# Quaero at TRECVID 2013
# Semantic Indexing Task

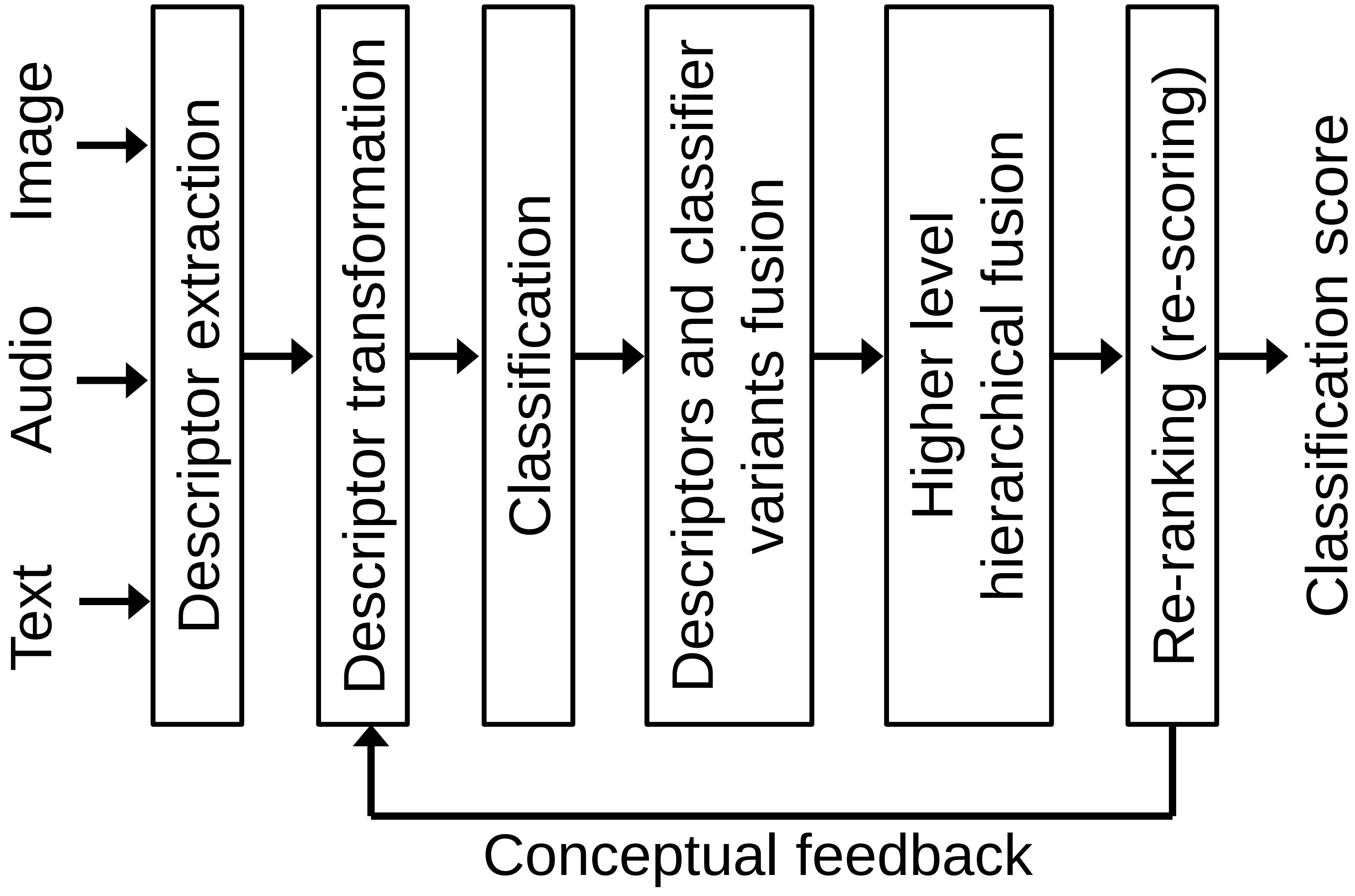
*Bahjat Safadi, Nadia Derbas, Abdelkader Hamadi,*
*Philippe Mulhem and Georges Quénot*
UJF-LIG

20 November 2013

# Outline

- Main task: almost nothing new
  - Use of semantic features: +8% relative gain
  - Result used for the pair and localization tasks
- Pair task:
  - Can we beat the baseline?
- Localization task:
  - Can we do it without local annotations?

