INRIA LEAR's COPYRIGHT DETECTION SYSTEM

Hervé Jégou

Joint work with Matthijs Douze & Cordelia Schmid

INRIA

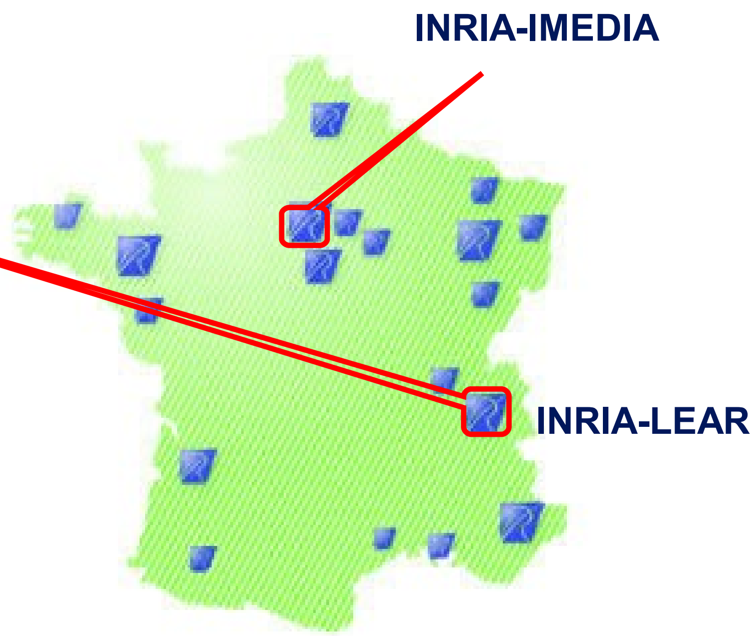
LEAR

- INRIA: National french institute
  - Computer sciences, electrical engineering, applied mathematics

  → two separate teams from INRIA have participated to this task

- INRIA-LEAR
  - located in Grenoble

  → close to the Alpes mountains



Outline

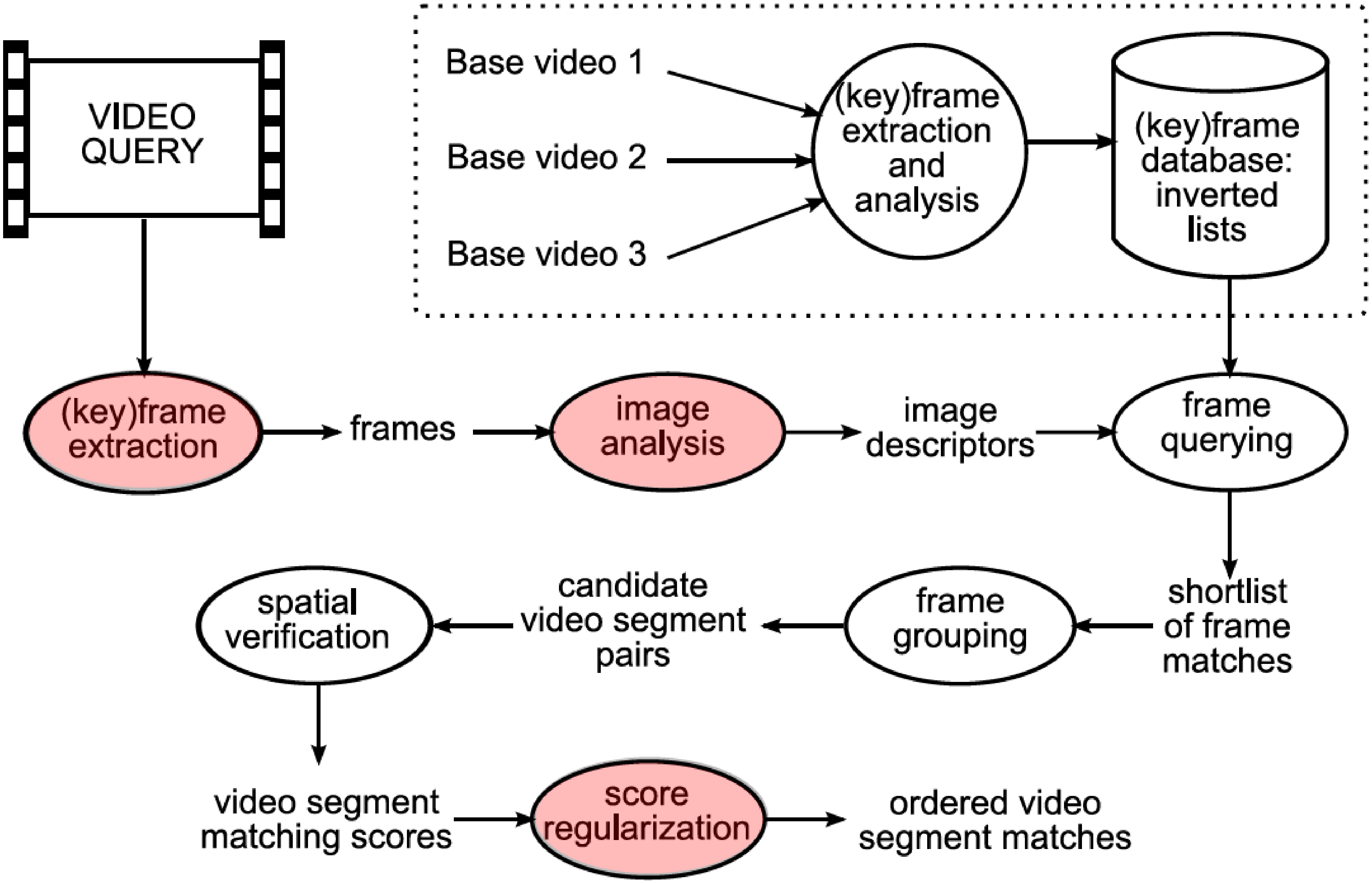
THREE KEY COMPONENTS

- Local descriptors with high invariance
- Hamming Embedding and new extensions
- Weak Geometry consistency

ABOUT THE COPYRIGHT DETECTION TASK

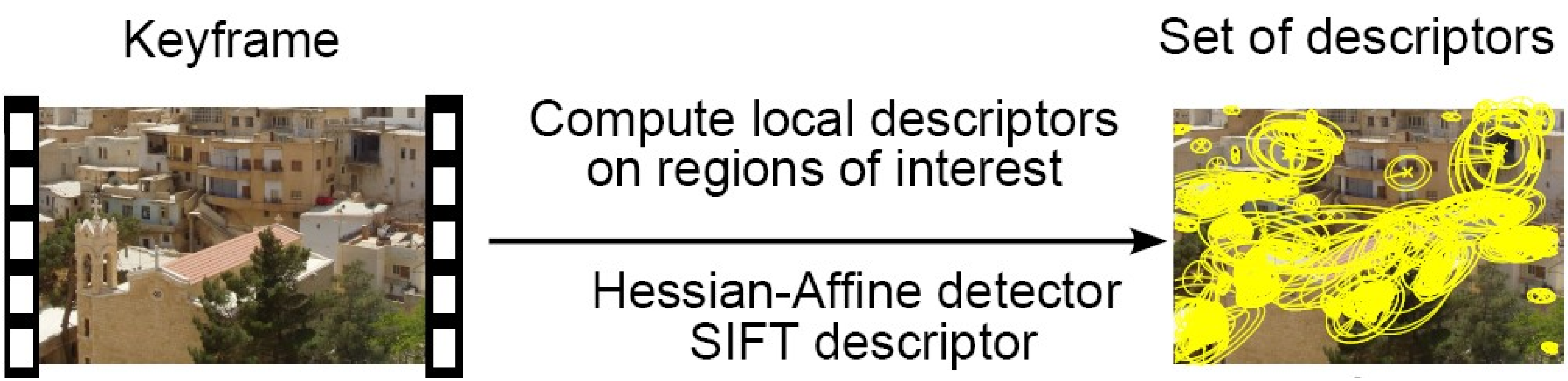
- Our validation dataset
- Our runs

Preliminary: a "short" overview of our system

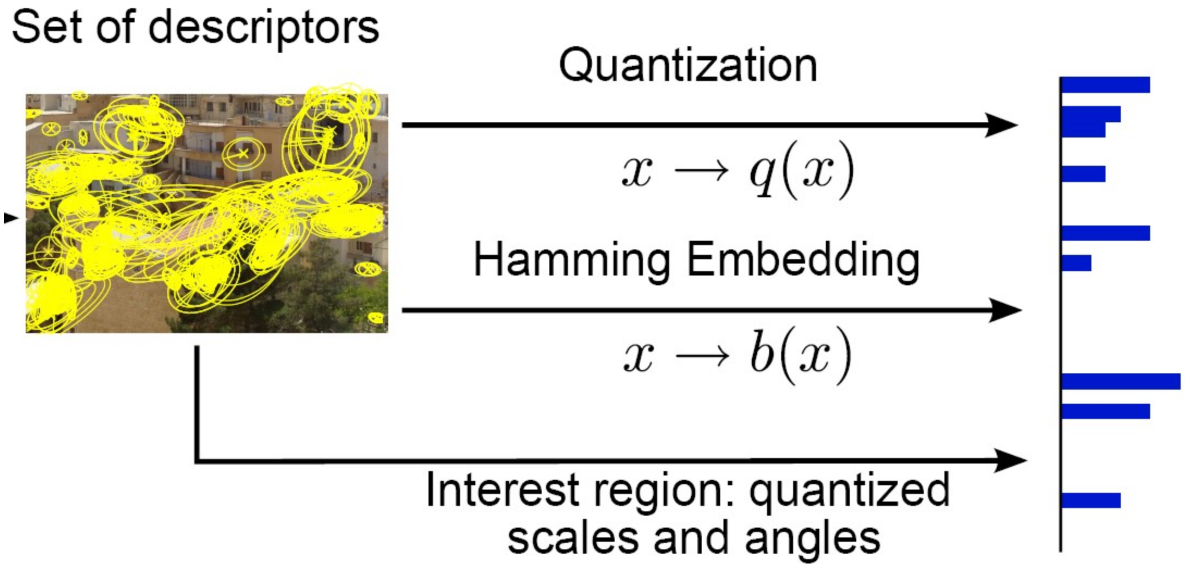


Local descriptors

- DETECTOR [Mikolajczyk Schmid, IJCV’05] : high degree of invariance
  - Hessian detector
  - scale invariance
  - orientation invariance
  - affine invariance
- DESCRIPTOR: SIFT [Lowe, IJCV’04]



