Head Pose Estimation Via Probabilistic High-Dimensional Regression

Vincent DROUARD [1] Silèye BA [1] Georgios EVANGELIDIS [1]
Antoine DELEFORGE [2] Radu HORAUD [1]

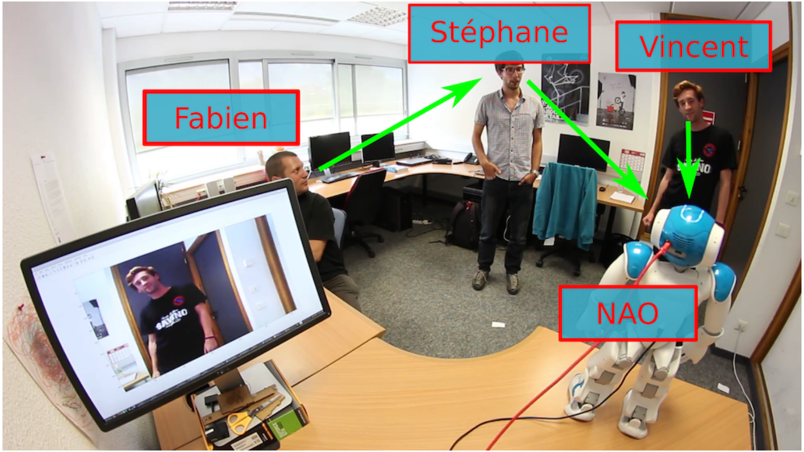
[1]Team Perception - Inria Grenoble Rhône-Alpes, France

[2]Friedrich-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



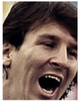
Visual cue in Human-Robot and Human-Human Interaction

Problem definition

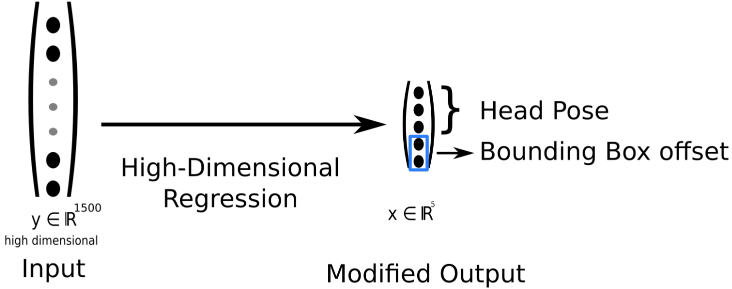
Data Input Output

Problem formulation

Annotated Bounding Box → Ideal Scenario

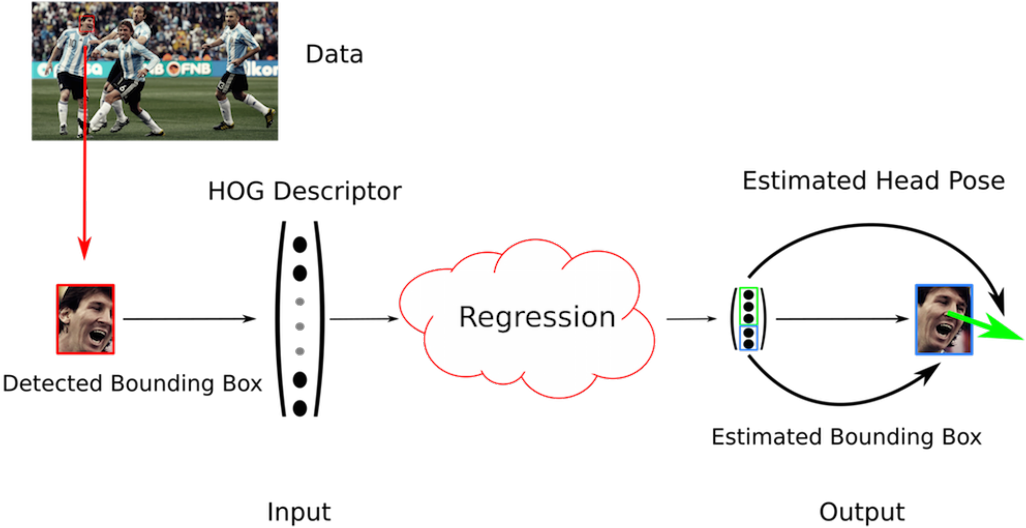


Detected Bounding Box → Reality



Affected by offset in detected bounding box

Method Pipeline



High-Dimensional Regression

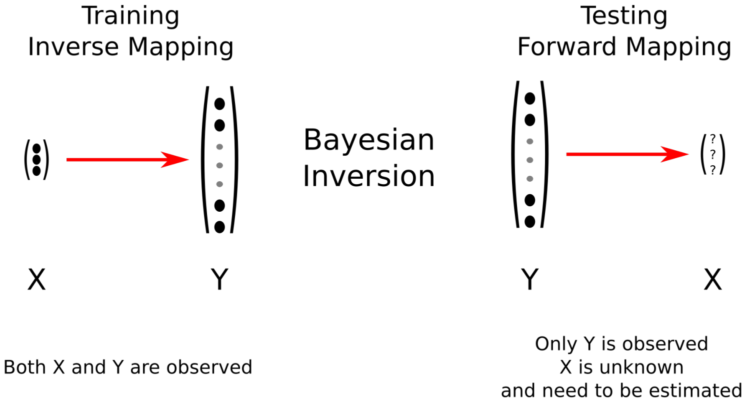
Problems:

- A lot of parameters to estimate
- $\boldsymbol{y} \to \boldsymbol{x}$ might be non linear
- $\boldsymbol{y}$ and $\boldsymbol{x}$ are obtained by measurement $\to$ might be noisy

Standard solution

- Step 1: dimension reduction, $\boldsymbol{y} \to \boldsymbol{x}'$
- Step 2: regression, $\boldsymbol{x}' \to \boldsymbol{x}$
- 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)

$$\boldsymbol{y} = \sum_{k=1}^{K} \mathbb{I}_{\{Z=k\}} \left(\mathbf{A}_k \boldsymbol{x} + \boldsymbol{b}_k + \boldsymbol{e}_k \right)$$

- A_k, b_k: parameters of the k^{th} affine transformation
- $\boldsymbol{e}_k$: zero-mean noise, $\boldsymbol{e}_k \sim \mathcal{N}\left(0, \boldsymbol{\Sigma}_k\right)$
- $\mathbb{I}$: indicator function
- Z: discrete latent variable selecting the affine transformation

Training: Inverse Regression (II)

- Probabilistic model

$$P\left(\boldsymbol{y}|\boldsymbol{x}, Z=k\right) = \mathcal{N}\left(\boldsymbol{y}; \mathbf{A}_k\boldsymbol{x} + \boldsymbol{b}_k, \boldsymbol{\Sigma}_k\right)$$
$$P\left(\boldsymbol{x}|Z=k\right) = \mathcal{N}\left(\boldsymbol{x}; \boldsymbol{c}_k, \boldsymbol{\Gamma}_k\right)$$
$$P\left(Z=k\right) = \pi_k$$

Training: Inverse Regression (III)

- Inverse Regression

$$P\left(\boldsymbol{y}|\boldsymbol{x};\theta\right) = \sum_{k=1}^{K} \underbrace{\frac{\pi_k \mathcal{N}\left(\boldsymbol{x};\boldsymbol{c}_k,\boldsymbol{\Gamma}_k\right)}{\sum_{j=1}^{K}\pi_j \mathcal{N}\left(\boldsymbol{x};\boldsymbol{c}_j,\boldsymbol{\Gamma}_j\right)}}_{\rho_k \rightarrow \textbf{Proportion}} \mathcal{N}\left(\boldsymbol{y};\mathbf{A}_k\boldsymbol{x}+\boldsymbol{b}_k,\boldsymbol{\Sigma}_k\right)$$

- $\theta = \{\mathbf{A}_k, \boldsymbol{b}_k, \boldsymbol{\Sigma}_k, \boldsymbol{c}_k, \boldsymbol{\Gamma}_k, \pi_k\}_{k=1}^{K}$
- Estimated using EM algorithm

Forward Testing

Bayesian inversion of the model

$$P(\boldsymbol{x}|\boldsymbol{y};\theta^*) = \sum_{k=1}^{K} \underbrace{\frac{\pi_k^* \mathcal{N}(\boldsymbol{y}; \boldsymbol{c}_k^*, \boldsymbol{\Gamma}_k^*)}{\sum_{j=1}^{K} \pi_j^* \mathcal{N}(\boldsymbol{y}; \boldsymbol{c}_j^*, \boldsymbol{\Gamma}_j^*)}}_{\rho_k^* \rightarrow \textbf{Proportion}} \mathcal{N}(\mathbf{A}_k^* \boldsymbol{y} + \boldsymbol{b}_k^*, \boldsymbol{\Sigma}_k^*)$$

- $\theta^* = \{\mathbf{A}_k^*, \boldsymbol{b}_k^*, \boldsymbol{\Sigma}_k^*, \boldsymbol{c}_k^*, \boldsymbol{\Gamma}_k^*, \pi_k^*\}_{k=1}^{K}$ obtained analytically using θ
- $\hat{x} = \mathbb{E}(\boldsymbol{x}|\boldsymbol{y};\theta^*) = \underbrace{\sum_{k=1}^{K} \rho_k^* (\mathbf{A}_k^* \boldsymbol{y} + \boldsymbol{b}_k^*)}_{\textbf{Fast evaluation}}$

Summary of the model

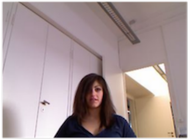
- Closed-form solution for estimating inverse regression parameters
  - high dimension = 1500 and Low dimension = 3 (or 5) (K=50)
  - Proposed inverse training: 375K parameters
  - Standard training: 56M parameters
- Forward testing parameters obtained in closed-form from the inverse regression parameters
- Estimation ($\hat{\boldsymbol{x}}$) is efficient (few computations)