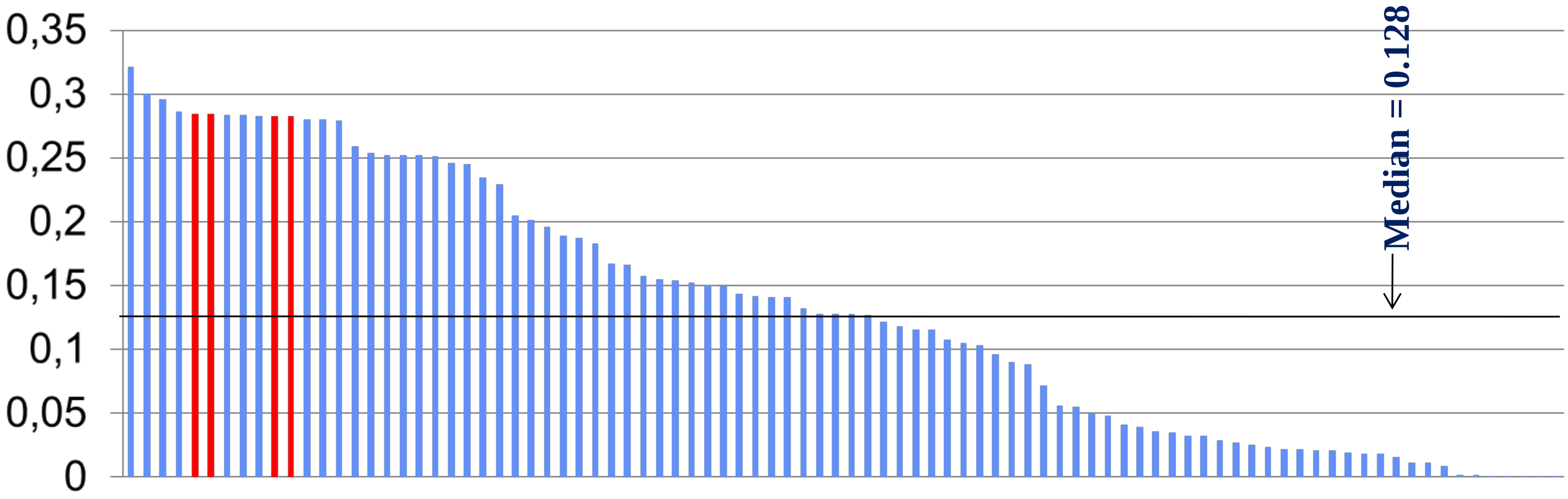
# The Quaero classification pipeline



# Main task

- As in 2011 and 2012 (see TV11 slides)
  - Six-stage pipeline including temporal re-ranking (actually re-scoring) and conceptual feedback
  - Use of a large number of descriptors shared by the IRIM group from GDR ISIS
- New descriptor:
  - Vectors of 1K and 10K concepts scores trained on ILSVRC10 and ImageNet and applied to key frames, kindly produced by Florent Perronnin from Xerox (XRCE)
  - Excellent individual descriptor (infAP of 0.2291, late fusion of both 1K and 10K versions)
  - Complementary to other descriptors: relative gain of 8% before conceptual feedback and temporal re-ranking (from 0.2387 to 0.2576; 0.2848 after feedback and re-scoring).

