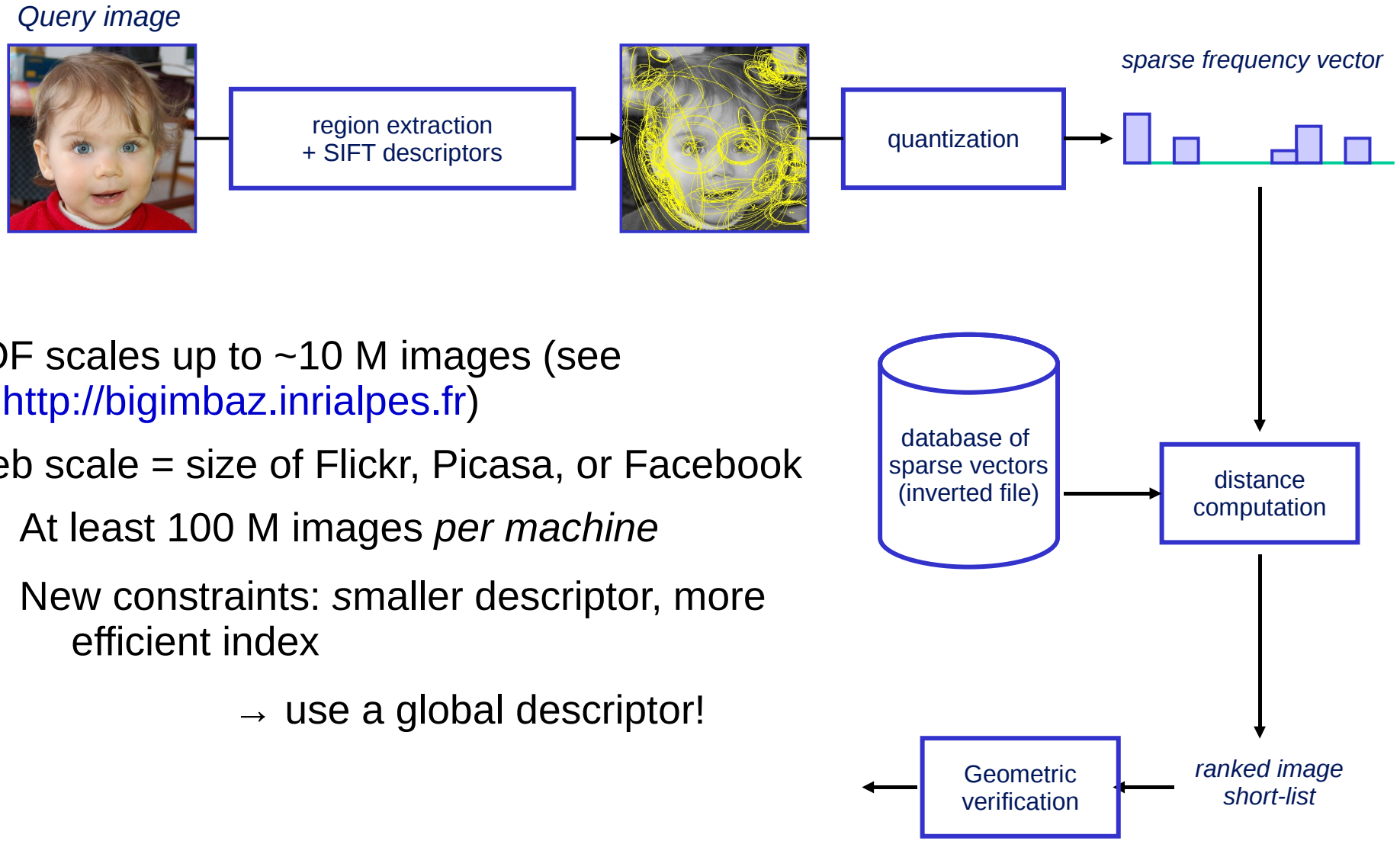
Evaluation of GIST descriptors for web-scale image search

Matthijs Douze

Hervé Jégou, Harsimrat Sandhawalia,
Laurent Amsaleg and Cordelia Schmid

INRIA Grenoble, France

Same-object recognition: bag-of-features



BOF scales up to ~10 M images (see http://bigimbaz.inrialpes.fr)

Web scale = size of Flickr, Picasa, or Facebook

At least 100 M images *per machine*

New constraints: smaller descriptor, more efficient index

→ use a global descriptor!