Extended Bag-of-features [Jegou al., ECCV'08]

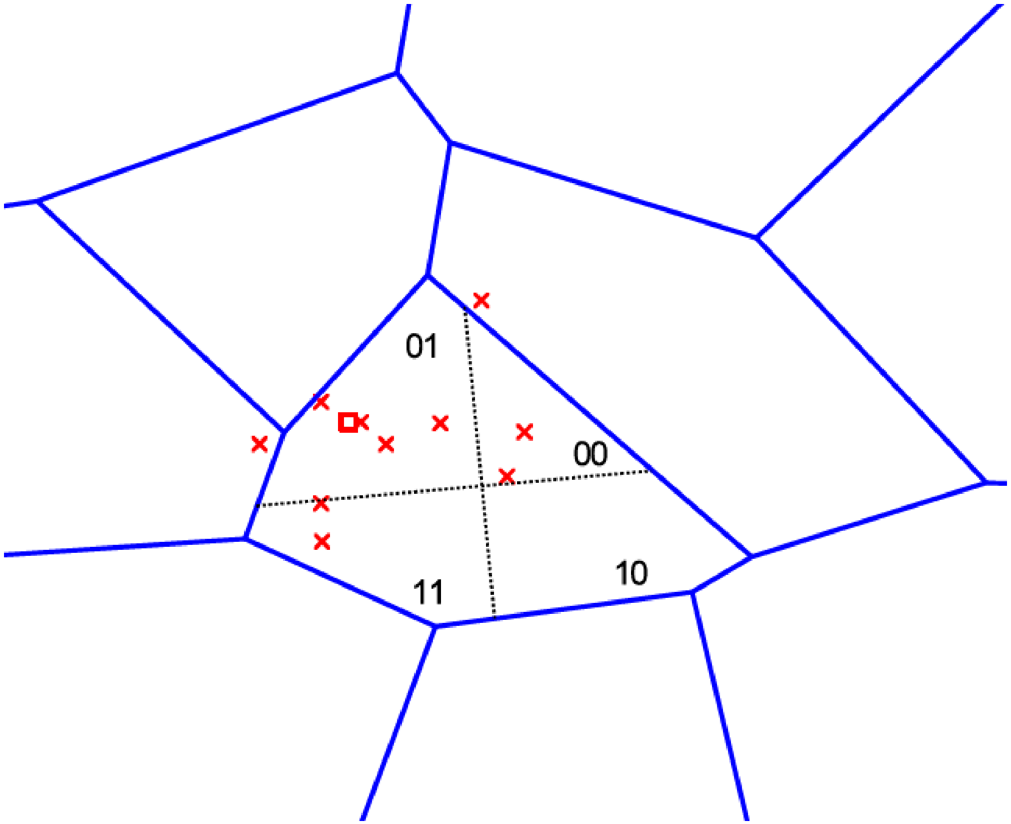


desc. x and *y* match iff

$$\begin{cases} q(x)=q(y) \\ \text{Hamdis}(b(x),b(y)) < h_t \end{cases}$$

- Output descriptor representation (12 bytes in memory):
  - id — frame identifier — 21 bits
  - q(x) — quantization cell — implicited coded by inverted lists
  - b(x) — binary signature — 64 bits
  - s(region) — characteristic scale — 5 bits
  - a(region) — dominent orientation — 6 bits

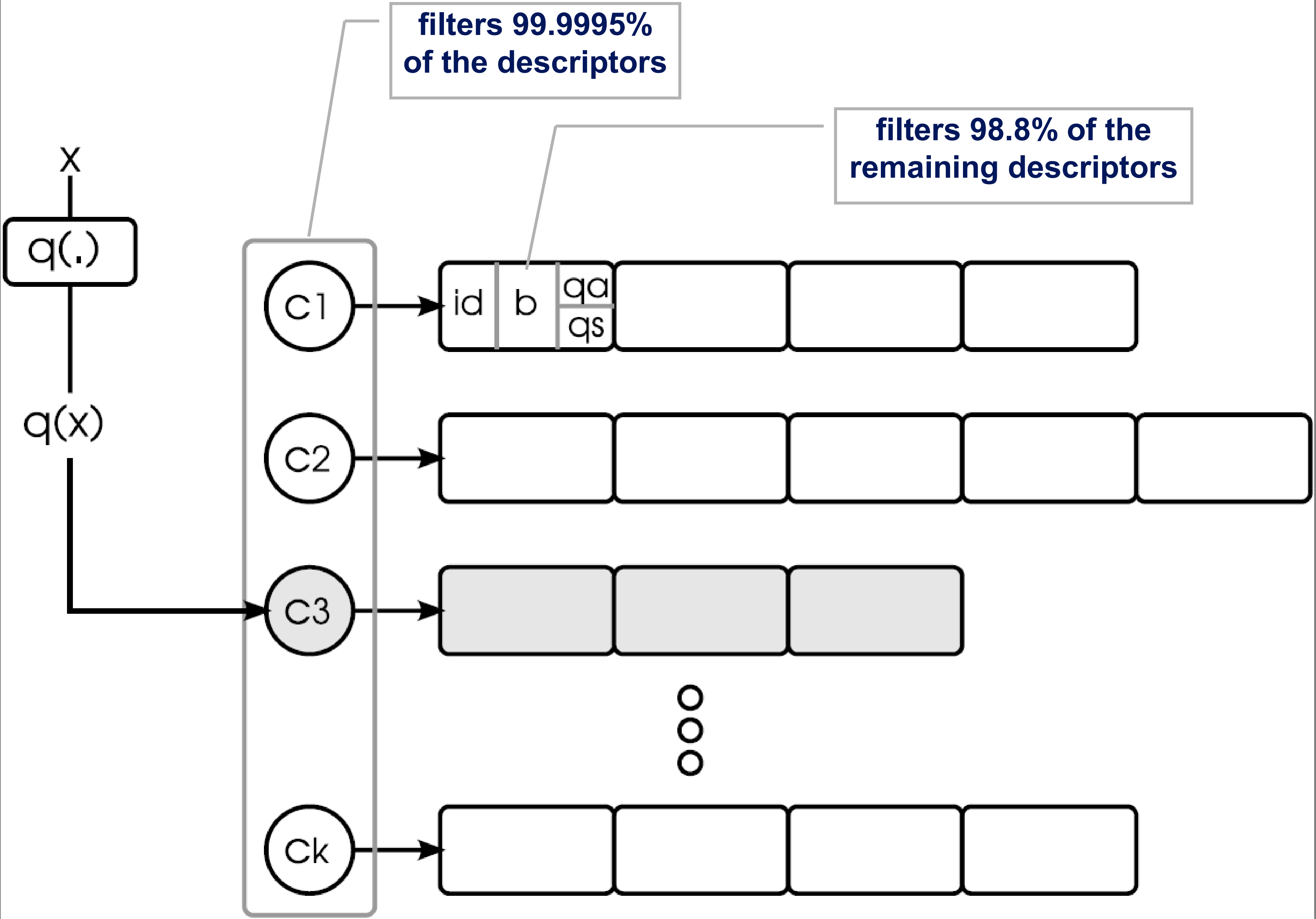
Hamming Embedding (HE) – [Jegou et al., ECCV'08]



