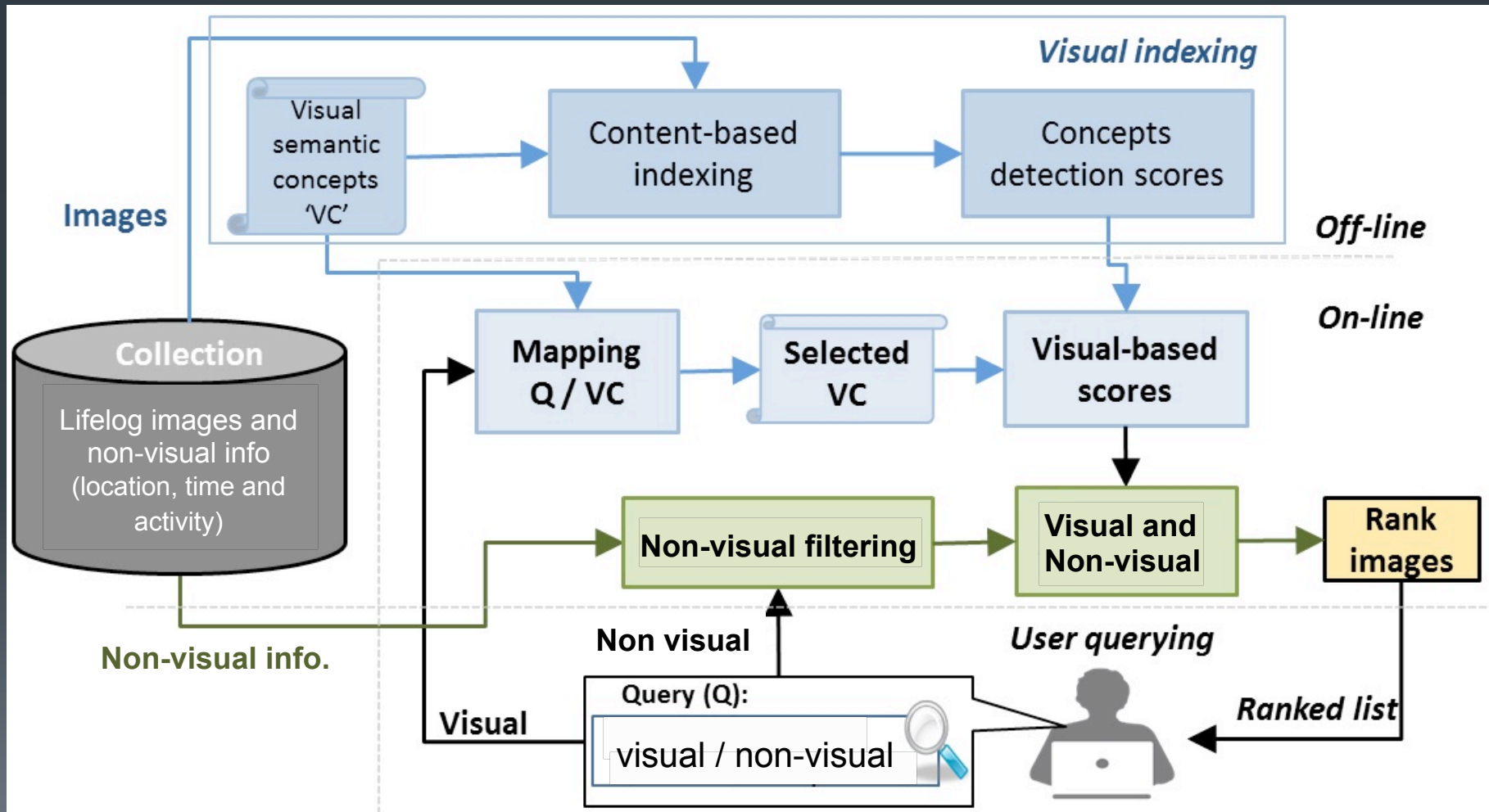
2.2. Temporal & Log indexing

- Time: Activity-related representation of time of the day

Time slot	Days	name
21:00 PM - 5:00 AM	All	night
5:00 AM - 7:15 AM	All	early morning, breakfast
7:30 AM - 11:30 AM	All	morning
11:30 AM - 2:00 PM	All	lunch
2:00 PM - 17:30 PM	All	afternoon
17:30 PM - 20:00 PM	All	early evening
20:00 PM - 23:00 PM	All	late evening
7:30 AM - 9:00 AM	Mon-Fri	trip from home to work
18:15 PM - 18:45 PM	Mon-Fri	trip from work

- Log: Store the location and activity log data from the sensors
 - Ex. home, work, walking...