Video Google: A text retrieval approach to object matching in videos. J. Sivic and A. Zisserman, ICCV 2003

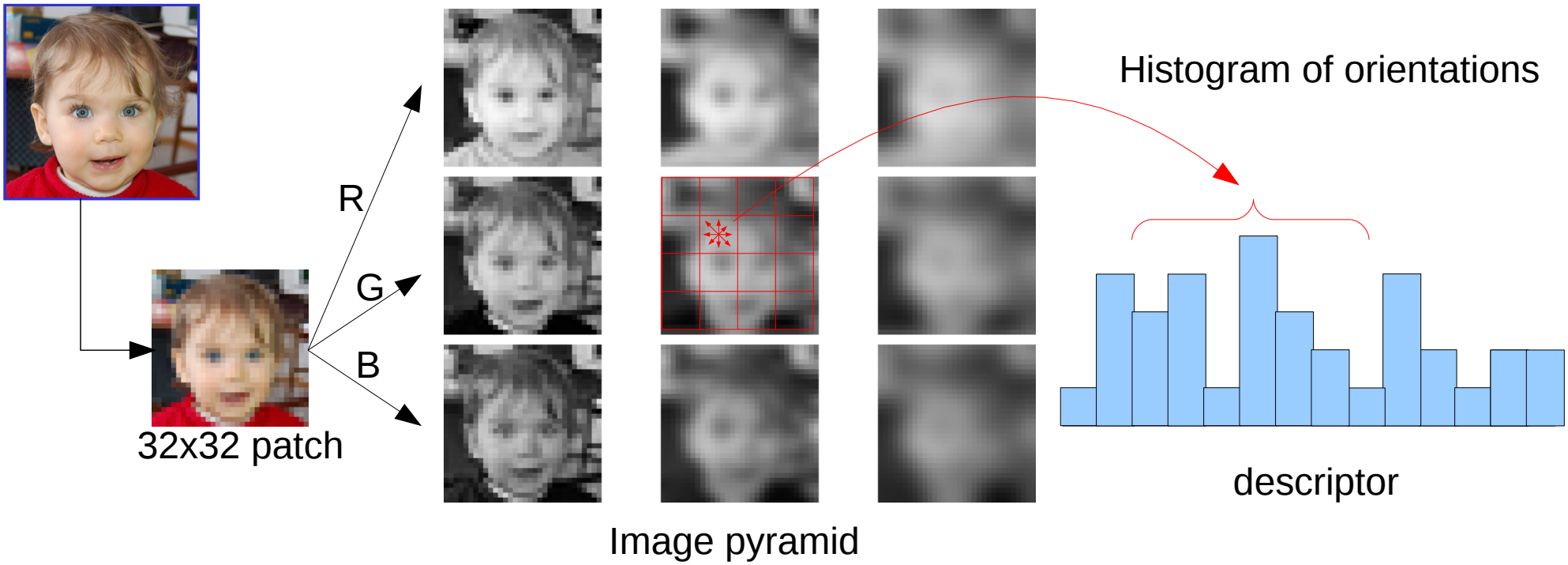
Overview

- GIST compared with BOF
  - Same-object recognition
  - Copy detection
- GIST at web scale
  - Indexing scheme
  - Results

GIST

Modeling the shape of the scene: a holistic representation of the spatial envelope, A. Olivia & A. Torralba, IJCV, 2001

Designed from perceptual experiments: properties like *naturalness* (vs. man-made), *roughness*, *openness*,...



Invariant to luminance transformations, blur, resize, etc.

Not invariant to translation, rotation, occlusion, crop, etc.