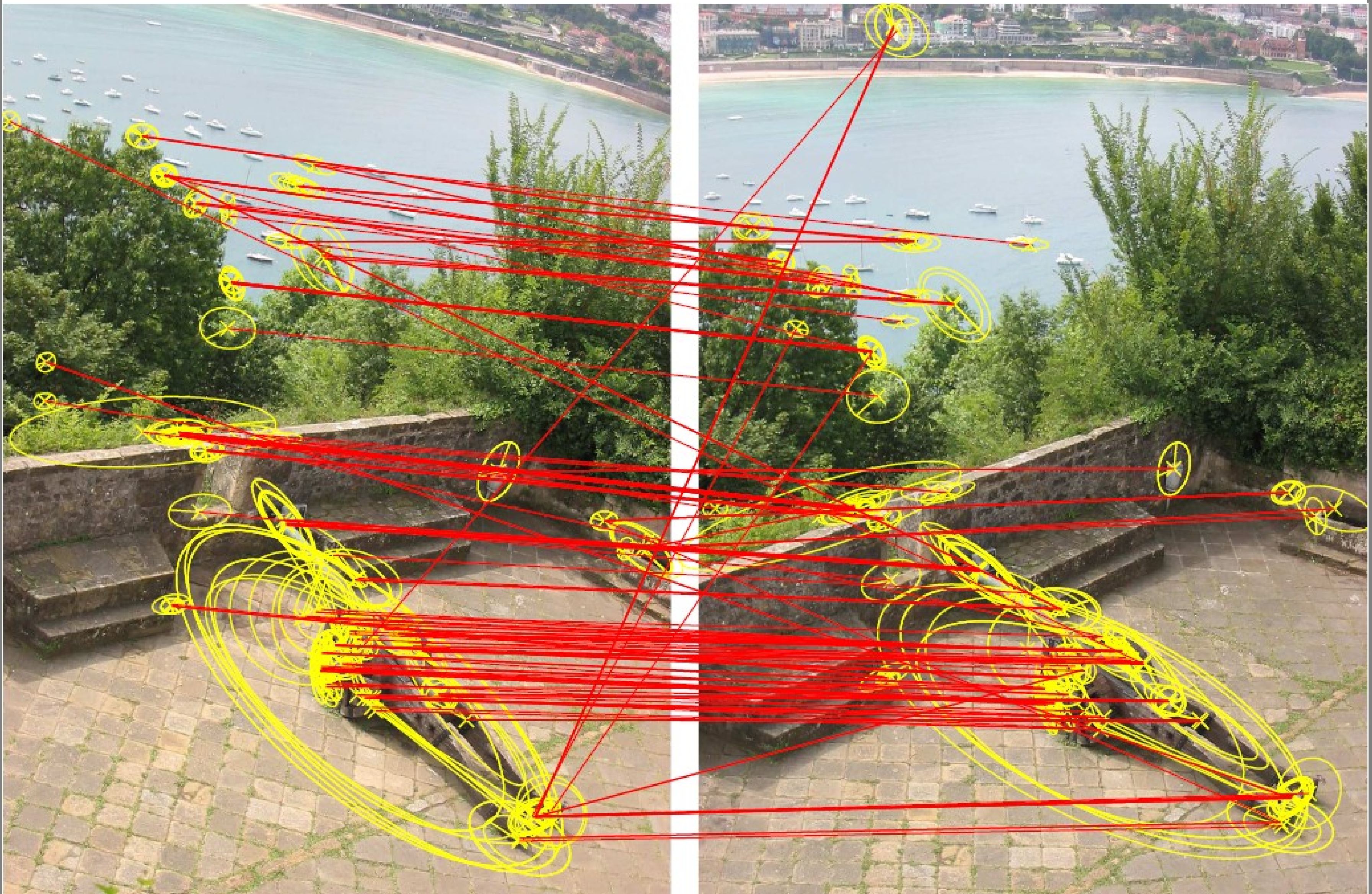
- **Improvement** for Trecvid's detection task:
  1. Multiple assignment of descriptors to quantization cells
  2. Weighting of the Hamming distance

  → based on *the Shannon information content*

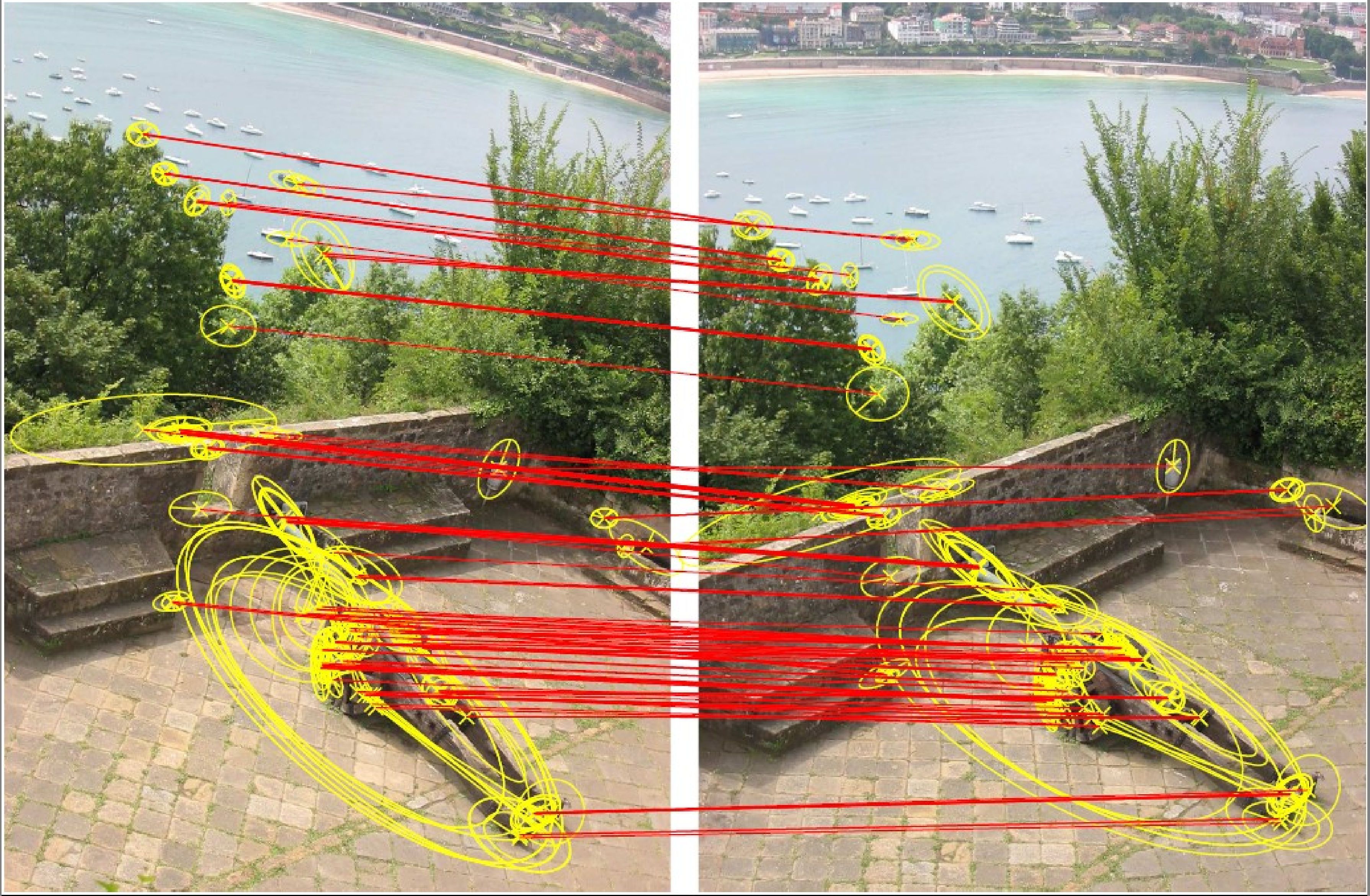
Indexing structure



Matching example: with Bag-of-features only



Hamming Embedding filters matches



Weak geometry consistency [Jegou et al., ECCV'08]

- We have used invariance to scale and orientation changes

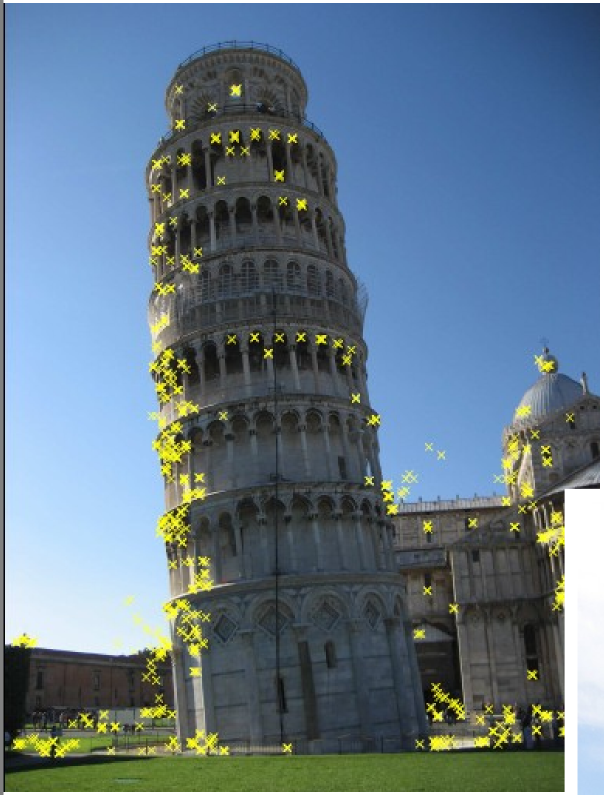
→ high invariance, but at the cost of lower discriminative power