3. Query processing (1/3)



3. Query processing (2/3)

- Examples:

- Topic001 – visual only

- The Red Taxi (when I boarded a Red taxi, only to get out again shortly afterwards)*

- Mapping:

- Visual: *Taxi* → TRECVID CONCEPT Taxi_Cab

- Topic015 – visual & non-visual

- The Rugby Match (when I was watching rugby football on a television when not at home)*

- Mapping:

- Visual:

- *Television* → TRECVID CONCEPT Computer_Or_Television_Screens – Computers
 - Activity: *static* → not in the predefined list of activities (not walking, no transport, no bus)
 - Location: *not at home* → location unknown or Café

- Topic025 – visual & non-visual

- Cycling Home (when I was cycling home from work)*

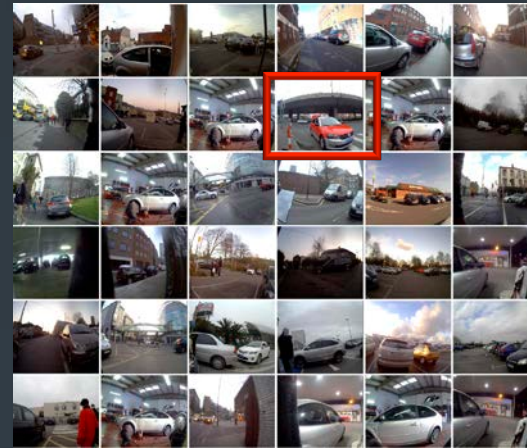
- Mapping:

- Visual: *Cycling* → TRECVID CONCEPT Bicycling
 - Activity: *Cycling* → cycling
 - Time: *from work* → trip from work (6:15PM-6:45PM, Mon_Fri)

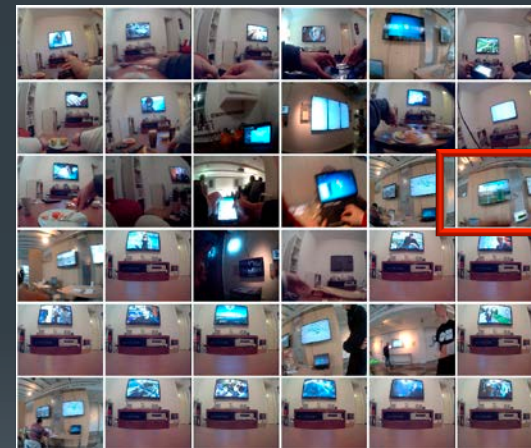
3. Query processing (3/3)

- Results:

- Topic001 – The Red Taxi
 - Event-based nDCG=0.32



- Topic015 –The Rugby Match
 - Event-based nDCG=0.14



- Topic025 – Cycling
 - Event-based nDCG=0.00
 - Need for adjudication ?



4. Official results

- Comparison of several runs (official evaluation)

metric	nDCG		mAP	
	Event	Image	Event	Image
Official submission	0.3896	0.2455	0.2940	0.1667
Without TRECVID	0.2841	0.1808	0.2069	0.1249
Without metadata	0.2955	0.1509	0.1948	0.0916
Metadata only	0.0753	0.0783	0.0490	0.0561

- Need for many visual concepts (ImageNet + TRECVID)
- Need of metadata
- Metadata only not enough

5. Conclusion

- We proposed a simple use of
 - Visual concepts from Deep Learning
 - Conceptualized temporal intervals
 - Available metadata
- The mapping of initial query is manual
 - Many efforts to map query terms into visual/temporal/time concepts
- The results obtained are promising, but certainly below human expectations (Precision = 0.4 at 0.0 pt. Recall)
- Linking queries to the available data must be studied more, and the LSAT corpus is of great help for that

6. Perspectives

- Automatization of term mapping
 - NLP/QA techniques (TRECVID Multimedia Event Detection, 0-shot event detection)
 - Should be personalized (visual / non visual)
- More precise spatial metadata should help the retrieval
 - knowing the path home/work for instance
 - impact on privacy
- Additionnal activity metadata should help the retrieval
 - Like “static” for instance (usefull for working, tv watching, having a drink...)
- May be defining query by examples
 - “Who did I meet last time I was here?”, “Where did I put my mug?”