Better matching with fewer features:
The selection of *useful* features
in large database recognition problems

Panu Turcot, David G. Lowe
University of British Columbia

IEEE WS-LAVD
Oct. 4, 2009
A brief history

Large database image recognition
e.g. 1 million images, 2 billion descriptors = 285 GB

- **2003**: Sivic & Zisserman  
  Video Google
- **2005**: Nister & Stewenius  
  Scalable recognition
- **2007**: Philbin et al.  
  Oxford buildings
- **2007**: Chum et al.  
  Query expansion
- **2008**: Jegou et al.  
  Hamming embedding
- **2008**: Philbin et al.  
  Lost in quantization
- ...and more!
Bag-of-Words

- Cluster descriptors and represent with single center
- Order of magnitude reduction in memory
- Images ($I_j$) represented using TF-IDF weights ($x_{ij}$) (weighted visual word histogram)

\[ I_j = \left( x_{1j}, x_{1j}, \ldots, x_{Kj} \right) \]

\[ x_{ij} = \frac{n_{ij}}{\sum_i n_{ij}} \log \frac{N}{\sum_j |n_{ij}| > 0} \]

\[ n_{ij} \]

\[ i \]

\[ tf_{ij} \]

\[ idf_i \]
Can we do better?

- When generating image features we obtain many non-useful features
  - Unstable features
  - Non-distinctive features
  - Features from transient objects
Useful features

We define a **useful** feature to be an image feature which is:
- **robust** enough to be matched with a corresponding feature in the same object
- stable enough to exist in **multiple viewpoints**
- distinctive enough that the corresponding features are assigned to the **same visual word**
## Dataset

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Images</th>
<th>Descriptors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oxford Buildings ¹</td>
<td>5063</td>
<td>15,886,524</td>
</tr>
<tr>
<td>Flickr100K ²</td>
<td>100,000</td>
<td>206,748,897</td>
</tr>
</tbody>
</table>

- **Oxford Buildings**
  - 11 buildings
  - **Good**: building fully visible
  - **OK**: building at least 25% visible
  - **Junk**: building present but < 25% visible

- **Vocabulary²**: 200,000 words

¹ – Available at: [http://www.robots.ox.ac.uk/~vgg/data/oxbuildings/index.html](http://www.robots.ox.ac.uk/~vgg/data/oxbuildings/index.html)
² – Available at: [http://lear.inrialpes.fr/~jegou/data.php](http://lear.inrialpes.fr/~jegou/data.php)
Useful feature generation

1. A BOW image database containing the full feature set is constructed
Useful feature generation

1. A BOW image database containing the full feature set is constructed

2. Each image in the database is used as a query
Useful feature generation

1. A BOW image database containing the full feature set is constructed
2. Each image in the database is used as a query
3. The best $M$ images are each geometrically verified (we use $M=30$)
Useful feature generation

1. A BOW image database containing the full feature set is constructed
2. Each image in the database is used as a query
3. The best M images are each geometrically verified (we use M=30)
4. Features which are geometrically consistent are labelled as **useful** and preserved
Image graph

- Vertices represent database images
- Edges represent verified image matches
- Matching regions and geometric relationships not stored
Colours correspond to ground truth building labels

Image generated using GraphViz (http://www.graphviz.org)
Query Expansion

- Chum et. al. (2007)
- Following spatial verification, initial matches are used as queries to find further matches
- Relies on good initial TF-IDF search

Initial Query

TF-IDF matching result

Additional Queries

Merge results from all queries to improve recognition
Image augmentation

- Image graph: Leverage known relationships between database images to improve recognition
- Images represented using own *useful* features as well as those of neighbours
- Similar to average query expansion, but applied at initial query time
Additional considerations

• Singleton images
  – Some images result in no confirmed matches
    i.e. no useful features found
  Solution: Preserve a subset of initial features

• Near duplicate images
  – Images taken from the same viewpoint
    i.e. too many useful features found
  Solution: Use only the most useful

• Throttle # of descriptors per image
Testing procedure

• 5-fold cross validation
  – **Good** & **OK**: split into 5-folds
  – **Junk** images included in background
  – **Good** & **OK** images used to query resulting image database

• Average Precision
## Descriptor reduction

<table>
<thead>
<tr>
<th></th>
<th>Images</th>
<th>Original Descriptors</th>
<th>Useful Descriptors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Singleton images: 300 largest</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;Useful&quot;</td>
<td>16,033</td>
<td>49.14 M</td>
<td>1.92 M (3.92%)</td>
</tr>
<tr>
<td>Singleton</td>
<td>88,917</td>
<td>173.13 M</td>
<td>25.13 M (14.52%)</td>
</tr>
<tr>
<td>&quot;Useful&quot; + Singleton</td>
<td>104,950</td>
<td>222.27 M</td>
<td>27.06 M (12.17%)</td>
</tr>
<tr>
<td><strong>Singleton Images: discarded</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;Useful&quot;</td>
<td>16,033</td>
<td>49.14 M</td>
<td>1.92 M (3.92%)</td>
</tr>
<tr>
<td>Singleton</td>
<td>88,917</td>
<td>173.13 M</td>
<td>0 M (0%)</td>
</tr>
<tr>
<td>&quot;Useful&quot; + Singleton</td>
<td>104,950</td>
<td>222.27 M</td>
<td>1.92 M (0.87%)</td>
</tr>
</tbody>
</table>
Recognition

- **Methods:**
  - Original: All image descriptors
  - UF: Only *useful* features (no augmentation)
  - UF+1: *Useful* features + 1-adjacent neighbours
  - UF+2: *Useful* features + 2-adjacent neighbours
Recognition
Recognition

0.87% of orig. descriptors
All Souls
Summary

• Selection of *useful* features
  – Reduce the number of features stored
  – Ability to throttle memory usage

• Image augmentation
  – Leverage the image graph
  – Single query performance improvements

• Unsupervised
Arigato!