MULTI-SCALE KEYPOINT MATCHING

Sina Lotfian, Hassan Foroosh
University of Central Florida
Computer Science Department
Single-Scale Matching

- There is not enough information for matching in a single scale.
- Humans use clues from different scales when matching points.
- How can we use data from multiple scales with least computational penalty?
Multi-Scale Matching

- Exploiting data from high and low scales
- Biologically plausible
- High scales contain Holistic information
- Lower Scale contain More localized data
- Iterative pruning
Multi-Scale Matching Pipeline

- Keypoints are extracted using Difference of Gaussian (DoG).
Multi-Scale Matching Pipeline

- Each keypoint is mapped into a feature space
- Descriptors can be generated from any method
  - Hand-Engineered features
  - Learned features
- This is done for $N$ scales for each keypoint
Multi-Scale Matching Pipeline

- Different scales are concatenated to form a pyramid
- Higher Scales are on Top → Lower Resolution
- Lower Scales Bottom → Higher Resolution
Multi-Scale Matching Pipeline
Top-Bottom Matching

- The goal is to choose the best match among points in the second image for a point in the first image.
- Start by considering all points in the second image.
- Use higher scales for early rejection.
- Repeat until one point is remaining.
Maximum Margin Nearest Neighbor

- Each point has several possible matches
- Which point should be rejected?
Maximum Margin Nearest Neighbor

- Threshold on distance
- Ant point inside the circle is a possible match
- Any point outside the circle is rejected
- Margin of confidence: Difference between distance of nearest rejected point $d_1$ and most distance accepted point $d_2$
Results

- Mean average Precision

- We choose SIFT\([1]\) as hand-crafted and SOSNet\([2]\) as Learned feature

- ASV-SIFT and DSP-SIFT as competing multi-scale approaches

### MAP for different datasets

<table>
<thead>
<tr>
<th>Method</th>
<th>Oxford</th>
<th>Webcam</th>
<th>HPVatches</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIFT ([1])</td>
<td>58.82</td>
<td>16.29</td>
<td>42.28</td>
</tr>
<tr>
<td>SOSNet ([2])</td>
<td>63.54</td>
<td>18.94</td>
<td>46.39</td>
</tr>
<tr>
<td>Root SIFT ([3])</td>
<td>60.11</td>
<td>18.29</td>
<td>44.05</td>
</tr>
<tr>
<td>Raw Patch</td>
<td>30.67</td>
<td>3.97</td>
<td>21.11</td>
</tr>
<tr>
<td>LIOP ([4])</td>
<td>40.54</td>
<td>1.79</td>
<td>33.87</td>
</tr>
<tr>
<td>DSP-SIFT ([5])</td>
<td>60.43</td>
<td>22.28</td>
<td>45.17</td>
</tr>
<tr>
<td>ASV-SIFT ([6])</td>
<td>60.94</td>
<td>23.16</td>
<td>45.53</td>
</tr>
<tr>
<td>SIFT + ours(bottom-top)</td>
<td>61.02</td>
<td>25.64</td>
<td>46.87</td>
</tr>
<tr>
<td>SIFT + ours(top-bottom)</td>
<td>60.41</td>
<td>25.10</td>
<td>47.04</td>
</tr>
<tr>
<td>SOSNet ([2]) + ours(top-bottom)</td>
<td>68.03</td>
<td>27.05</td>
<td>51.31</td>
</tr>
<tr>
<td>SOSNet ([2]) + ours(bottom-top)</td>
<td>68.82</td>
<td>27.37</td>
<td>50.63</td>
</tr>
</tbody>
</table>

Precision-Recall plot for Oxford dataset (View-Point change)
Result

1. Feature extraction time:
   - Grows linearly with number of keypoints
   - Grow linearly with number of sampled scales

2. Matching time:
   - Grows quadratically with number of keypoints
   - For the proposed method increases linearly with number of sampled scales
Results

- Time Analysis:
References