# Category A results (Main runs)




0.2835 All with one iteration of feedback

0.2848 All with two iteration of feedback

0.2846 All with two iteration of feedback + uploader weak (bug)

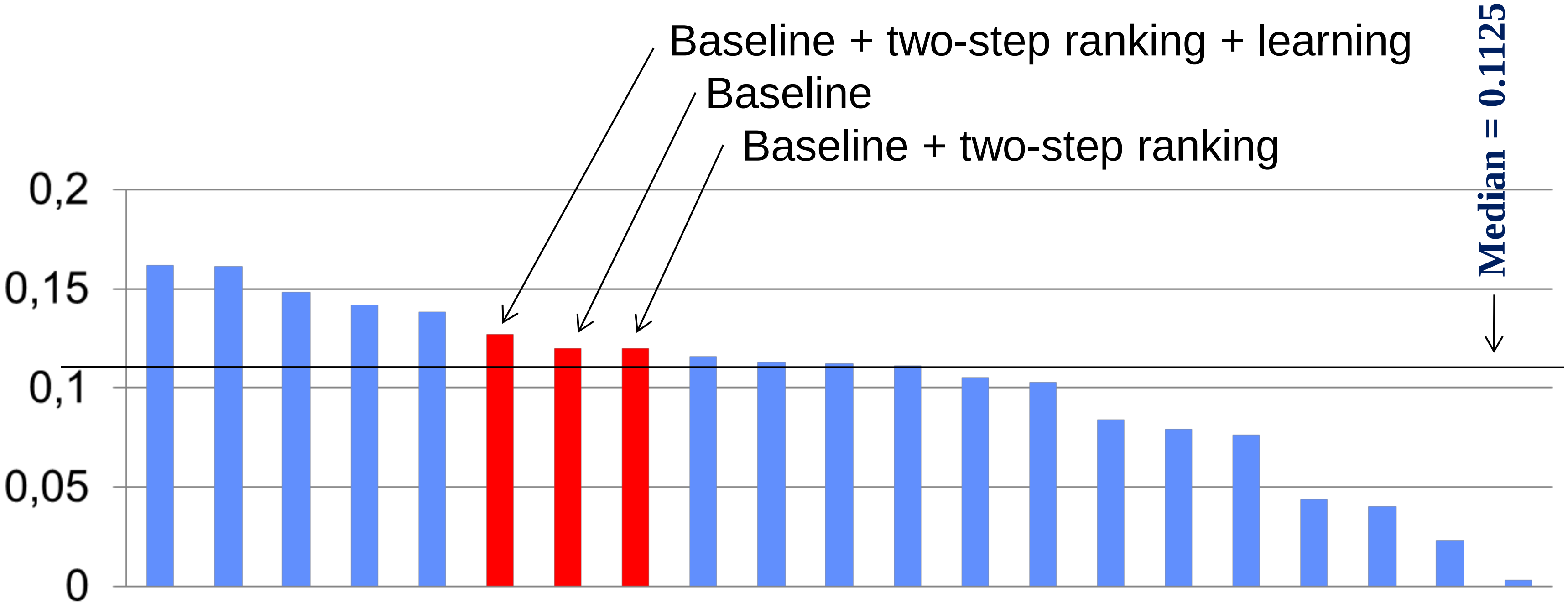
0.2827 All with two iteration of feedback + uploader strong (bug)

Differences not statistically significant

# Concept pairs: can we beat the baseline?

- Which baseline?
  - Single concept scores approximately calibrated as probabilities (e.g. Platt's method)
  - Sum or product (arithmetic of geometric mean) or minimum of the single concept scores
  - Best (worst) individual classifier performance
  - Most (least) frequent single concept
- What alternatives?
  - Direct learning: very imbalanced, extremely few positive samples, but possible for most pairs
  - Other and possibly more complex methods for single concept score fusion

# Category A results (Concept Pairs)




Quaero official submissions on concept pair:

- Not using the final version of single concept scores (late)
- Two-step ranking: ranking the top list of one concept with the ranking of the other + symmetrization, not so goof idea
- Direct learning incomplete relative to the concept learning
- Not bad but not significant results

# “Baselines” from best Quaero submission (NOT official submissions)

- Use of one of the two scores:
  - Most frequent (dev): 0.1096
  - Least frequent (dev) : 0.1130
  - Higher infAP (CV): 0.1222
  - Lower infAP (CV) : 0.1004
- Use of both scores:
  - Sum (arithmetic mean): 0.1613
  - infAP weighted sum (CV): 0.1613
  - infAP weighted sum with power (CV): 0.1637
  - Product (geometric mean): 0.1761 (makes sense)
- Best official submission (UvA): 0.1616