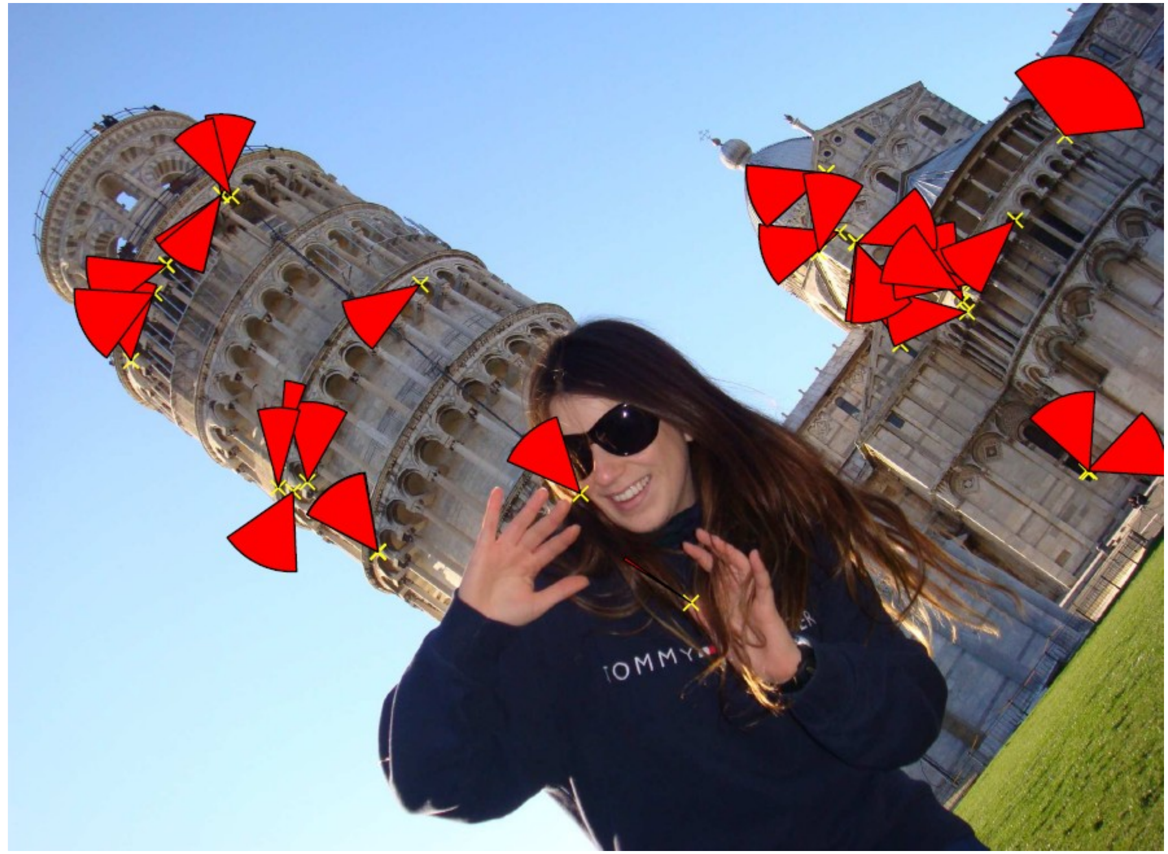
- Weak geometry consistency (WGC):
  - use the angle and scale information provided by the region detector
  - to filter the descriptors which are not scaled/rotated consistently

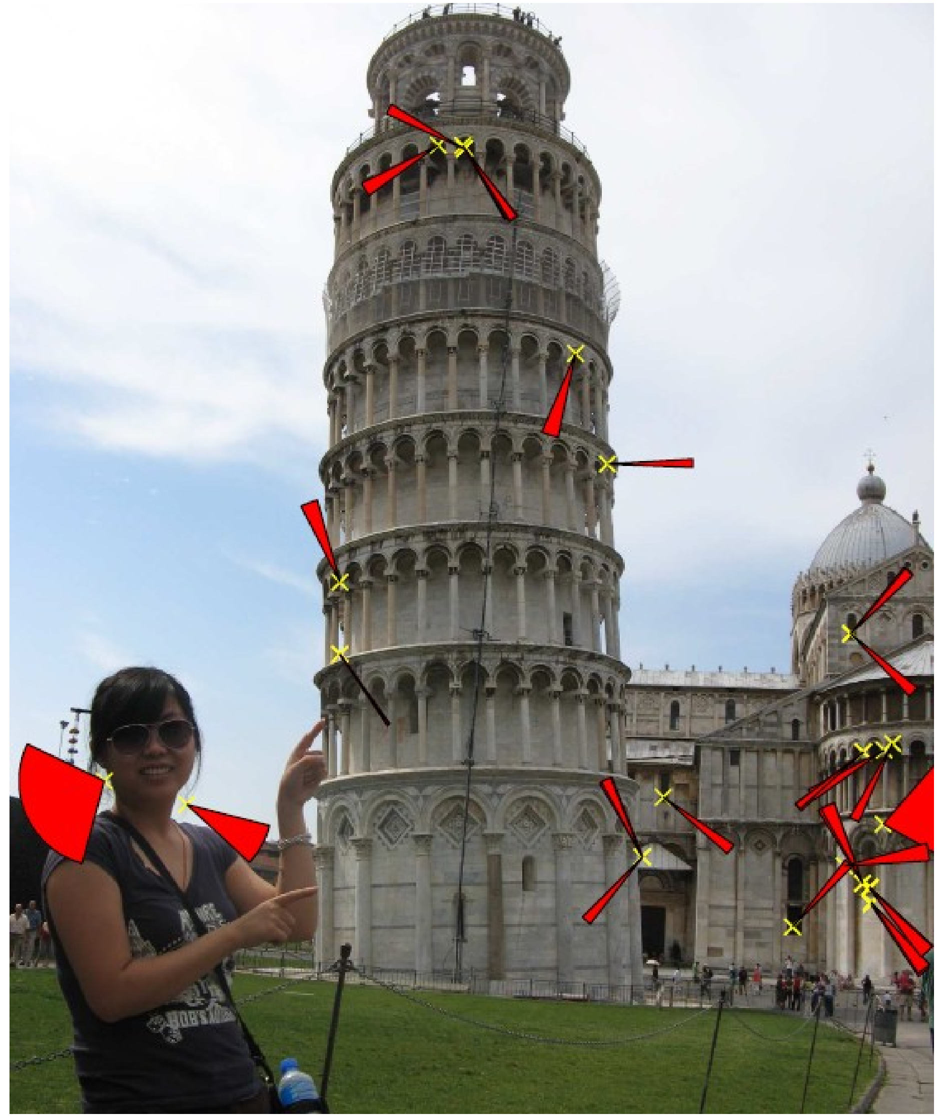
- Strong points
  - descriptors are now **consistently** invariant
  - without explicitly estimating a transform mapping a frame to another