L2 distance to compare, NN-search

Evaluation for same-scene/object recognition

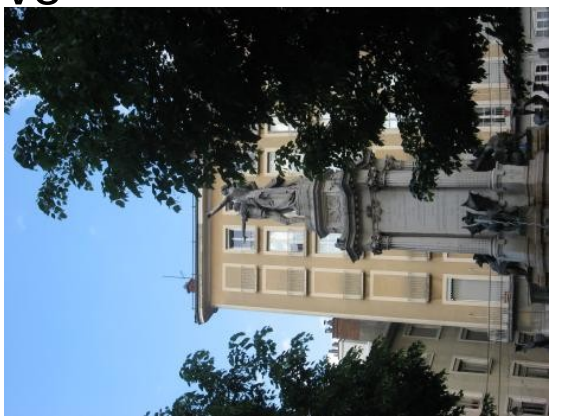
Test on the INRIA Holidays dataset

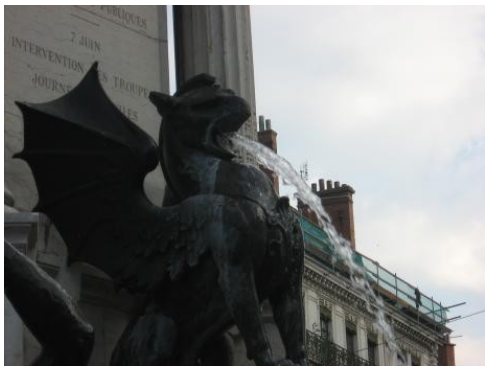
500 groups of images of the same scene

Hide them in up to 1 M distractor images from Flickr

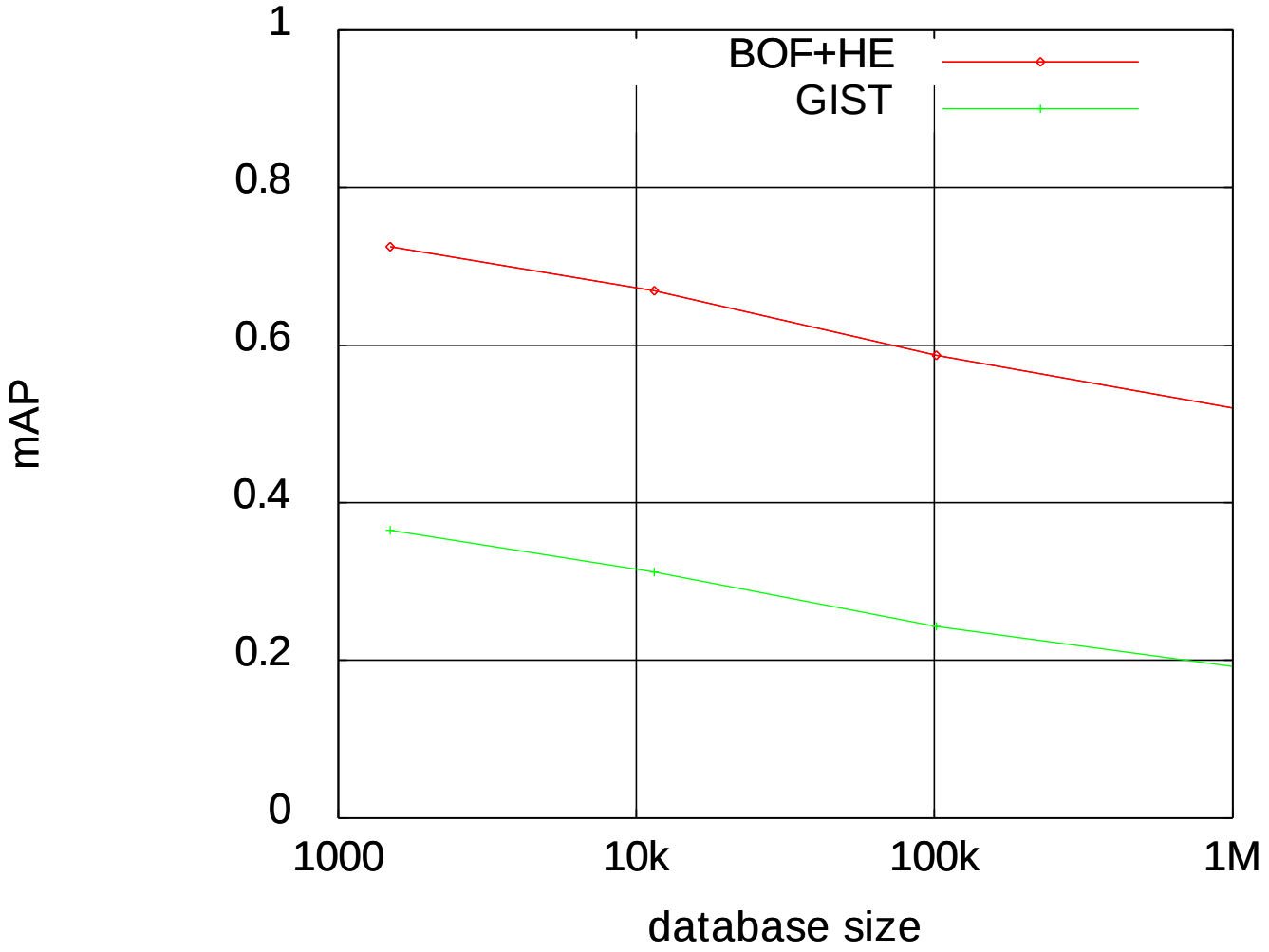
Queries = 1 image per group

Measure = mAP = mean area under precision-recall curve