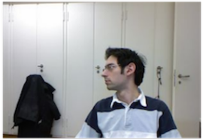
Datasets

Training Sets

Prima Head Pose

Biwi-kinect Head Pose





Testing Sets

Leave one out test (ground truth available)

Prima

Biwi

Out of datasets test

Results with face annotations

Biwi-Kinect Head Pose			
Method	yaw	pitch	roll
Fanelli *et al.* (use 3D information)	$\mathbf{3.5} \pm 5.8$	$\mathbf{3.8} \pm 6.5$	5.4 ± 6.0
Wang *et al.* (use 2D-3D information)	8.8 ± 14.3	8.5 ± 11.1	7.4 ± 10.8
Our method (use 2D information)	$4.9 \pm \mathbf{4.1}$	$5.9 \pm \mathbf{4.8}$	$\mathbf{4.7} \pm \mathbf{4.6}$

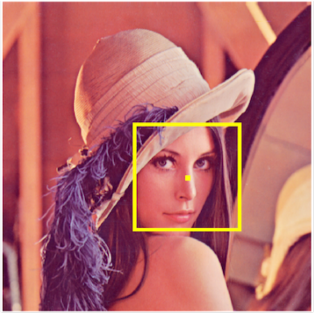
Mean Absolute Error (MAE) in degrees for head pose estimation

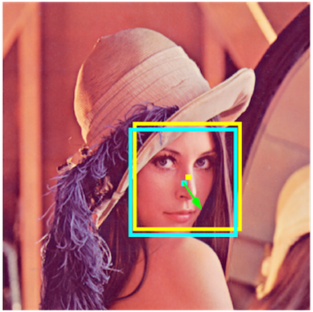
Results with face detection

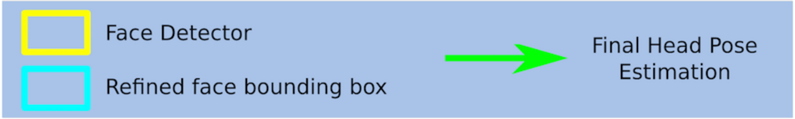
Prima Head Pose		
Method	yaw	pitch
Gourier *et al.*	10.3	15.9
Ricci & Odobez	9.1	10.5
Our method	**8.7**	**8.85**

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

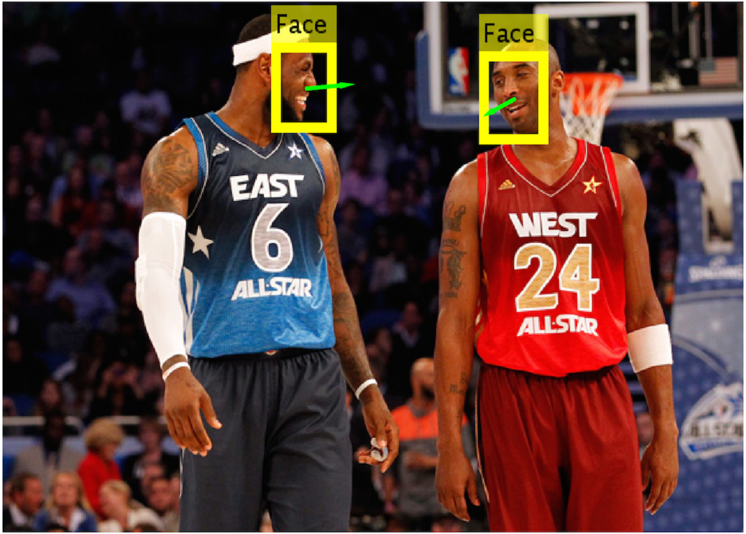
Examples of head pose estimation



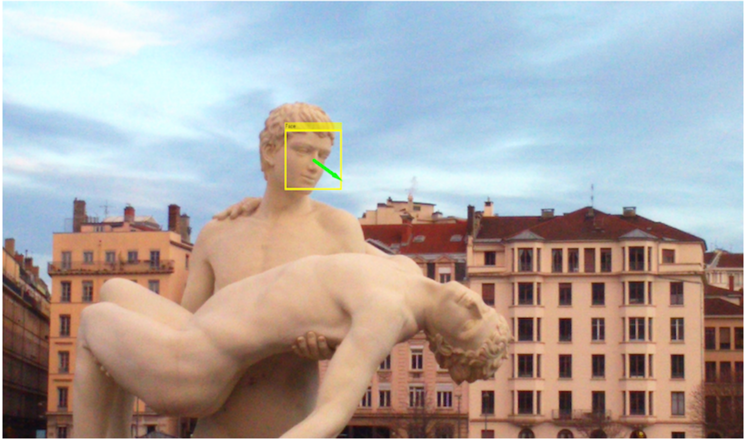




Examples of head pose estimation



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