  → very efficient for millions of frames

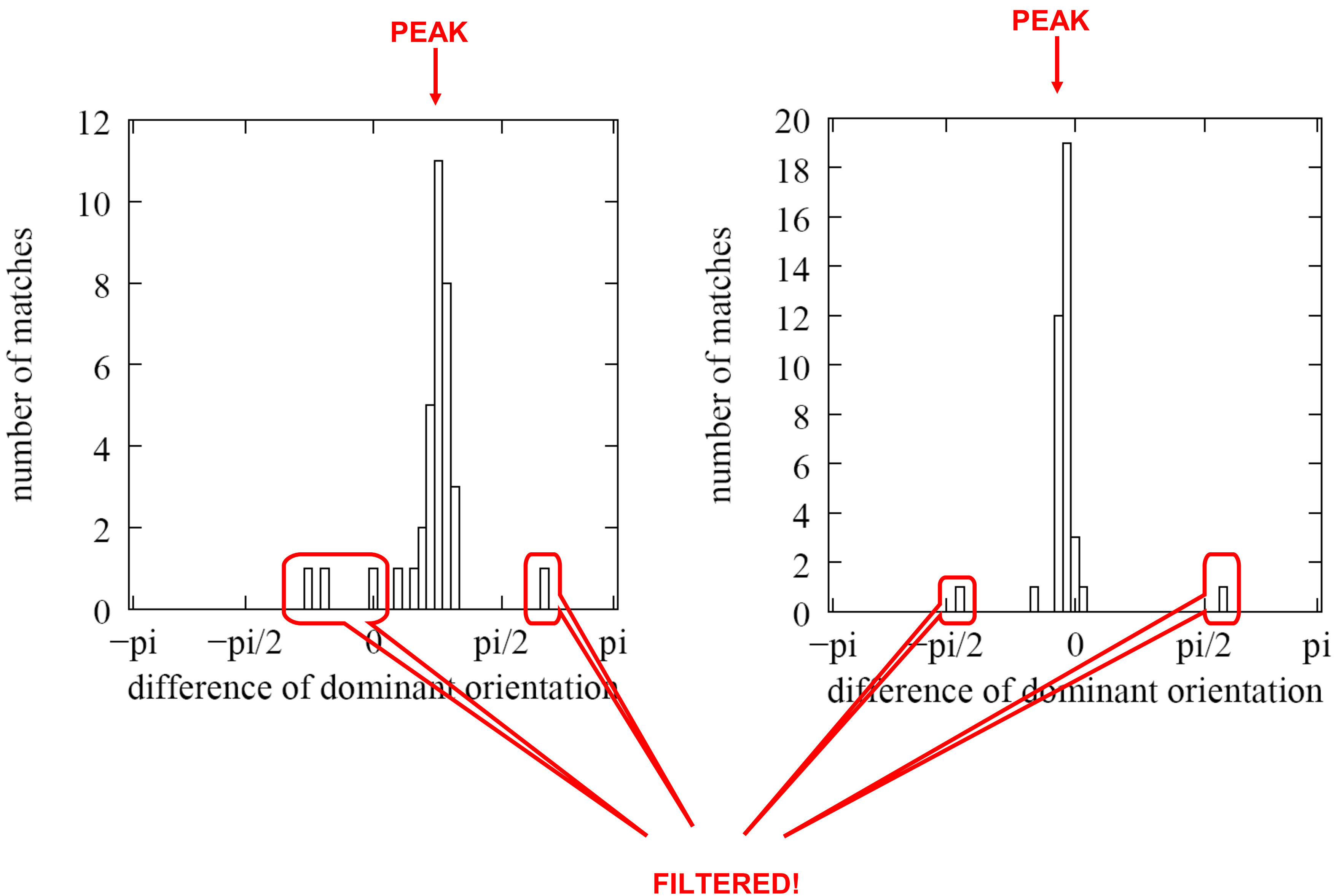
WGC: orientation



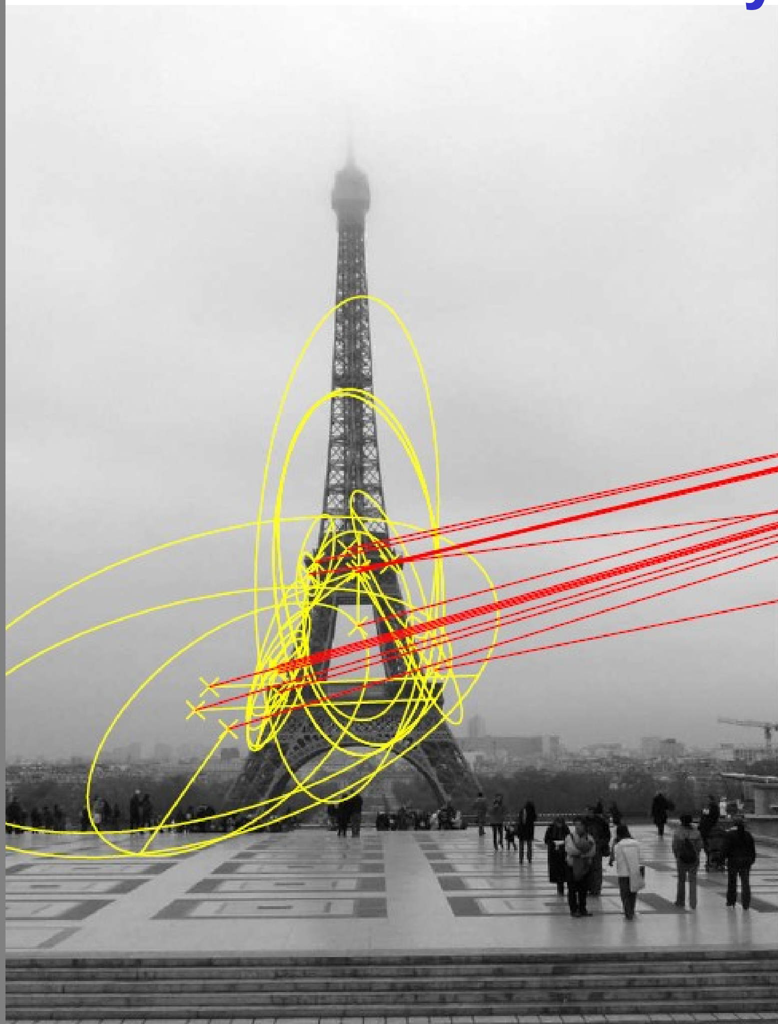

TOMMY

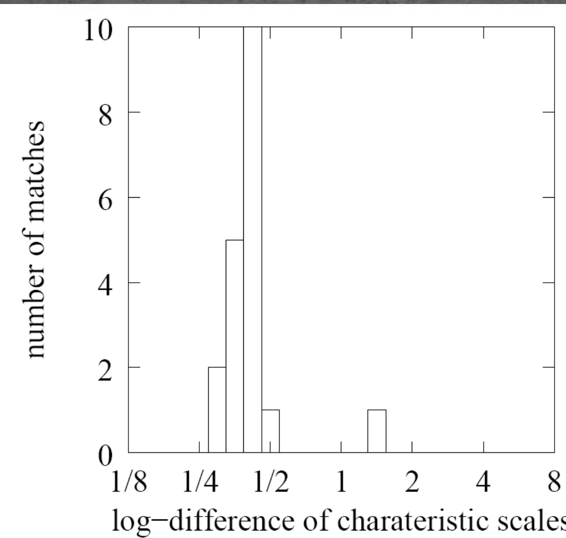


Find peak of orientation differences

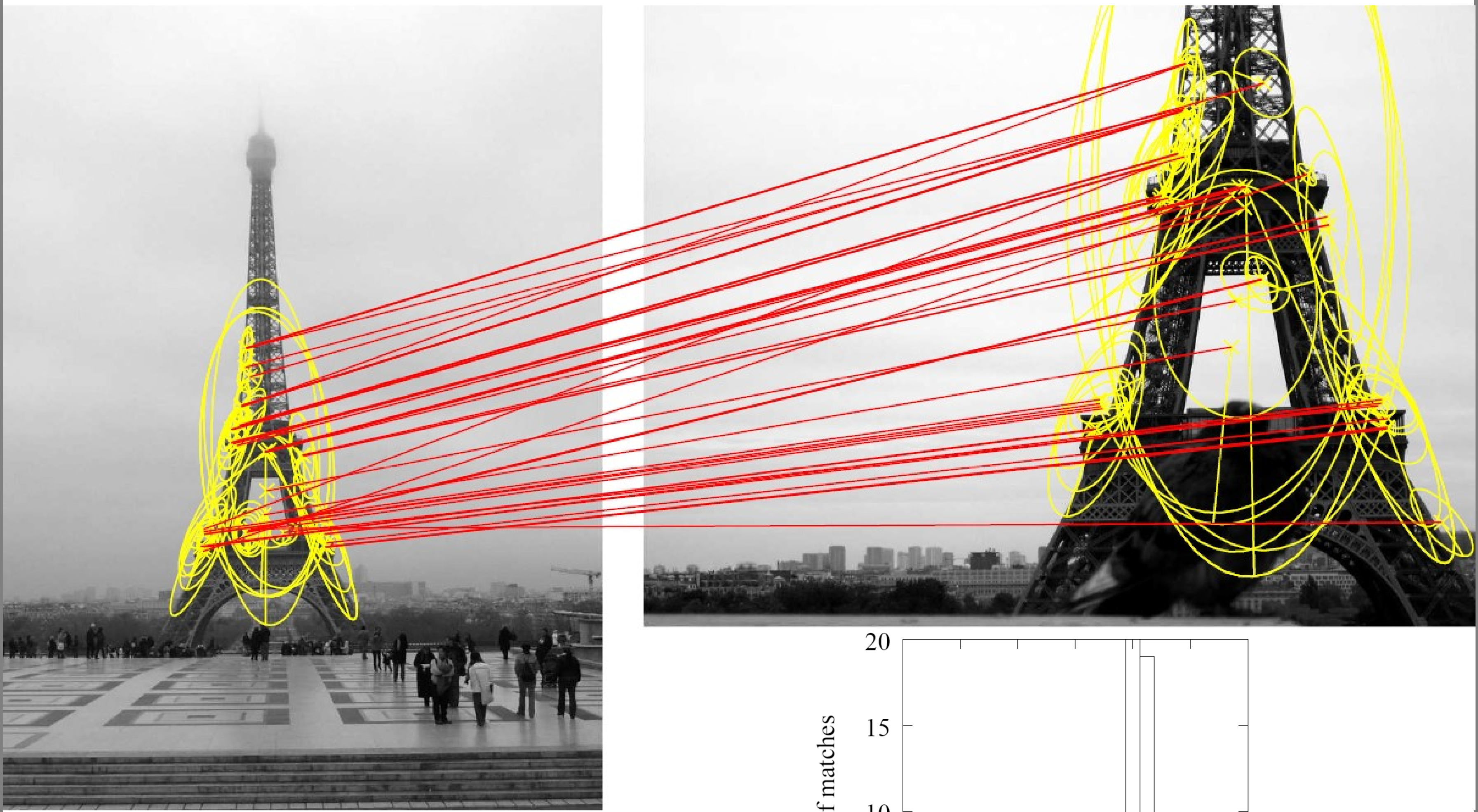


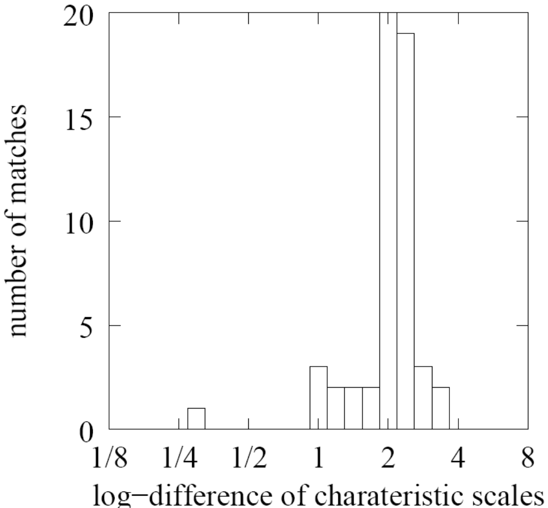
WGC: scale consistency



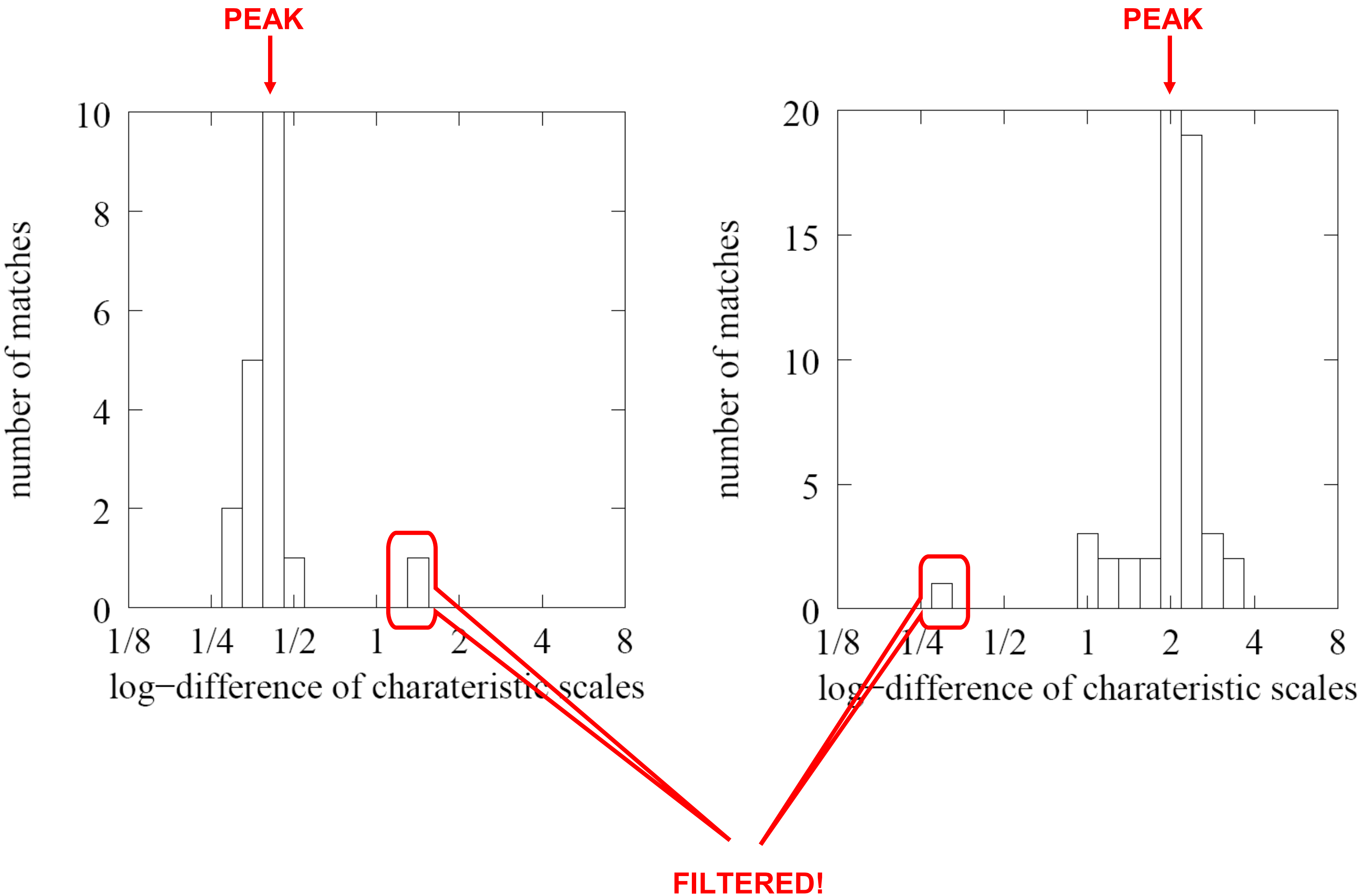


WGC: scale consistency





Find peak of log-scale differences



A priori on transformations

- Frame scores are penalized by
  - strong rotations
  - important changes in scale

→ done directly on the previous histograms of orientation and angle

- For the picture-in-picture

→ the indexed dataset is stored twice: normal size and **half-size**

- To handle flipped videos

→ submit the flipped video query

Creating data for the evaluation

- The proposed validation set was not difficult enough to optimize our system