Results on Holidays



Copy detection scenario

Simpler sub-problem: recognize transformed pictures

Useful for copy detection

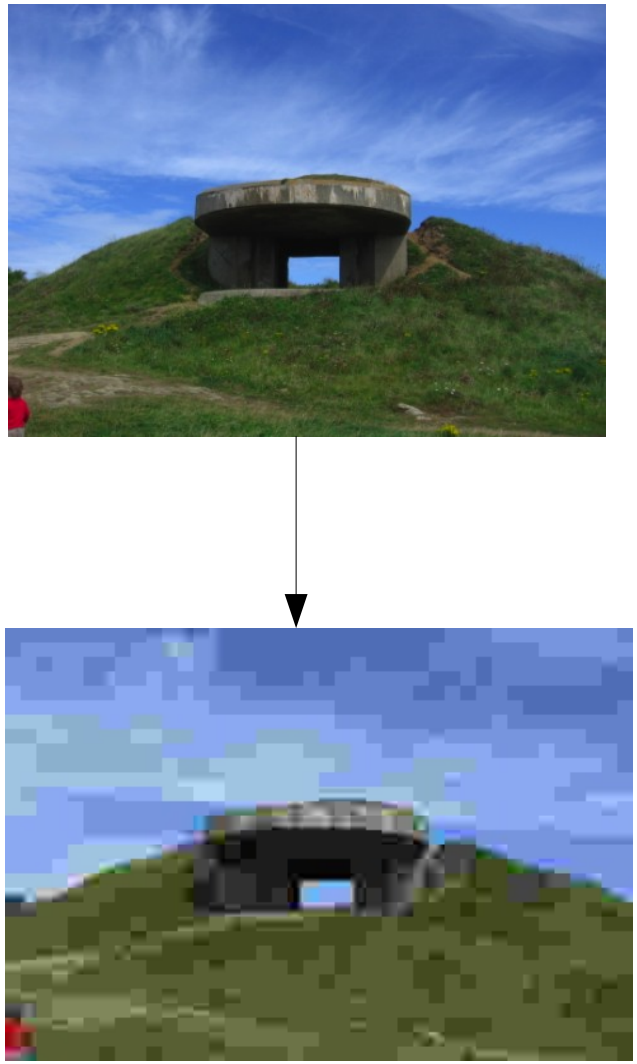
Our test dataset: Copydays

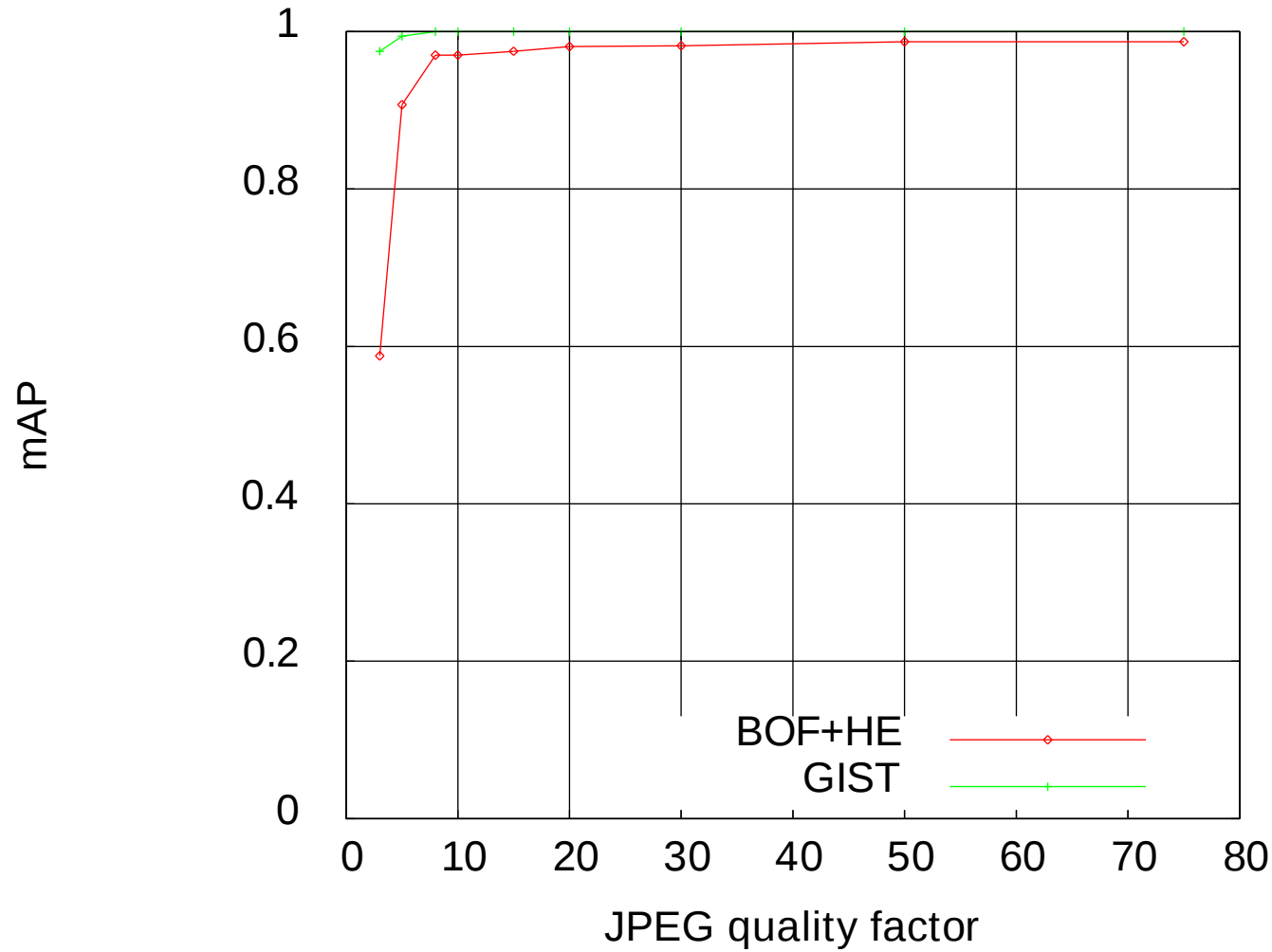
- 156 base images
- merged with 1 M distractors

Attacks

- JPEG compression
- Crop
- Queries = attacked images
- Performance measure is mAP

