Head Pose Estimation Via Probabilistic High-Dimensional Regression

Vincent Drouard ¹ Silèye Ba ¹ Georgios Evangelidis ¹
Antoine Deleforge ² Radu Horaud ¹

¹Team Perception - Inria Grenoble Rhône-Alpes, France
²Friedrich-Alexander-Universität, Erlangen, Germany

September 28, 2015, Québec, Canada

Work supported by EU-FP7 ERC Advanced Grant VHIA (#340113) and STREP project EARS (#609645)
Motivation

Visual cue in Human-Robot and Human-Human Interaction
Problem definition

Data  Input  Output

Face bounding box

$x \in \mathbb{R}^3$

$x$: pose vector
Problem formulation

Annotated Bounding Box → Ideal Scenario

Detected Bounding Box → Reality

\[
y \in \mathbb{R}^{1500}_{\text{high dimensional}}\]

High-Dimensional Regression

\[
x \in \mathbb{R}^k
\]

Head Pose → Bounding Box offset

Affected by offset in detected bounding box
High-Dimensional Regression

Problems:

- A lot of parameters to estimate
- \( y \rightarrow x \) might be non linear
- \( y \) and \( x \) are obtained by measurement \( ightarrow \) might be noisy

Standard solution

- Step 1: dimension reduction, \( y \rightarrow x' \)
- Step 2: regression, \( x' \rightarrow x \)
- Head-pose information may be lost when dimensionality reduction is performed
High-Dimensional Regression - Solution

- Inverse problem (training): easier to solve
- Forward solution (testing): closed-form
Training: Inverse Regression (I)

\[ y = \sum_{k=1}^{K} \mathbb{I}\{Z=k\} \left( A_k \mathbf{x} + b_k + e_k \right) \]

- \(A_k, b_k\): parameters of the \(k^{th}\) affine transformation
- \(e_k\): zero-mean noise, \(e_k \sim \mathcal{N} (0, \Sigma_k)\)
- \(\mathbb{I}\): indicator function
- \(Z\): discrete latent variable selecting the affine transformation
Training: Inverse Regression (II)

- Probabilistic model

\[
P (y | x, Z = k) = \mathcal{N} (y; \mathbf{A}_k x + b_k, \Sigma_k)
\]
\[
P (x | Z = k) = \mathcal{N} (x; \mathbf{c}_k, \Gamma_k)
\]
\[
P (Z = k) = \pi_k
\]
Training: Inverse Regression (III)

- Inverse Regression

\[
P (\mathbf{y} | \mathbf{x}; \theta) = \sum_{k=1}^{K} \frac{\pi_k \mathcal{N} (\mathbf{x}; \mathbf{c}_k, \Gamma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N} (\mathbf{x}; \mathbf{c}_j, \Gamma_j)} \mathcal{N} (\mathbf{y}; A_k \mathbf{x} + b_k, \Sigma_k)
\]

- \( \theta = \{ A_k, b_k, \Sigma_k, c_k, \Gamma_k, \pi_k \}_{k=1}^{K} \)

- Estimated using EM algorithm
Forward Testing

Bayesian inversion of the model

\[
P (x|y; \theta^*) = \sum_{k=1}^{K} \frac{\pi^*_k \mathcal{N} (y; c^*_k, \Gamma^*_k)}{\sum_{j=1}^{K} \pi^*_j \mathcal{N} (y; c^*_j, \Gamma^*_j)} \mathcal{N} \left( A^*_k y + b^*_k, \Sigma^*_k \right)
\]

- \( \theta^* = \{ A^*_k, b^*_k, \Sigma^*_k, c^*_k, \Gamma^*_k, \pi^*_k \}^{K}_{k=1} \) obtained analytically using \( \theta \)

- \( \hat{x} = \mathbb{E} (x|y; \theta^*) = \sum_{k=1}^{K} \rho^*_k (A^*_k y + b^*_k) \) Fast evaluation
Summary of the model

- Closed-form solution for estimating inverse regression parameters
  - high dimension = 1500 and Low dimension = 3 (or 5) (K=50)
  - Proposed inverse training: 375K parameters
  - Standard training: 56M parameters

- Forward testing parameters obtained in closed-form from the inverse regression parameters

- Estimation ($\hat{x}$) is efficient (few computations)
Datasets

Training Sets

Prima Head Pose

Biwi-kinect Head Pose

Testing Sets

Leave one out test (ground truth available)

Out of datasets test

Prima

Biwi
## Results with face annotations

<table>
<thead>
<tr>
<th>Method</th>
<th>yaw</th>
<th>pitch</th>
<th>roll</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fanelli <em>et al.</em> (use 3D info)</td>
<td>3.5 ± 5.8</td>
<td>3.8 ± 6.5</td>
<td>5.4 ± 6.0</td>
</tr>
<tr>
<td>Wang <em>et al.</em> (use 2D-3D info)</td>
<td>8.8 ± 14.3</td>
<td>8.5 ± 11.1</td>
<td>7.4 ± 10.8</td>
</tr>
<tr>
<td>Our method (use 2D info)</td>
<td>4.9 ± 4.1</td>
<td>5.9 ± 4.8</td>
<td>4.7 ± 4.6</td>
</tr>
</tbody>
</table>

Mean Absolute Error (MAE) in degrees for head pose estimation
## Results with face detection

<table>
<thead>
<tr>
<th>Prima Head Pose</th>
<th>yaw</th>
<th>pitch</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gourier <em>et al.</em></td>
<td>10.3</td>
<td>15.9</td>
</tr>
<tr>
<td>Ricci &amp; Odobez</td>
<td>9.1</td>
<td>10.5</td>
</tr>
<tr>
<td>Our method</td>
<td>8.7</td>
<td>8.85</td>
</tr>
</tbody>
</table>

Mean Absolute Error (MAE) in degrees for head pose estimation
Examples of head pose estimation

- Face Detector
- Refined face bounding box
- Final Head Pose Estimation
Examples of head pose estimation
Examples of head pose estimation
Conclusion

Probabilistic piece-wise linear regression for high dimensional data:

- Efficient and accurate solution based on inverse training
- Head pose estimation in the presence of face localization errors

Next step:

- Apply to other problems: articulated motion capture, human-body pose, etc
- Extend the model to track head pose
Thank you for your attention