  - near perfect results in our first run!

- We have used two home-made validation datasets

  $\rightarrow$ available online

- **Holidays dataset**
  - pure image dataset
  - to have shorter feedback for our core image system

- **Video validation dataset**
  - query generation tool created for the purpose of the Trecvid evaluation
  - we have generated (very) difficult queries

Video query generator
http://lear.inrialpes.fr/people/douze/trecvid_generator

Video query generator
http://lear.inrialpes.fr/people/douze/trecvid_generator




Our runs: KEYSADVES – STRICT - SOFT

Same algorithm for all runs

KEYSADVES

- *Asymmetric* frame extraction: 1 frame/6s on dataset side, 2/s on query side
- 95K frames indexed = 39M descriptors

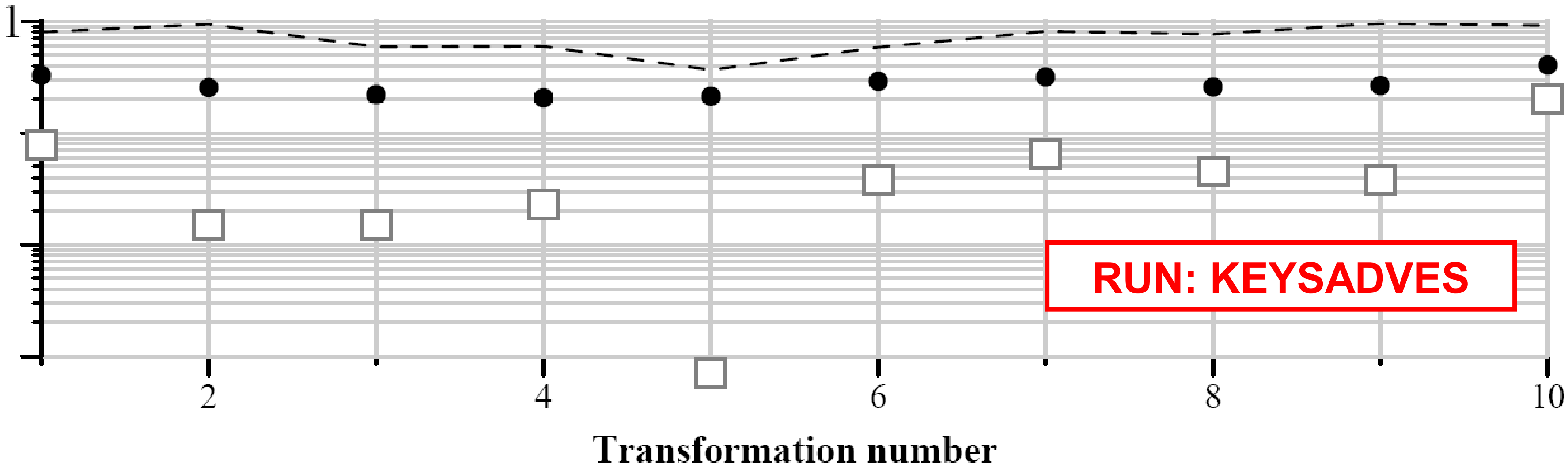
SOFT

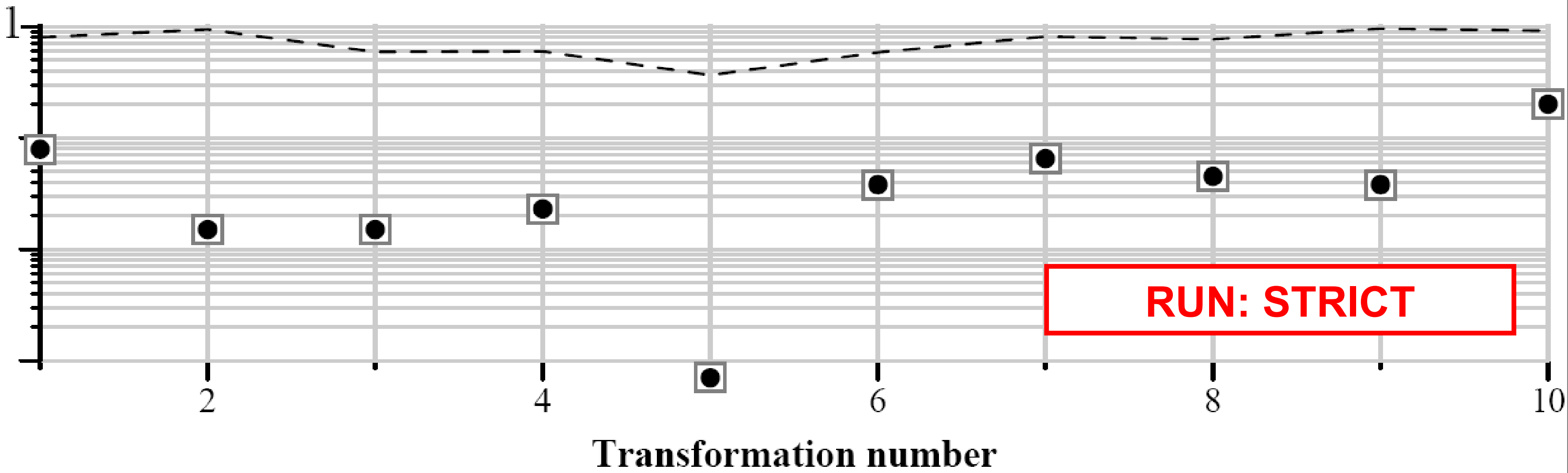
- symmetric frame extraction: 2 frames/s on both sides
- 2M frames indexed = 875M descriptors