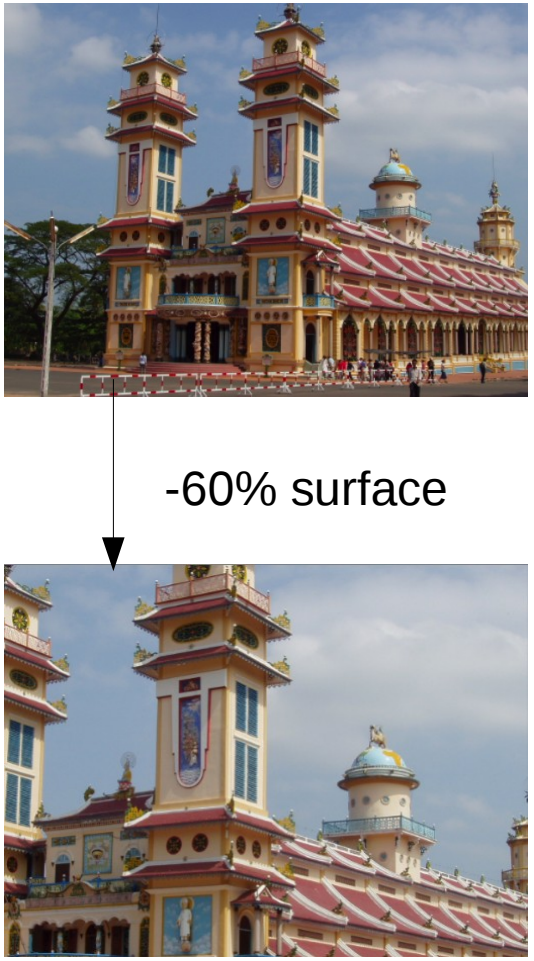
Results, JPEG compression

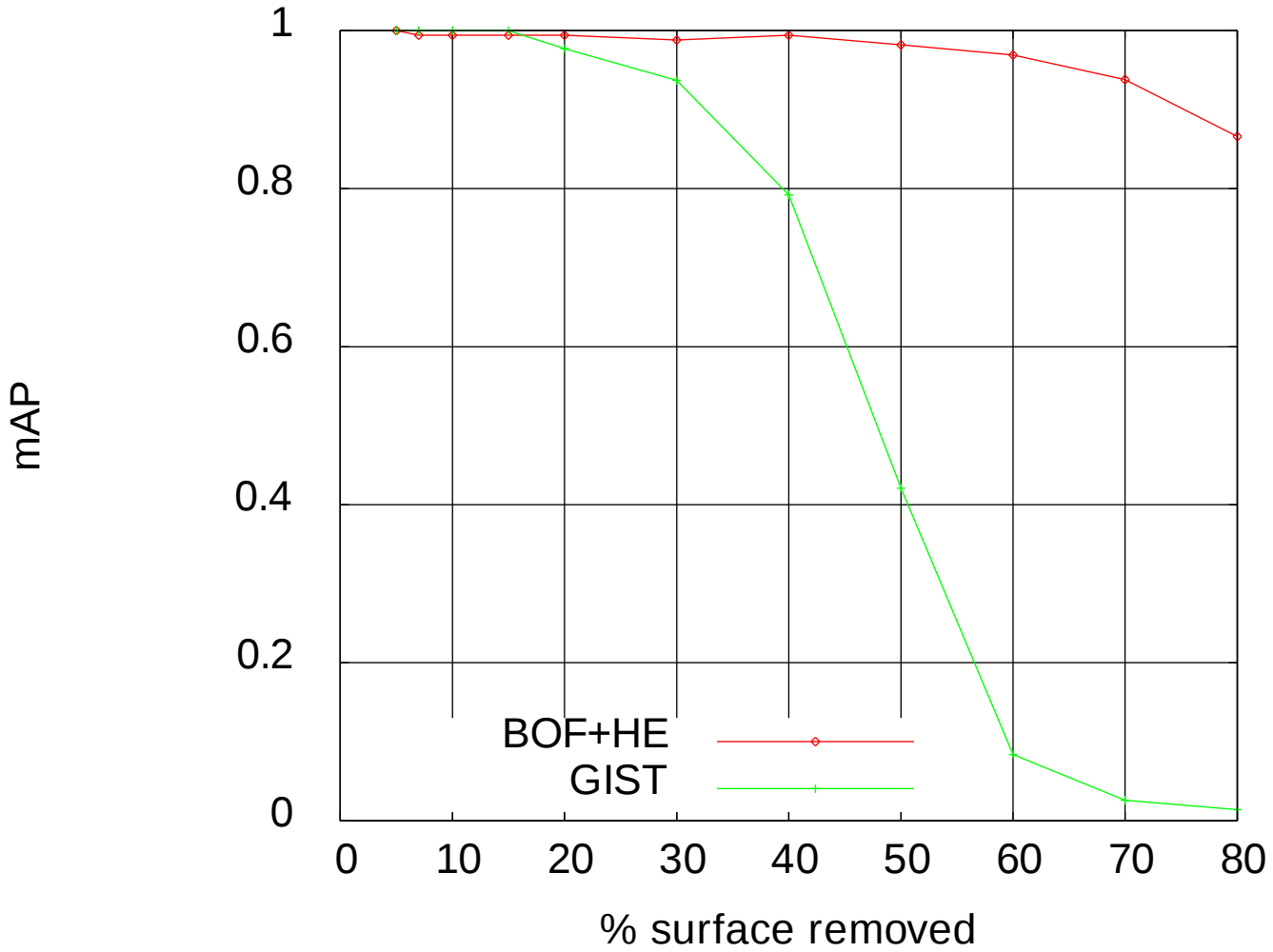




Results, crop

Remove (random) margin from image





Overview

- GIST compared with BOF
  - Same-object recognition
  - Copy detection
- GIST at web scale
  - Indexing scheme
  - Results

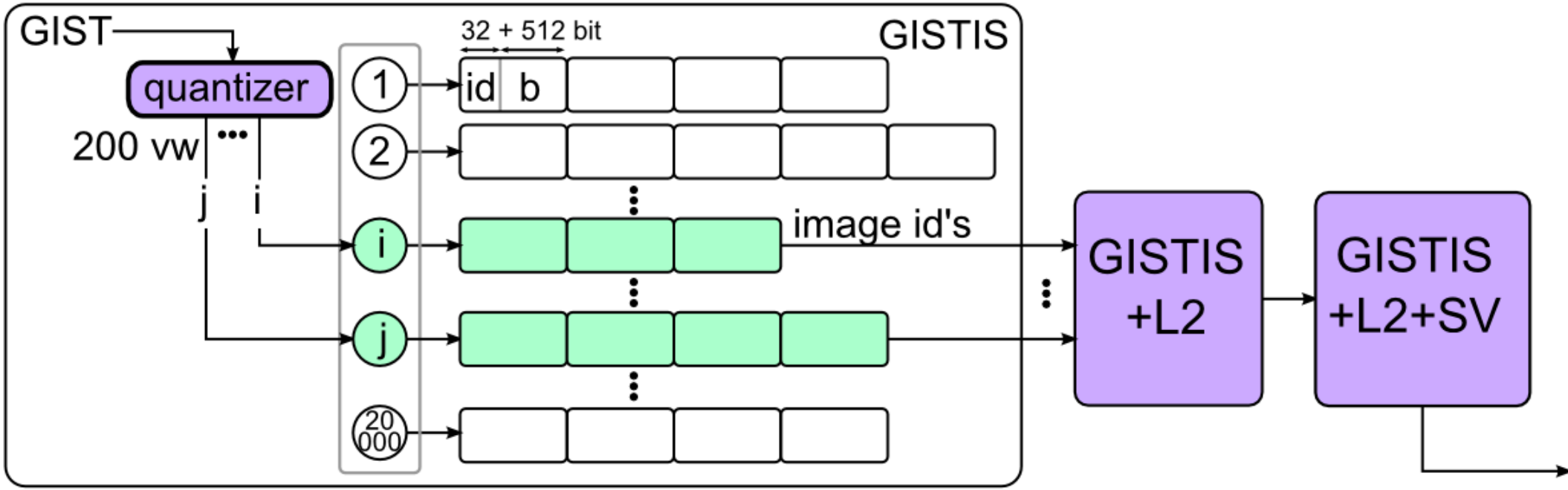
GIST indexing structure (GISTIS)