# Alternatives (non official values)

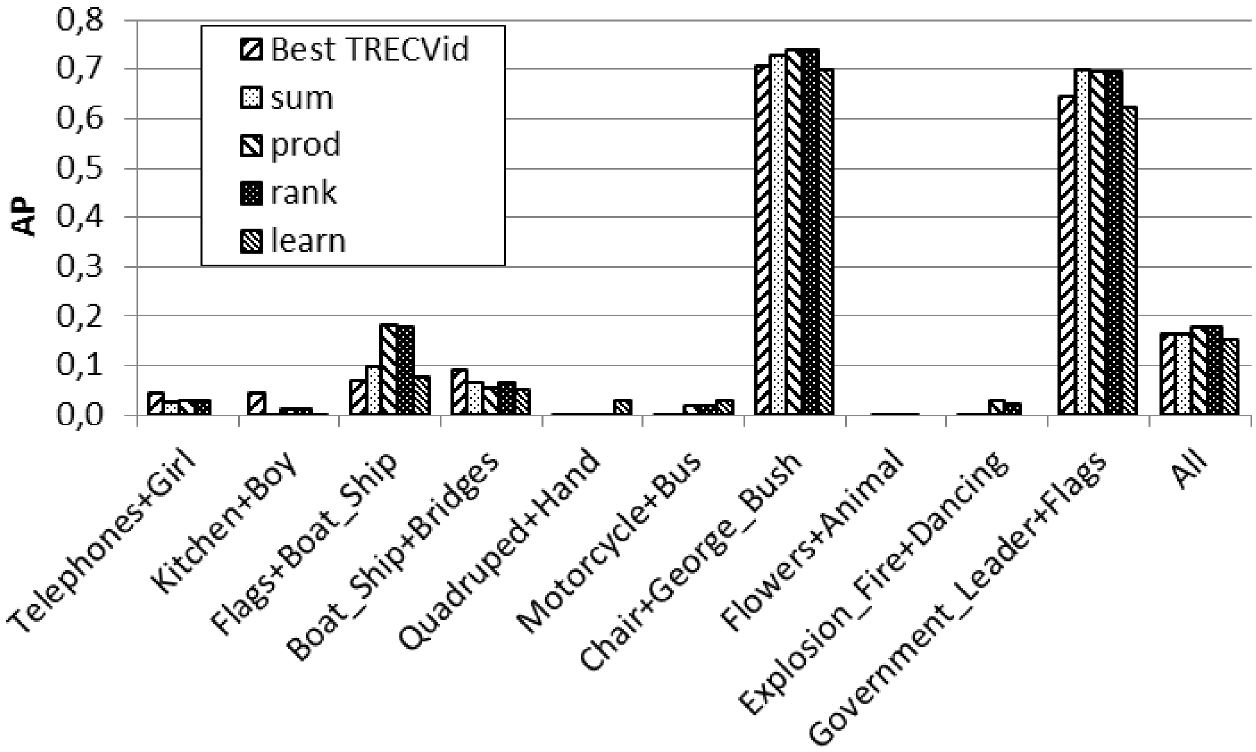
- Rank fusion: arithmetic mean of shot ranks
- Boolean fusion (extended Boolean approach [9]):

$$p(i, c1, c2) = 1 - \sqrt{((1 - p(i, c1))^2 + (1 - p(i, c2))^2)/2}$$

- Direct learning: handle imbalance with MSVM

| System/run | MAP |
|---|---|
| Best submission TRECVid 2013 | 0.1616 |
| linFus | 0.1613 |
| prodFus | 0.1761 |
| rankFus | 0.1767 |
| boolFus | 0.1724 |
| learnDouble | 0.1514 |

# By concept pair results




- Rank fusion is the best, very close to product fusion
- But: most of the MAP is supported by only two concepts
- Almost no difference is statistically significant ☹

# Localization task:
# Can we do it without local annotations?

Motivation:

- Annotations are costly and boring
- Local annotations are even more
- We had no time and support to do any

