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

Query image

region extraction + SIFT descriptors

quantization

sparse frequency vector

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!

database of sparse vectors (inverted file)

distance computation

ranked image short-list

Geometric verification

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

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

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

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

![Graph showing mAP vs database size for BOF+HE and GIST]

- BOF+HE
- GIST
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

-60% surface

![Graph showing mAP vs. % surface removed for BOF+HE and GIST methods. BOF+HE remains relatively stable, while GIST degrades significantly with increasing surface removal.](image-url)
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 $\rightarrow$ 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

![Graph showing mAP vs database size for different methods: BOF+HE, GIST, GISTIS, GISTIS+L2, GISTIS+L2+SV.]

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

Evaluation of GIST for web-scale image search
GISTIS results on Copydays+110 M images
## Timings & memory usage

<table>
<thead>
<tr>
<th>Method</th>
<th>Dataset size</th>
<th>RAM usage</th>
<th>Search time</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOF + HE</td>
<td>1 M</td>
<td>24 GB</td>
<td>1.8 s</td>
</tr>
<tr>
<td>GIST</td>
<td>1 M</td>
<td>3.8 GB</td>
<td>1.3 s</td>
</tr>
<tr>
<td>GISTIS</td>
<td>1 M</td>
<td>68 MB</td>
<td>38 ms</td>
</tr>
<tr>
<td>GISTIS</td>
<td>100 M</td>
<td>6.8 GB</td>
<td>0.18 s</td>
</tr>
<tr>
<td>GISTIS+L2</td>
<td>100 M</td>
<td>6.8 GB</td>
<td>2.1 s</td>
</tr>
</tbody>
</table>
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!