Raw GIST for 100 M images: 384 GB : too big!

→ quantize

Nearest-neighbor quantizer, 20000 centroids (learnt on independent dataset)

Index = Inverted file



Returns 1% of dataset: too many images!

Binary signatures & Hamming Embedding

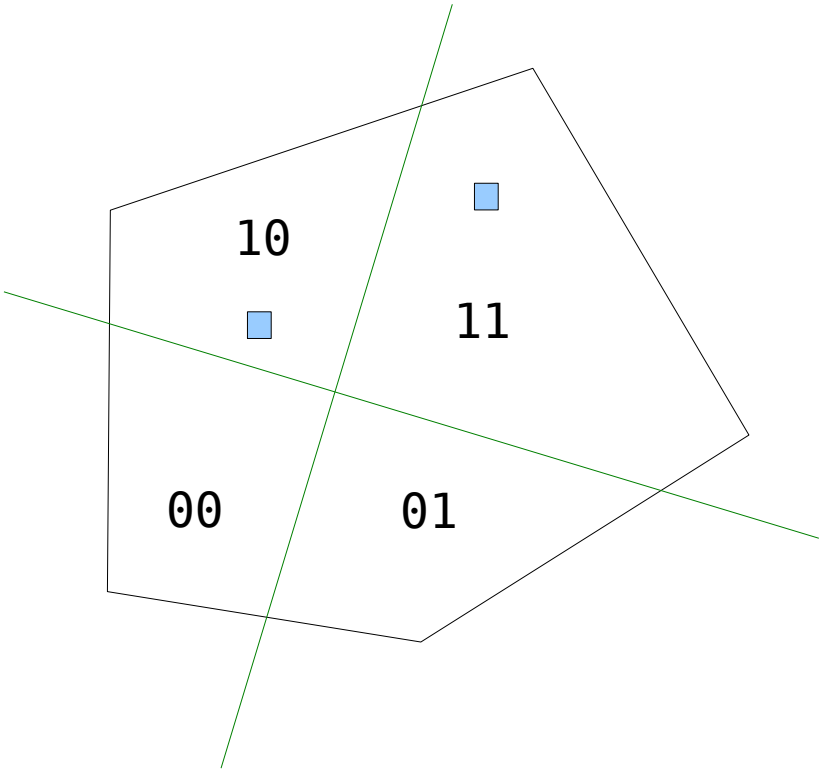
Quantization index is too coarse: need information about the point's position inside the quantization cell