STRICT

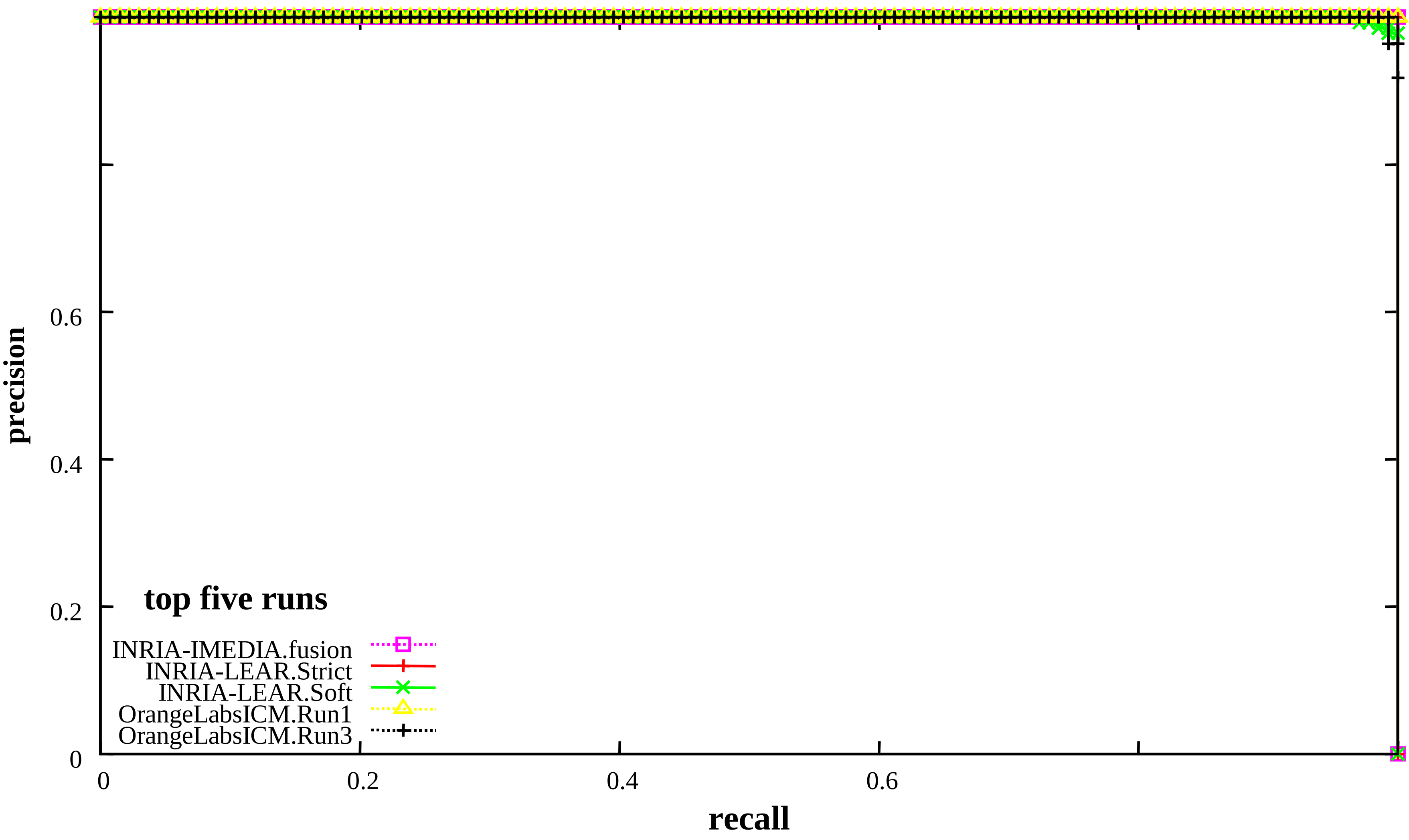
- almost the same as SOFT
- returns 1 result or none

