Image Retrieval 2.0

Filip Radenović

Center for Machine Perception
Czech Technical University in Prague
Outline

• 1.0: Standard image retrieval problems
  • Visually most similar
  • All visually similar

• 2.0: Beyond similarity retrieval
  • New (unseen) information
  • What/where is this?
  • What is interesting here?
  • Where should I look?

• 2.1: Image retrieval for 3D reconstruction
Standard Image Retrieval Evaluation

Query

Database size: 10 images
Relevant (total): 5 images

Results (ordered):

precision = #relevant / #returned
recall = #relevant / #total relevant

area under the curve
Average Precision (AP)
Is this what we want?

• Visually most similar
  – Results identical to query for large datasets

• All visually similar
  – Output of varying length
  – Ground truth hard to obtain

  – Users will never take a look at more than few tens of near-duplicate images!!!
1.0: Bag of Words (BoW) Image Retrieval
Bag of Words: Off-line Stage

Keypoint Detection

Local Appearance

Local Geometry

Geom. Vocabulary

$x_1, y_1, B_1$

$x_2, y_2, B_5$

$x_3, y_3, B_3$

...$

x_N, y_N, B_N$

SIFT Description [Lowe’04]

Visual Vocabulary

Visual Words

word_1, word_2, word_8, ...

word_948534, word_998125

graftiti

graftiti

graftiti
Bag of Words Image Representation

Term-frequency (tf) – visual word D is twice in the image

Images are represented by sparse vector / histogram of visual words present in them
Bag of Words: On-line Stage

**IN:** $q$

**BOW**

1
3
7
...
15999565

**geometries**

**1. Inverted file: posting list per visual word**

<table>
<thead>
<tr>
<th>word</th>
<th>image ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1 5 10 ... 7350125</td>
</tr>
<tr>
<td>2</td>
<td>2 7 12 ... 7399121</td>
</tr>
<tr>
<td>3</td>
<td>1 4 15 ... 7200190</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>3 7 10 ... 7012245</td>
</tr>
</tbody>
</table>

**2. Image ranking**

<table>
<thead>
<tr>
<th>score</th>
<th>image ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.87</td>
<td>5</td>
</tr>
<tr>
<td>0.75</td>
<td>1573</td>
</tr>
<tr>
<td>0.52</td>
<td>11202</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>0.001</td>
<td>32</td>
</tr>
</tbody>
</table>

**Shortlist: top N images**

**3. Spatial verification**

**image 5**

**4. Re-ranked shortlist**

<table>
<thead>
<tr>
<th>#inliers</th>
<th>image ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>247</td>
<td>1573</td>
</tr>
<tr>
<td>105</td>
<td>5</td>
</tr>
<tr>
<td>17</td>
<td>11202</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>2</td>
<td>75213</td>
</tr>
</tbody>
</table>

**5. Query expansion**

**query**

**image 1573**

**image 45**

**OUT:** $R$

...
Bag of Words Scoring

\[ \text{score} = \frac{q^\top x}{\|x\|} \]

Posting lists

- B query visual word 1
- D query visual word 2
- G query visual word 3
Geometric Re-ranking

Re-rank top ranked images (removing false positives)
- RANSAC

**NOTE:** Standard BoW score ranking performed without geometric information

**IMPORTANT:** Geometric verification crucial for query expansion

Sivic, Zisserman: Video Google, ICCV 2003

Philbin, Chum, Isard, Sivic, Zisserman: Object retrieval with large vocabularies and fast spatial matching, CVPR’07
Query Expansion

Results

Spatial verification

Query image

New results

New query

Chum, Philbin, Sivic, Isard, Zisserman: Total Recall..., ICCV 2007
Query Expansion: Step by Step

Query Image

Retrieved image

Originally not retrieved
Query Expansion: Step by Step
Query Expansion: Step by Step
2.0: Beyond Similarity Retrieval
Other Retrieval Problems

What is this?  ... and what is that?

