Regression-based Hand Pose Estimation from Multiple Cameras
Teófilo E. de Campos and David W. Murray
Department of Engineering Science, University of Oxford
www.robots.ox.ac.uk/ActiveVision

Introduction
- Objective: to estimate hand pose in 3D for real-time applications.
- Current Approaches:
  - Model-based generative methods: require (re-)initialisation and tracking can easily get lost with sudden movements;
  - Classification-based methods: require a large training set to be loaded into memory for classification;
  - Regression-based methods: provide a continuous map from image to 3D pose. They are fast and can work with global image descriptors.
- Problem: hands present a high degree of self-occlusion.
- Our approach: extend a regression-based method for multiple views.

Multiple View Image Descriptors

Rotation Invariant Shape Contexts
- Pre-processing: skin is detected using a histogram-based classifier in the CbCr colour space (from YCbCr).
- Shape Contexts
  input image silhouette contour → a shape context

Methods for Rotation Invariance
Classification Results
input
not-invariant tangent centroid principal axis

Image Descriptor Manifolds

Rotation Invariant Shape Contexts
- Pre-processing: skin is detected using a histogram-based classifier in the CbCr colour space (from YCbCr).
- Shape Contexts
  input image silhouette contour → a shape context

Methods for Rotation Invariance
Classification Results
input
not-invariant tangent centroid principal axis

Image Descriptor Manifolds

Multiple View Image Descriptors

Evaluting the Descriptor with Real Images

Regressor Learning
- Given a set of pairs of image measurements $x_i$ and 3D poses $y_i$, the system is described as:
$$y_i = \sum_{j} a_j \phi(x_i) + \epsilon = \Lambda(x_i) + \epsilon$$
where $\phi$ is a basis function and $\Lambda$ is a weights vector which determines the relevance of each basis function.
- Using a matrix representation, the probem is described as:
$$\Lambda = \arg \min_{\Lambda} \left\{ \| Y - \Lambda \Phi \| + \epsilon \right\}$$
(2)
- To solve this for multi-dimensional target spaces $y$, we use Agarwal and Triggs’ method, which first initialises $\Lambda$ with ridge regression and iteratively adjusts vectors of $\Lambda$ using a quadratic residual function $R(\Lambda)$ that penalises columns of $\Lambda$ with large norm.

Weights Matrices
- A matrix from unidimensional regressions
- A matrix from A&T regression method

Experiments and Results

Number of Relevance Vectors

Feature Selection
- Selecting two features from a multiple view database (open-close)

Quantitative Results with Synthetic Images

Open-close data set
- Using two features
- Using 12 features and 29 samples

Summary of the Quantitative Results

<table>
<thead>
<tr>
<th># Views</th>
<th>Kernal</th>
<th>PCs</th>
<th>Spks</th>
<th>Error</th>
<th>STD</th>
<th>RRMSE</th>
<th>Target</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Lin</td>
<td>33</td>
<td>26</td>
<td>6.8</td>
<td>5.5</td>
<td>0.64</td>
<td></td>
<td>115.7</td>
</tr>
<tr>
<td></td>
<td>Gauss.</td>
<td>40</td>
<td>24</td>
<td>4.9</td>
<td>4.3</td>
<td>0.59</td>
<td></td>
<td>192</td>
</tr>
<tr>
<td>2</td>
<td>Lin</td>
<td>40</td>
<td>20</td>
<td>8.5</td>
<td>7.3</td>
<td>0.65</td>
<td></td>
<td>115.7</td>
</tr>
<tr>
<td></td>
<td>Gauss.</td>
<td>40</td>
<td>20</td>
<td>6.5</td>
<td>5.7</td>
<td>0.65</td>
<td></td>
<td>192</td>
</tr>
<tr>
<td>3</td>
<td>Lin</td>
<td>40</td>
<td>20</td>
<td>9.5</td>
<td>8.2</td>
<td>0.65</td>
<td></td>
<td>115.7</td>
</tr>
<tr>
<td></td>
<td>Gauss.</td>
<td>40</td>
<td>20</td>
<td>7.5</td>
<td>6.8</td>
<td>0.65</td>
<td></td>
<td>192</td>
</tr>
</tbody>
</table>

Qualitative Results with Real Images

Using all Features and Samples with Gaussian Kernels

Using 32 Features and 38 Samples

Time Measurements

Training Phase
- Kernel | Features | Samples | Time (s) |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Lin</td>
<td>32</td>
<td>38</td>
<td>115.7</td>
</tr>
<tr>
<td>Gauss.</td>
<td>32</td>
<td>38</td>
<td>192</td>
</tr>
</tbody>
</table>

Application

Overview

Conclusion
The proposed method to combine multiple views for regression improves the results for estimation of hand pose in 3D. This method also increases the sparsity of the regressor.

Acknowledgments
This work was supported by the Brazilian government CAPES Foundation and by EPSRC (UK).
Contact: [teo,dwm]@robots.ox.ac.uk