Experimental study on performance of view-based pose estimation

Toru Tamaki
Hiroyuki Okugawa
Toshiyuki Amano
Kazufumi Kaneda
View-based pose estimation

Learning

Images $x_1, x_2, \ldots$

poses $\theta_1, \theta_2, \ldots, \theta_n$

Estimation

$x$

$\hat{\theta}$
Learning relations

- **Learning set**
  - \( \{ \theta_j, x_j \} \)
  - \((i = 1, 2, \ldots, n)\)

- **Relations**
  - Nonlinear: \( \theta_j = f(x_j) \)
  - Linear: \( \theta_j = Fx_j \)

- **Estimation**
  - Nonlinear: \( \theta = f(x) \)
  - Linear: \( \theta = Fx \)

- **Nonlinear methods**
  - Parametric
    - Eigenspace method
      - (Murase, 1995)
  - Kernels
    - (Melzer, 2003)
    - (Ando, 2005)
  - Manifold learning
Learning relations

- Learning set
  - \( \{ \theta_j, x_j \} \)
    - \((i=1, 2, \ldots, n)\)

- Relations
  - Nonlinear: \( \theta_j = f(x_j) \)
  - Linear: \( \theta_j = Fx_j \)

- Estimation
  - Nonlinear: \( \theta = f(x) \)
  - Linear: \( \theta = Fx \)

- Linear methods
  - Linear regression
    - (Okatani, 2000)
  - Cyclic permutation
    - (Tamaki, 2007)
  - EbC
    - (Amano, 2006/2007)
Overview of EbC

- EbC: "Estimation-by-Completion"

- Learn
  - Image part $x_j$
  - Parameter part $p_j$
  - Compute Eigenspace

- Estimate pose
  - A test image has no parameter part
  - Completed as missing image area
Questions to investigate

- Performance depends on the number of learning images.
  - Few images: bad estimation
  - Many images: better performance
- Is it really? How many images are enough?
Questions to investigate

- Performance depends on the number of learning images.
- What is an appropriate set of images when we fix the number of images?
  - Any set is enough?
### Learning image set

**Definition of a learning set:**

\[ S_{i,s} = \{ x_{ik+s} \} \]

- \( x_{\theta} \) : images at \( \theta \)
- \( i \) : sample span [deg]
- \( k = 0, 1, \ldots, n_i - 1 \)
- \( n_i = 360/i \)
- \( s \) : start angle [deg]