Add a *binary signature* to the descriptor's quantization index

- Set of orthogonal hyperplanes
- Bit *i* = on which side of hyperplane *i* is the descriptor?
- 512 hyperplanes → 512 bits

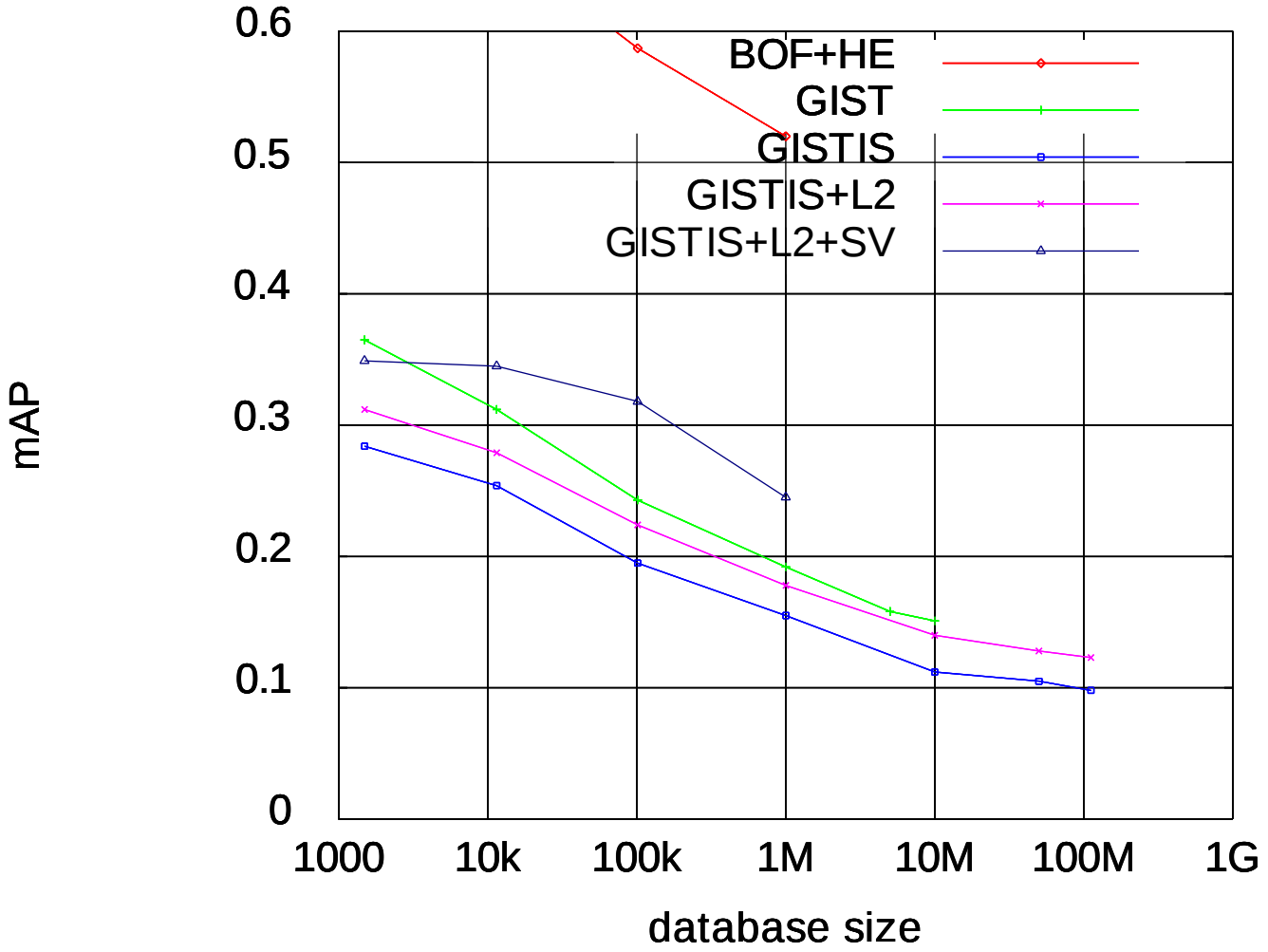
Signatures stored in inverted file, compared with Hamming distance