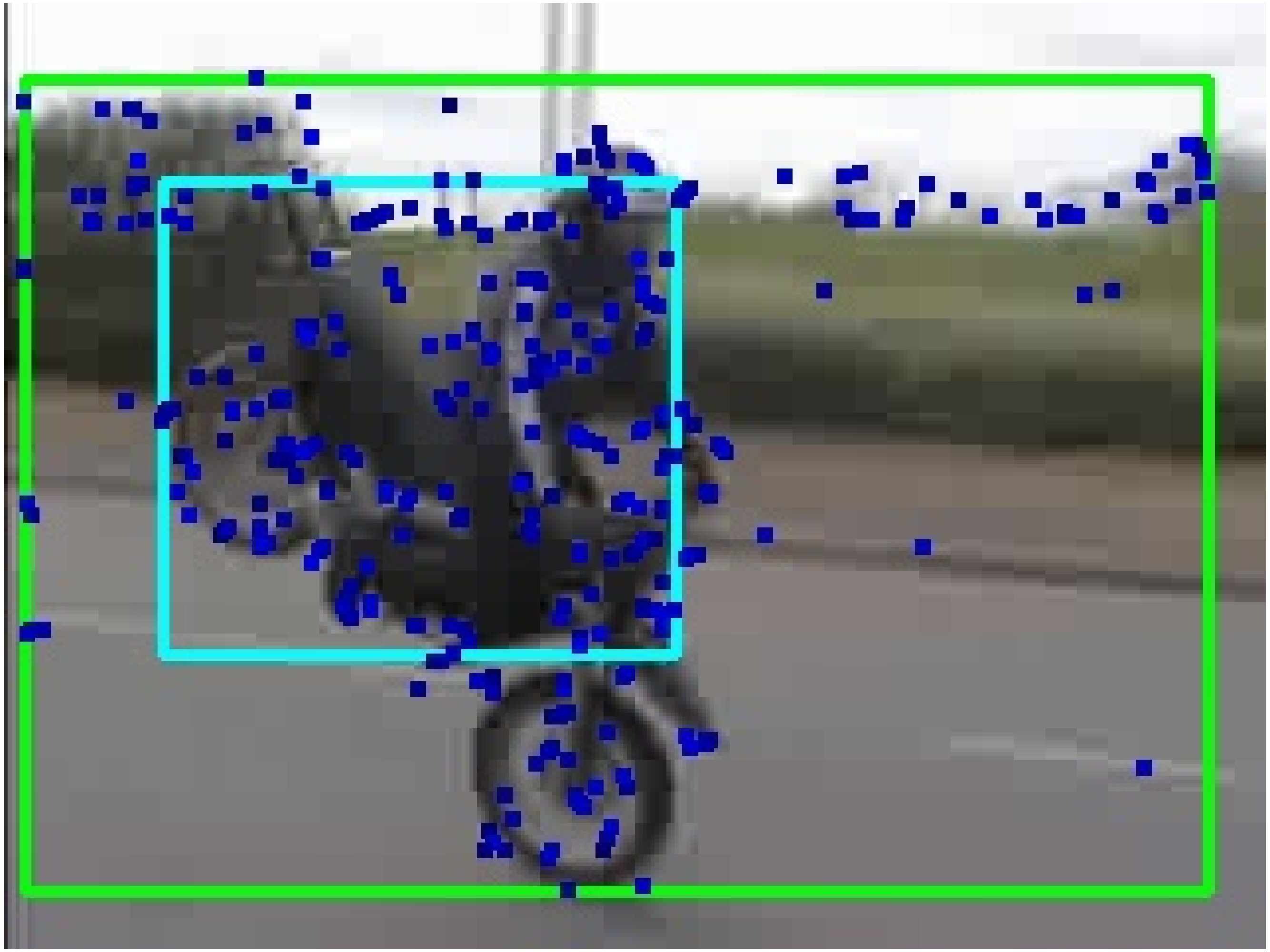
# Localization task proposed approach

Inspired from (Ries and R. Lienhart, 2012):

- Compute local descriptors (opponent SIFT fro UvA tool)
- Cluster local descriptors (k-means)
- Learn discriminative models for clusters based on relative occurrence frequencies using global image annotations only
- Filter points in a an image predicted as globally positive
- Select a rectangle according to the density of points using horizontal and vertical projections
- Main problem:
  - no training data for parameter tuning (e.g. threshold selection);

C. X. Ries and R. Lienhart. Deriving a discriminative color model for a given object class from weakly labeled training data. In Proceedings of the 2nd ACM International Conference on Multimedia Retrieval, page 44. ACM, 2012.