Let’s zoom-in!
Different Retrieval Problems

Top: visually most similar

Query 1

Bottom: zoom-in

Query 2

Standard Retrieval and Details

query

rank: 1 2 32 64 65

DIFFICULT

query

rank: 1 2048 16384 81368

EASY
Zoom-in: On-line Stage

**IN:** q

BOW
1
3
7
... 15999565

geometries

1. Inverted file: posting list per visual word

<table>
<thead>
<tr>
<th>word</th>
<th>image ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1 5 10 ... 7350125</td>
</tr>
<tr>
<td>2</td>
<td>2 7 12 ... 7399121</td>
</tr>
<tr>
<td>3</td>
<td>1 4 15 ... 7200190</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>16777216</td>
<td>3 7 10 ... 7012245</td>
</tr>
</tbody>
</table>

2. Image ranking

<table>
<thead>
<tr>
<th>score</th>
<th>image ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.87</td>
<td>5</td>
</tr>
<tr>
<td>0.75</td>
<td>1573</td>
</tr>
<tr>
<td>0.52</td>
<td>11202</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>0.001</td>
<td>32</td>
</tr>
</tbody>
</table>

Geometry compressed in inverted file taken into account during scoring

Shortlist: top N images

3. Spatial verification

4. Re-ranked shortlist

<table>
<thead>
<tr>
<th>#inliers</th>
<th>zoom</th>
<th>image ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>37x</td>
<td>1573</td>
</tr>
<tr>
<td>105</td>
<td>17x</td>
<td>5</td>
</tr>
<tr>
<td>17</td>
<td>7x</td>
<td>11202</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>247</td>
<td>2x</td>
<td>75213</td>
</tr>
</tbody>
</table>

Problem specific ranking function, e.g. maximize scale change

5. Query expansion

Query expansion from already zoomed images
Zoom-out: On-line Stage

IN: q

1. Inverted file: posting list per visual word

<table>
<thead>
<tr>
<th>word</th>
<th>image ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1 5 10 ... 7350125</td>
</tr>
<tr>
<td>2</td>
<td>2 7 12 ... 7399121</td>
</tr>
<tr>
<td>3</td>
<td>1 4 15 ... 7200190</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>16777216</td>
<td>3 7 10 ... 7012245</td>
</tr>
</tbody>
</table>

2. Image ranking

<table>
<thead>
<tr>
<th>score</th>
<th>image ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.87</td>
<td>5</td>
</tr>
<tr>
<td>0.75</td>
<td>1573</td>
</tr>
<tr>
<td>0.52</td>
<td>11202</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>0.001</td>
<td>32</td>
</tr>
</tbody>
</table>

Geometry compressed in inverted file taken into account during scoring

3. Spatial verification

Shortlist: top N images

4. Re-ranked shortlist

<table>
<thead>
<tr>
<th>#inliers</th>
<th>zoom</th>
<th>image ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>1/37x</td>
<td>1573</td>
</tr>
<tr>
<td>105</td>
<td>1/17x</td>
<td>5</td>
</tr>
<tr>
<td>17</td>
<td>1/7x</td>
<td>11202</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>247</td>
<td>1/2x</td>
<td>75213</td>
</tr>
</tbody>
</table>

OUT: R

Problem specific ranking function, e.g. maximize scale change

5. Query expansion

Query expansion from already zoomed images
Zoom-in: Example
Zoom-in: Query Expansion
Zoom-in: Example
Zoom-in: Query Expansion
Zoom-in: Query Expansion
Zoom-out: Iterate
Zoom-out: Iterate
Zoom-out: Iterate
What is interesting here?
What should you not miss?
Highest Resolution Transform

Given a query and a dataset, for every pixel in the query image:
Find the database image with the maximum resolution depicting the pixel

Mikulík, Radenović, Chum, Matas: Efficient Image Detail Mining, ACCV 2014
What most people find interesting?

Most commonly photographed parts
Given a query and a dataset, for every pixel in the query image:
Find the frequency with which it is photographed in detail

Mikulík, Radenović, Chum, Matas: Efficient Image Detail Mining, ACCV 2014
All Details: On-line Stage

**IN: q**

1. Inverted file: posting list per visual word

- **BOW**
  - 1
  - 3
  - 7
  - ...
  - 15999565

- **geometries**

2. Image ranking

- score | image ID
- --- | ---
- 0.87 | 5
- 0.75 | 1573
- 0.52 | 11202
- ... | ...
- 0.001 | 32

3. Spatial verification

- #inliers | zoom | image ID
- --- | --- | ---
- 247 | 7x | 1573
- 105 | 2x | 5
- 17 | 37x | 11202
- ... | ... | ...
- 2 | 17x | 75213