**Example:**

<table>
<thead>
<tr>
<th></th>
<th>( x_0 )</th>
<th>( x_5 )</th>
<th>( x_{10} )</th>
<th>( x_{15} )</th>
<th>( x_{20} )</th>
<th>( x_{25} )</th>
<th>( x_{30} )</th>
<th>( x_{35} )</th>
<th>( x_{40} )</th>
<th>( x_{45} )</th>
<th>( x_{50} )</th>
<th>( x_{55} )</th>
<th>( x_{60} )</th>
<th>\ldots</th>
</tr>
</thead>
<tbody>
<tr>
<td>( S_{20,0} )</td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
<td><img src="image9.png" alt="Image" /></td>
<td><img src="image10.png" alt="Image" /></td>
<td><img src="image11.png" alt="Image" /></td>
<td><img src="image12.png" alt="Image" /></td>
<td><img src="image13.png" alt="Image" /></td>
<td><img src="image14.png" alt="Image" /></td>
</tr>
<tr>
<td>( S_{20,5} )</td>
<td><img src="image15.png" alt="Image" /></td>
<td><img src="image16.png" alt="Image" /></td>
<td><img src="image17.png" alt="Image" /></td>
<td><img src="image18.png" alt="Image" /></td>
<td><img src="image19.png" alt="Image" /></td>
<td><img src="image20.png" alt="Image" /></td>
<td><img src="image21.png" alt="Image" /></td>
<td><img src="image22.png" alt="Image" /></td>
<td><img src="image23.png" alt="Image" /></td>
<td><img src="image24.png" alt="Image" /></td>
<td><img src="image25.png" alt="Image" /></td>
<td><img src="image26.png" alt="Image" /></td>
<td><img src="image27.png" alt="Image" /></td>
<td><img src="image28.png" alt="Image" /></td>
</tr>
<tr>
<td>( S_{20,10} )</td>
<td><img src="image29.png" alt="Image" /></td>
<td><img src="image30.png" alt="Image" /></td>
<td><img src="image31.png" alt="Image" /></td>
<td><img src="image32.png" alt="Image" /></td>
<td><img src="image33.png" alt="Image" /></td>
<td><img src="image34.png" alt="Image" /></td>
<td><img src="image35.png" alt="Image" /></td>
<td><img src="image36.png" alt="Image" /></td>
<td><img src="image37.png" alt="Image" /></td>
<td><img src="image38.png" alt="Image" /></td>
<td><img src="image39.png" alt="Image" /></td>
<td><img src="image40.png" alt="Image" /></td>
<td><img src="image41.png" alt="Image" /></td>
<td><img src="image42.png" alt="Image" /></td>
</tr>
<tr>
<td>( S_{20,15} )</td>
<td><img src="image43.png" alt="Image" /></td>
<td><img src="image44.png" alt="Image" /></td>
<td><img src="image45.png" alt="Image" /></td>
<td><img src="image46.png" alt="Image" /></td>
<td><img src="image47.png" alt="Image" /></td>
<td><img src="image48.png" alt="Image" /></td>
<td><img src="image49.png" alt="Image" /></td>
<td><img src="image50.png" alt="Image" /></td>
<td><img src="image51.png" alt="Image" /></td>
<td><img src="image52.png" alt="Image" /></td>
<td><img src="image53.png" alt="Image" /></td>
<td><img src="image54.png" alt="Image" /></td>
<td><img src="image55.png" alt="Image" /></td>
<td><img src="image56.png" alt="Image" /></td>
</tr>
</tbody>
</table>
Performance evaluation

Root mean square error (RMSE):

$$RMSE_{i,s} = \sqrt{\frac{1}{72 - n_i} \sum_{x_j \in S_{i,s}} (\hat{\theta}_j - \theta_j)^2}$$

\(\theta\): true angle
\(\hat{\theta}\): estimated angle

Exclude learned images

sample spans:

\(i = 5, 10, 15, 20, 30, 40, 45, 60, 90, 120\)

(divisors of 360 [deg])
Experimental results 1: moderate case

COIL-20 Object 4

- Error increases monotonically
- Starting angle doesn’t affect the performance

<table>
<thead>
<tr>
<th>Sample Span [deg]</th>
<th>RMSE [deg]</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td></td>
</tr>
<tr>
<td>45</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td></td>
</tr>
<tr>
<td>90</td>
<td></td>
</tr>
<tr>
<td>120</td>
<td></td>
</tr>
</tbody>
</table>

$i$: sample span [deg]

RMSE [deg]
Experimental results 2: performance dip at 40 deg.

More the images, Better the performance

- Not monotonically
- Error at $i=40$ [deg] is very large: 9 images are learned

The number of images is important!
Examples of learning sets

$S_{60.0}$  6 images

$S_{45.0}$  8 images

$S_{40.0}$  9 images

Worst!

$S_{30.0}$  12 images
Objects that have performance dip at 40 deg.

What property affect the performance?
Future work....
Experimental results 3: keeping good performance

- Error increases monotonically
- Error at $i=120$ [deg] is so small: only 3 images are used for learning
Objects that keep good performance

- Round shape may affect the performance
- Also future work...

COIL-20 Object 15

COIL-20 Object 12

COIL-20 Object 20
Conclusions

- Performance evaluation of EbC
  - a view-based pose estimation
- Experimental results:
  - Some objects have the performance dip
  - Some objects keep good performance
- Future work
  - To investigate the relationship between performance and object shape