# Localization task proposed approach



# Filtering SIFT points

- Relative Occurrence Frequency (ROF):

  $ROF_p(y) = p_y/p$ and $ROF_n(y) = n_y/n$ with:

  $p_y$ (resp. $n_y$) = number of positive (resp. negative) images in which at least one point belonging to the cluster y is present in the image and:

  p (resp. n) = total number of positive and negative images

- Filter a point associated to a cluster y according to $ROF_p(y)/ROF_n(y)$ or simply to $ROF_p(y)$ (better)

# Finding Rectangles

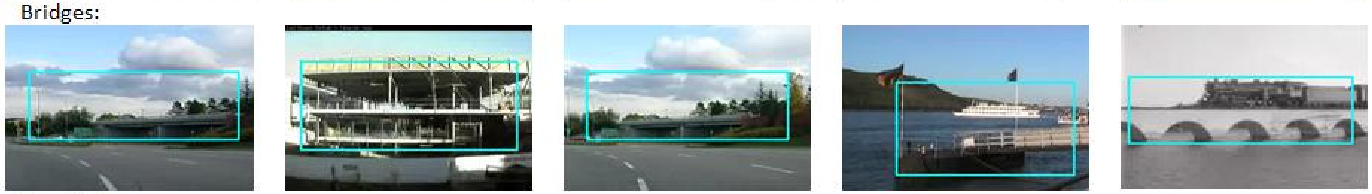
- Compute horizontal and vertical histograms of filtered points (32 bins for each projection)
- Remove bins from left and right (resp. top and bottom) as long as the bin value is below a given threshold $\beta$
- Keep the rectangle covering the remaining bins
- $\beta$ is manually tuned separately for each concept by looking at the top 500 results within the development set (human intervention but not exactly annotation)
- Limitation: approach suited for finding a single rectangle

# Sample results (1)

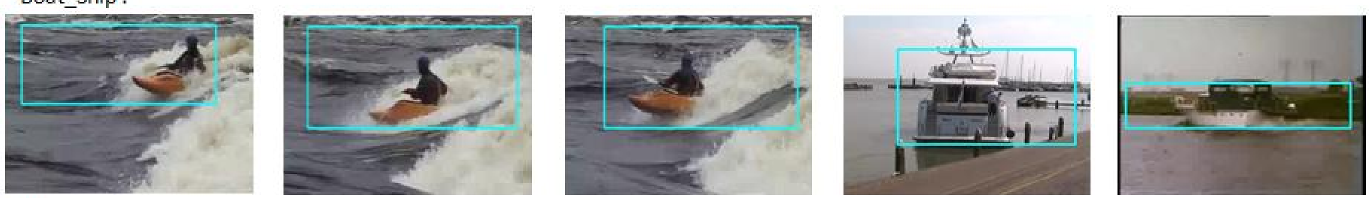
Motocycle:

Airplane:

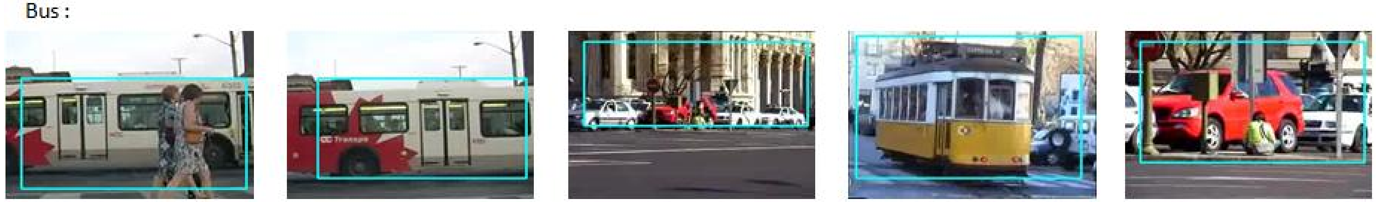
Bridges:



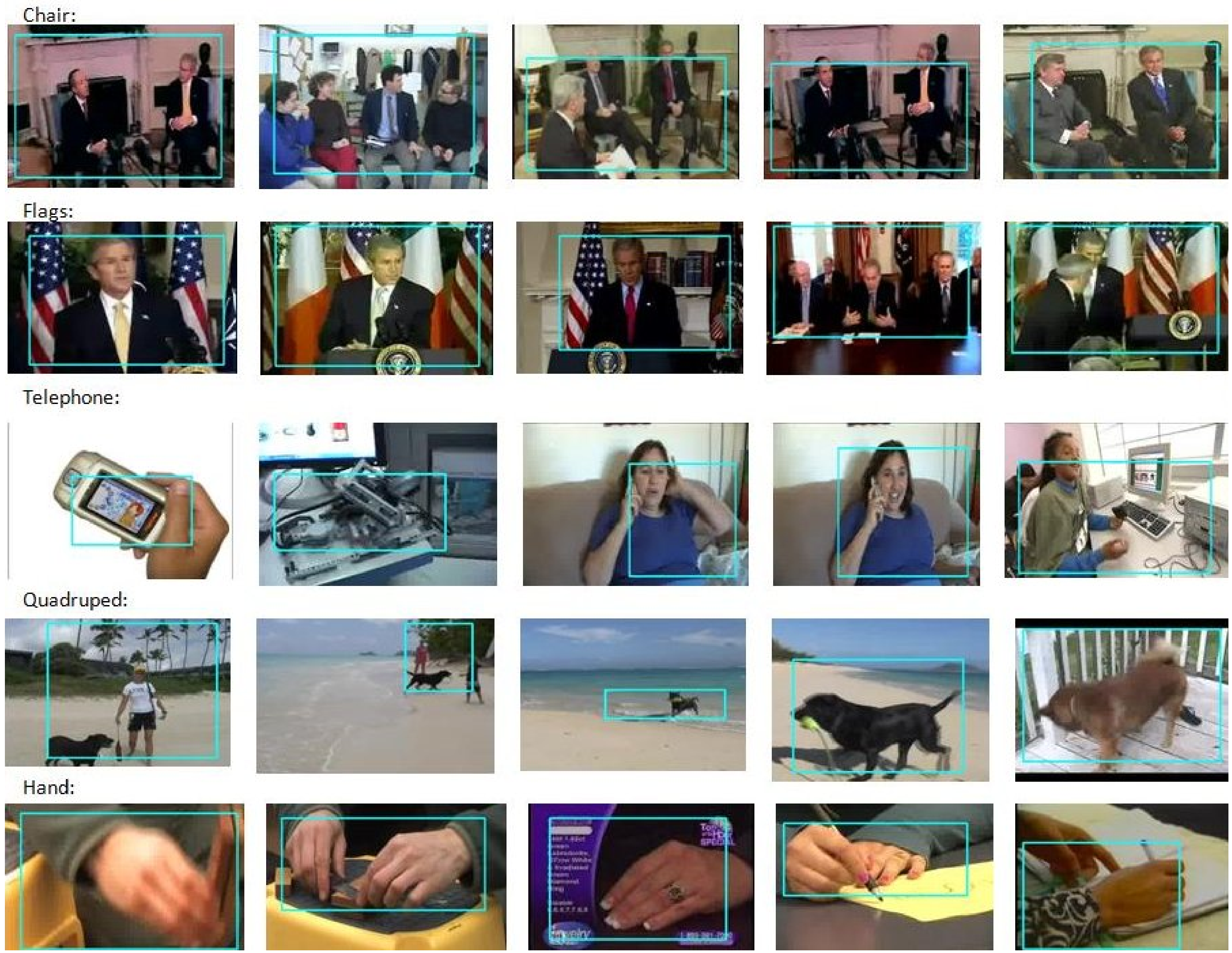
Boat_Ship :



Bus :



# Sample results (2)



# Only one submitted run

- Quite good in temporal detection but mostly comes from the concept detector developed for the main task
- Less good for the spatial localization but not so bad
- The recall versus precision compromise was not optimized
- No region annotation was used
- Many possible improvement
- TV13 assessment will allow a better tuning for next issues or other applications

# Thanks