- Threshold: filter 94% images
- Output ordered by Hamming distance



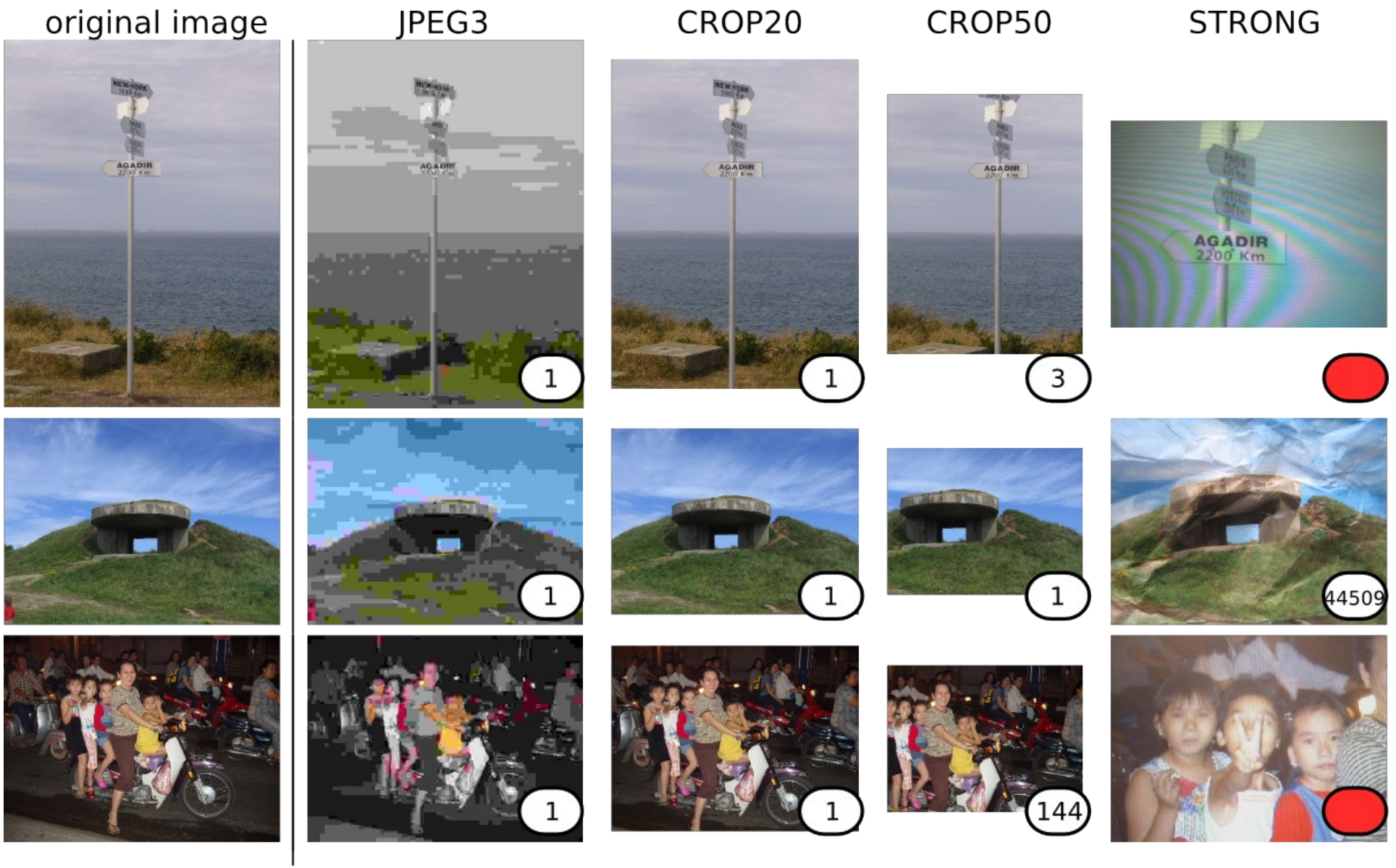
Hamming embedding and weak geometric consistency for large scale image search, H. Jégou, M. Douze, C. Schmid, ECCV 2008

Evaluation on Holidays



Thanks to A. Torralba for 80 M distractor pictures (32*32 pixels)...

GISTIS results on Copydays+110 M images



Timings & memory usage

Method	Dataset size	RAM usage	Search time
BOF + HE	1 M	24 GB	1.8 s
GIST	1 M	3.8 GB	1.3 s
GISTIS	1 M	68 MB	38 ms
GISTIS	100 M	6.8 GB	0.18 s
GISTIS+L2	100 M	6.8 GB	2.1 s

Conclusion

GIST recognizes:

- Image copies with small geometrical changes
- Scenes with small occlusions, same overall color...

GISTIS (quantized GIST + Hamming Embedding)

- Fast and compact
- Results close to full GIST descriptor

Current work: compact BOF representation

- Web-scale
- Handles geometrical changes
- Better results! (see our ICCV paper)

Thank you

See our demo tomorrow!