Results: NDCR

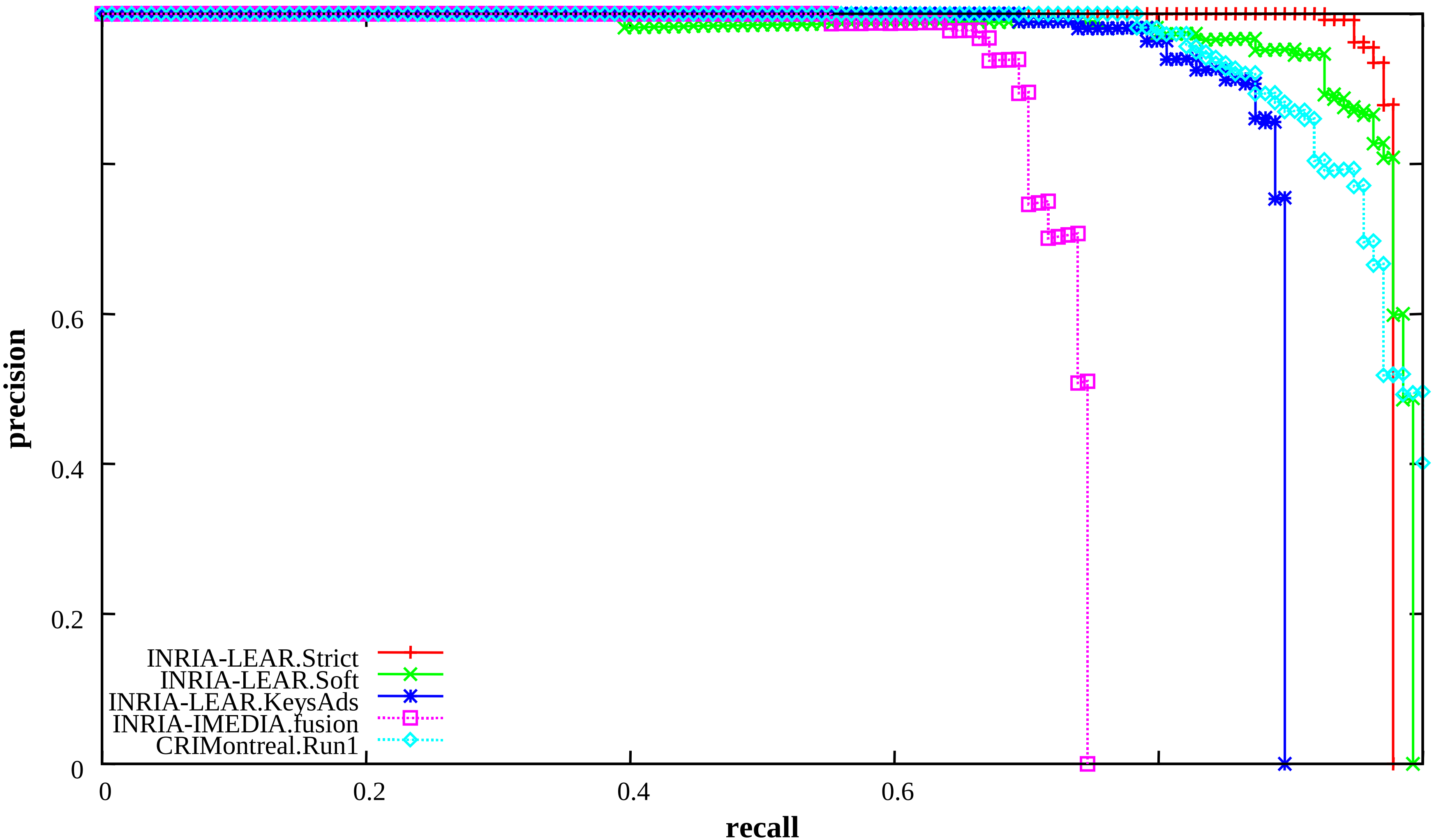




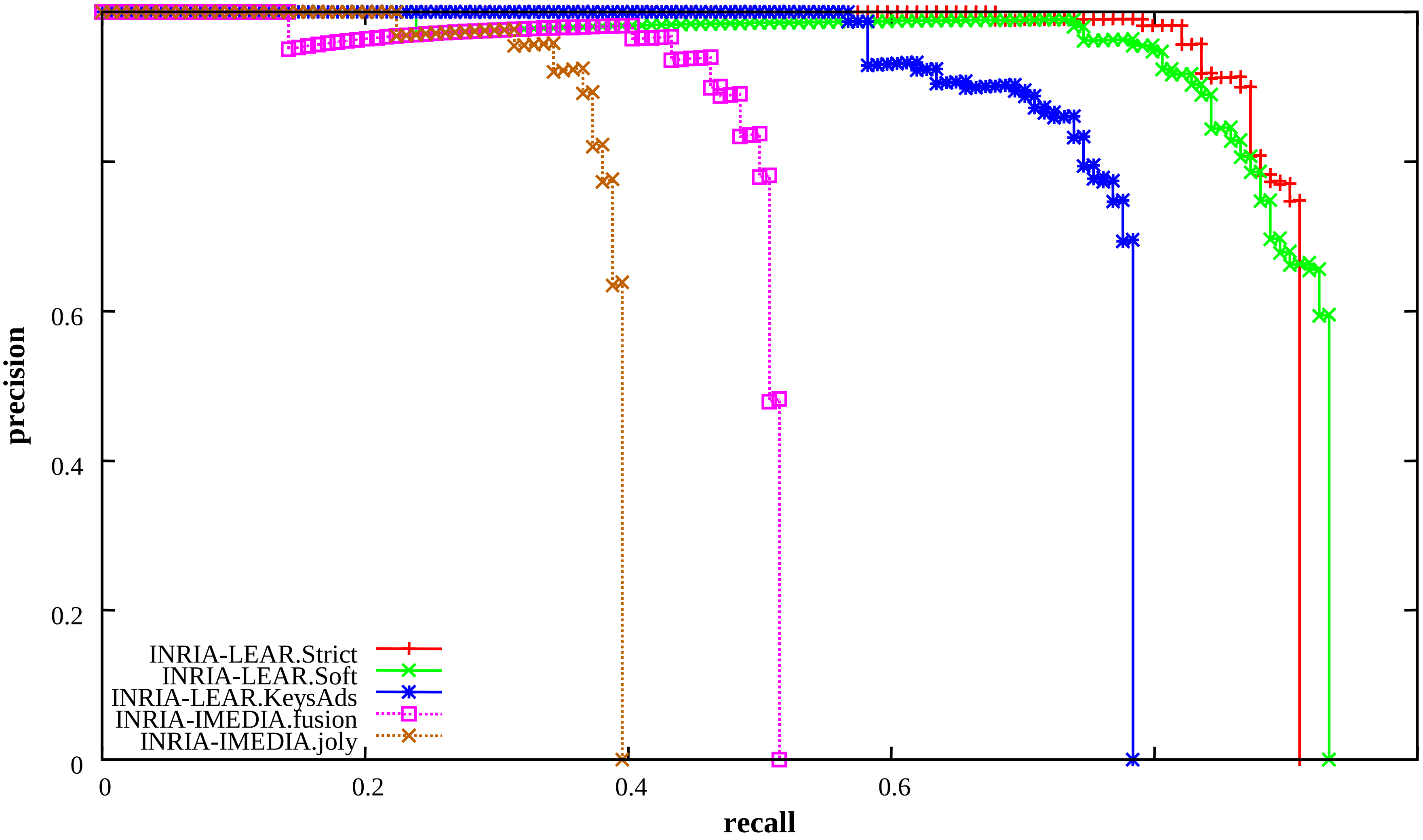
Precision-recall: change of gamma



Precision-recall: camcording



Precision-recall: combined transformation (10)



Sample result 1

Sample result 2

Sample result 3

Conclusion

- Important points for high search quality
  - high number of extracted frames
  - the quality of the approximate nearest neighbor search

  → extended version of Hamming Embedding

- Trecvid specific methods
  - the scoring strategy has an high impact on the NDCR measure

  → detailed in our notebook paper
  - submitted flipped videos
  - store half-size videos

References

- Papers
  - System overview: Trecvid'2008 notebook paper
  - Hamming Embedding: ECCV'08 paper
  - Weak Geometry Consistency: ECCV'08 paper
  - Multiple assignment and HE weighting: INRIA technical report (online)

- Online Ressources
  - Local invariant descriptors: binary software by Mikolajczyk
    http://www.robots.ox.ac.uk/~vgg/research/affine/
  - Holidays dataset for optimizing the core image system
    http://lear/inrialpes.fr/people/jegou/data.php
  - Video query generator tool and sample examples
    http://lear.inrialpes.fr/people/douze/trecvid_generator

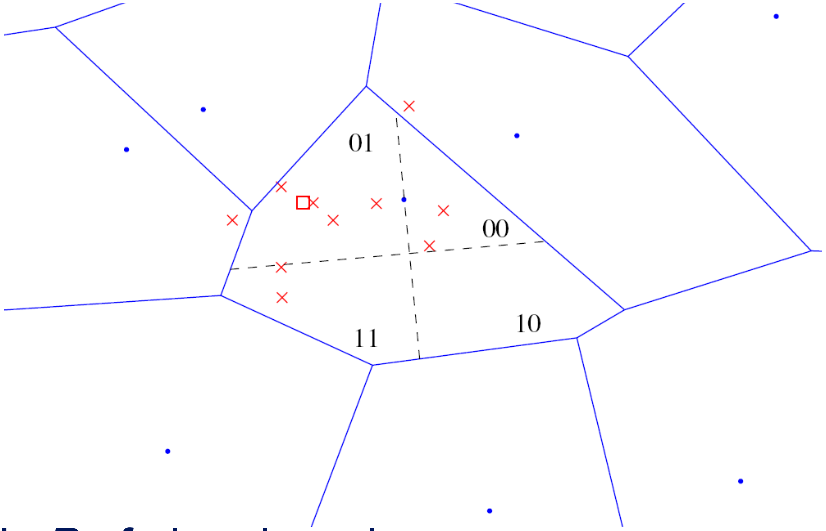
END: ?

Hamming Embedding

- Representation of a descriptor *x*
  - Vector-quantized to *q(x)* as in standard BOF
  - \+ short binary vector *b(x)* for an additional localization in the Voronoi cell
- Two descriptors x and y match iif $\begin{cases} q(x) = q(y) \\ \mathrm{h}(b(x),\ b(y)) < \tau \end{cases}$

  where h(*a*,*b*) is the Hamming distance
- Nearest neighbors for Hamming distance ≈ those for Euclidean distance
  → a metric in the embedded space reduces dimensionality curse effects
- Efficiency
  - Hamming distance = very few operations
  - Fewer random memory accesses: 3 x faster that standard BOF with same dictionary size!

Hamming Embedding

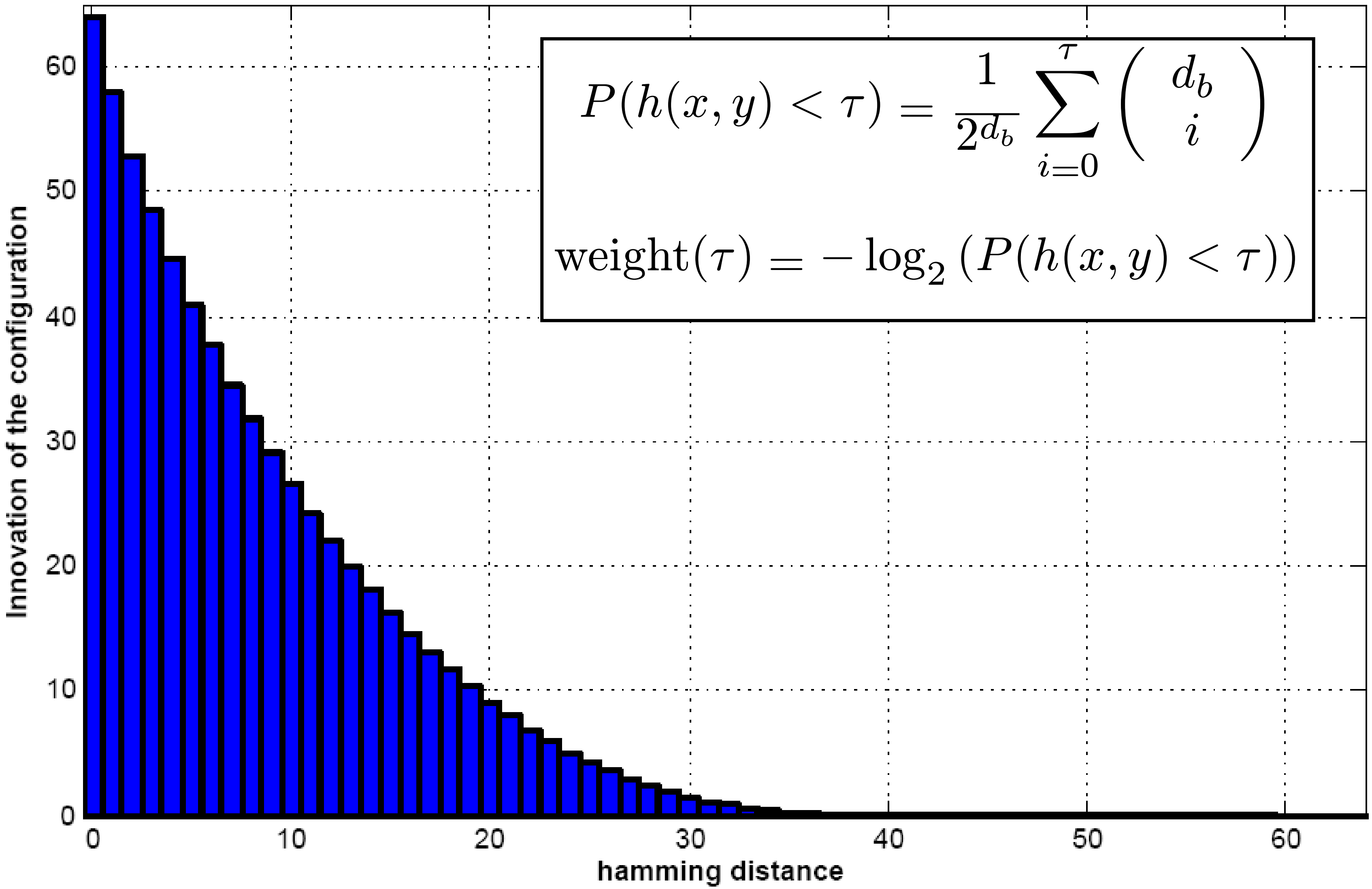


- **Off-line** (given a quantizer)
  - draw an orthogonal projection matrix P of size $d_b \times d$

  $\rightarrow$ this defines d_b random projection directions
  - for each Voronoi cell and projection direction, compute the median value for a learning set

- **On-line**: compute the binary signature b(x) of a given descriptor
  - project x onto the projection directions as $z(x) = (z_1, \ldots z_{db})$
  - $b_i(x) = 1$ if $z_i(x)$ is above the learned median value, otherwise 0

HE: Entropic weighting of Hamming distances



$$P(h(x,y) < \tau) = \frac{1}{2^{d_b}} \sum_{i=0}^{\tau} \binom{d_b}{i}$$

$$\mathrm{weight}(\tau) = -\log_2 \left(P(h(x,y) < \tau) \right)$$

Holidays dataset
http://lear/inrialpes.fr/people/jegou/data.php