4. Re-ranked shortlist

5. Query expansion

**OUT: R**
All Details: Hierarchical Query Expansion

IN: R

1. Grouped images
   - Group $G_i$
   - Group $G_n$

2. Geometric consistency
   - $A_{q,i} \approx A_{q,j} A_{j,i}$

OUT: $q_1, q_2, \ldots, q_n$

- Query $q_1$
  - Image 1573
  - Image 45

- Query $q_n$
  - Image 1761
  - Image 33
2.1: Image Retrieval for 3D Reconstruction
Structure-from-Motion 3D Reconstruction

• Few thousand images
  Exhaustive matching of all image pairs
  [Snavely, Seitz, Szeliski: Photo tourism, SIGGRAPH 2006]
  + High level of details reconstructed
  - Unfeasible for larger photo collections

• Few million images
  Matching images through standard image retrieval
  [Heinly, Schonberger, Dunn, Frahm: Reconstructing the World in Six Days, CVPR 2015]
  + Efficient and scalable image matching
  - Details not reconstructed
Retrieval for 3D Reconstruction

• Visually most similar search
  – Many near duplicates
  – Details lost

• Zoom-in and details search
  – Details retrieved
  – Transition images to match the details

• Zoom-out search
  – Viewpoint change
  – More context

• Sideways crawl
  – Significant viewpoint change
  – More context

Schoenberger, Radenović, Chum, Frahm: From Single Image Query to Detailed 3D Reconstruction, CVPR 2015
Sideways image crawl

Schoenberger, Radenović, Chum, Frahm: From Single Image Query to Detailed 3D Reconstruction, CVPR 2015
Sideways crawl: On-line Stage

1. Inverted file: posting list per visual word

<table>
<thead>
<tr>
<th>word</th>
<th>image ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1 5 10 ... 7350125</td>
</tr>
<tr>
<td>2</td>
<td>2 7 12 ... 7399121</td>
</tr>
<tr>
<td>3</td>
<td>1 4 15 ... 7200190</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>16777216</td>
<td>3 7 10 ... 7012245</td>
</tr>
</tbody>
</table>

2. Image ranking

<table>
<thead>
<tr>
<th>score</th>
<th>image ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.87</td>
<td>5</td>
</tr>
<tr>
<td>0.75</td>
<td>1573</td>
</tr>
<tr>
<td>0.52</td>
<td>11202</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>0.001</td>
<td>32</td>
</tr>
</tbody>
</table>

3. Spatial verification

4. Re-ranked shortlist

<table>
<thead>
<tr>
<th>#inliers</th>
<th>zoom</th>
<th>image ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>247</td>
<td>7x</td>
<td>1573</td>
</tr>
<tr>
<td>105</td>
<td>2x</td>
<td>5</td>
</tr>
<tr>
<td>17</td>
<td>37x</td>
<td>11202</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>2</td>
<td>17x</td>
<td>75213</td>
</tr>
</tbody>
</table>

5. Query expansion

Using geometry to find adequate features for expansion (left-right)

Building an expanded query using only sideways features
Sideways Left: Step by Step
Sideways Left: Step by Step
Retrieval for 3D Reconstruction

See our video at:

https://youtu.be/Dlv1aGKqSlk

VIDEO

Schoenberger, Radenović, Chum, Frahm: From Single Image Query to Detailed 3D Reconstruction, CVPR 2015
Localization: Most Similar Retrieval

Application: Camera Elevation Estimation
• Automatic elevation estimation from image content
• Location recognition in Alps
• Inferring height from a training dataset by using recognized location

<table>
<thead>
<tr>
<th>Method</th>
<th>test dataset (13148 images)</th>
<th>user experiment set (50 images)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>801.49; 786.42</td>
<td>1383.64; 1154.43</td>
</tr>
<tr>
<td>Human</td>
<td>-</td>
<td>879.95</td>
</tr>
<tr>
<td>CNN</td>
<td>537.11</td>
<td>709.10</td>
</tr>
<tr>
<td>BOW</td>
<td>601.63</td>
<td>757.76</td>
</tr>
<tr>
<td>mVocab</td>
<td>610.36</td>
<td>811.00</td>
</tr>
<tr>
<td>BOW+mVocab</td>
<td>564.14</td>
<td>646.89</td>
</tr>
<tr>
<td>BOW+CNN</td>
<td><strong>500.44</strong></td>
<td><strong>531.05</strong></td>
</tr>
</tbody>
</table>

Čadík, Vašíček, Hradiš, Radenović, Chum: Camera Elevation Estimation from a Single Mountain Landscape Photograph, BMVC 2015
Summary

Visually most similar

Zoom-in / details

Zoom-out

Sideways right
Thank you!

Questions?