Packing bag of features

Hervé Jégou, Matthijs Douze, Cordelia Schmid
LEAR, INRIA/LJK, Grenoble, France

Context: large scale image search

Same-scene search

Constraints:
- web-scale: index ~10-500 M images on a single server machine
- search in interactive time

Starting point: bag of features (BOF)

MiniBOF

Problems with BOF vectors:
- high-dimensional
- sparse - there is no efficient approximate nearest neighbor method for sparse vectors

Our solution: transform BOF vectors into dense and shorter vectors = miniBOFs

Which transformation? A simple one: aggregator

\[
A_1 = \begin{bmatrix}
1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
\end{bmatrix}
\]

- arbitrary choice of components to aggregate
- all aggregated components have same number of source components
- few collisions because input vectors are sparse

Why quantization:
- smaller memory footprint
- scan only fraction of dataset (via an inverted file)

2-level quantization and matching:
- nearest-neighbor quantizer trained with k-means.
- Multiple assignment for the query descriptors
  - images can match if they share a quantization index

- binary signature computed with Hamming Embedding
  - matching distance is Hamming distance between signatures

Indexing MiniBOF

Results

- Kentucky object recognition benchmark: 2550 groups of 4 images that represent the same object

<table>
<thead>
<tr>
<th>method</th>
<th>score</th>
<th>memory usage</th>
<th>image hits</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOF</td>
<td>2.92</td>
<td>66.62</td>
<td>9.926</td>
</tr>
<tr>
<td>compressed BOF</td>
<td>3.02</td>
<td>94.26</td>
<td>9.928</td>
</tr>
<tr>
<td>1 MiniBOF</td>
<td>2.07</td>
<td>20</td>
<td>94</td>
</tr>
<tr>
<td>8 MiniBOFs</td>
<td>2.72</td>
<td>160</td>
<td>383</td>
</tr>
<tr>
<td>64 MiniBOFs</td>
<td>2.93</td>
<td>1,280</td>
<td>1,076</td>
</tr>
</tbody>
</table>

- INRIA Holidays dataset + 1 million Flickr images: search 500 images of a scene among distractors

<table>
<thead>
<tr>
<th>method</th>
<th>mAP</th>
<th>memory usage</th>
<th>query time</th>
</tr>
</thead>
<tbody>
<tr>
<td>state-of-art</td>
<td>0.609</td>
<td>35.844</td>
<td>22656</td>
</tr>
<tr>
<td>BOF</td>
<td>0.313</td>
<td>8.885</td>
<td>2827</td>
</tr>
<tr>
<td>binary BOF</td>
<td>0.301</td>
<td>7.108</td>
<td>2562</td>
</tr>
<tr>
<td>1 MiniBOF</td>
<td>0.066</td>
<td>20</td>
<td>74</td>
</tr>
<tr>
<td>8 MiniBOFs</td>
<td>0.196</td>
<td>160</td>
<td>132</td>
</tr>
<tr>
<td>32 MiniBOFs</td>
<td>0.244</td>
<td>640</td>
<td>352</td>
</tr>
</tbody>
</table>

Example results:

Missing distance measures: fill in with Hamming distance expectation = half of binary signature length

MiniBOF combination

Combination = sum up Hamming distances

<table>
<thead>
<tr>
<th>db image</th>
<th>15036</th>
<th>72221</th>
<th>14394</th>
<th>53058</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOF</td>
<td>0</td>
<td>57</td>
<td>45</td>
<td>50</td>
</tr>
<tr>
<td>1 MiniBOF</td>
<td>0</td>
<td>48</td>
<td>53</td>
<td>50</td>
</tr>
<tr>
<td>2 MiniBOF</td>
<td>0</td>
<td>50</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>total</td>
<td>0</td>
<td>155</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

Query image

BOF descriptor

extraction of local descriptors + quantization

projection with vocabulary aggregators

迷你BOF descriptors

quantization + HE

query indexes

image scores

select smallest distances

Output: ordered image list