

On Partial Least Squares in Head Pose Estimation: How to simultaneously deal with misalignment

Murad Al Haj¹, Jordi Gonzàlez¹, Larry S. Davis²

¹Centre de Visió per Computador, Universitat Autònoma de Barcelona

²Institute for Advanced Computer Studies, University of Maryland

June 04, 2013

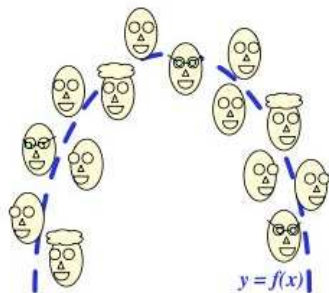
Introduction



Image Source: E. Murphy-Chutorian and M.M. Trivedi. "Head pose estimation in computer vision: A survey".
2009.

Introduction

Best performance in head pose estimation is obtained by non-linear regression methods.



Partial Least Squares, a regression technique, has been gaining much interest in computer vision lately.

Introduction

- In prior work, there is no study of the effect misalignment
- Propose head pose estimation method based on partial least squares (PLS) regression
- ... while solving the alignment problem simultaneously.

Linear and Kernel PLS

Consider a matrix of independent variables \mathbf{X} and a matrix of dependent variables \mathbf{Y} , obtained as a response to \mathbf{X} . PLS decomposes the matrices as follows:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T + \mathbf{E}$$

$$\mathbf{Y} = \mathbf{U}\mathbf{Q}^T + \mathbf{F}$$

Solving for \mathbf{T} and \mathbf{U} using the NIPALS algorithm, the regression coefficients can be expressed as:

$$\mathbf{B} = \mathbf{X}^T \mathbf{U} (\mathbf{T}^T \mathbf{X} \mathbf{X}^T \mathbf{U})^{-1} \mathbf{T}^T \mathbf{Y}$$

The kernel PLS applies the same decomposition but after a nonlinear transformation of the input vectors.

Experimental Setup

We tested the linear PLS and kPLS on two databases: Pointing'04 and CMU Multi-PIE.

\mathbf{X} was composed of the HOG features of each face.

\mathbf{Y} is composed of the corresponding pose: two dimensional (pitch and yaw) for Pointing'04 while one dimensional (yaw) for CMU Multi-PIE.

Pointing'04 Results

Method	Yaw Err	Pitch Err	Accuracy (Yaw,Pitch)	Notes
Ours (kernel PLS)	6.56^o	6.61^o	(67.36%, 80.36%)	-
Stiefelhagen	9.5 ^o	9.7 ^o	(52.0%, 66.3%)	1
Ours (linear PLS)	11.29^o	10.52^o	(45.57%, 58.70%)	-
Human Performance	11.8 ^o	9.4 ^o	(40.7%, 59.0%)	2
Gourier (Associative Memories)	10.1 ^o	15.9 ^o	(50.0%, 43.9%)	3
Tu (High-order SVD)	12.9 ^o	17.97 ^o	(49.25%, 54.84%)	4
Tu (PCA)	14.11 ^o	14.98 ^o	(55.20%, 57.99%)	4
Tu (LEA)	15.88 ^o	17.44 ^o	(45.16%, 50.61%)	4
Voit	12.3 ^o	12.77 ^o	—	-
Li (PCA)	26.9 ^o	35.1 ^o	—	5
Li (LDA)	25.8 ^o	26.9 ^o	—	5
Li (LPP)	24.7 ^o	22.6 ^o	—	5
Li (Local-PCA)	24.5 ^o	37.6 ^o	—	5
Li (Local-LPP)	29.2 ^o	40.2 ^o	—	5
Li (Local-LDA)	19.1 ^o	30.7 ^o	—	5

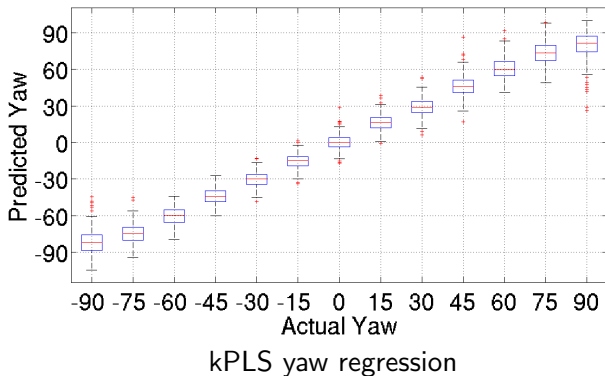
Notes:

- 1) Used 80% of Pointing'04 images for training, 10% for cross-evaluation, and 10% for testing.
- 2) Human performance with training.
- 3) Best results over different reported methods.
- 4) Better results have been obtained with manual localization.
- 5) Results for 32-dim embedding.

Ref: M. Al Haj, J. Gonzalez, and L.S. Davis. "On partial least squares in head pose estimation: How to simultaneously deal with misalignment". CVPR 2012.

Table Source: E. Murphy-Chutorian and M.M. Trivedi. "Head pose estimation in computer vision: A survey". April 2009.

Box-and-Whisker Pointing'04 Results



Multi-PIE

2700 face images from the CMU Multi-PIE database were manually annotated. These images belong to 144 subjects, under frontal illumination and varying expressions.

	kPLS	linear PLS	PCR
Mean Absolute Error (MAE)	5.31°	9.11°	11.03°
Accuracy	79.48%	57.22%	48.33%

The Alignment Problem

Misalignment is a problem for any regression/classification algorithm.

The Alignment Problem

Misalignment is a problem for any regression/classification algorithm.

Pose results are reported without studying this effect.

The Alignment Problem

Misalignment is a problem for any regression/classification algorithm.

Pose results are reported without studying this effect.

The detection output is not necessarily aligned with trained models.

Misalignment Effect



0% shift
pitch 0° and yaw -30°



0% shift
pitch ? and yaw ?

Error in Pointing'04 linear PLS: 11.29°
Error in Multi-PIE linear PLS: 9.11°
Error in Pointing'04 kernel PLS: 6.56°
Error in Multi-PIE kernel PLS: 5.31°

Misalignment Effect



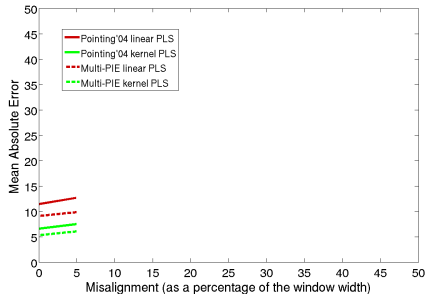
0% shift

pitch 0° and yaw -30°



5% shift

pitch ? and yaw ?



Misalignment Effect



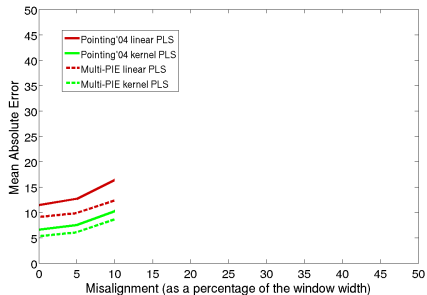
0% shift

pitch 0° and yaw -30°



10% shift

pitch ? and yaw ?



Misalignment Effect



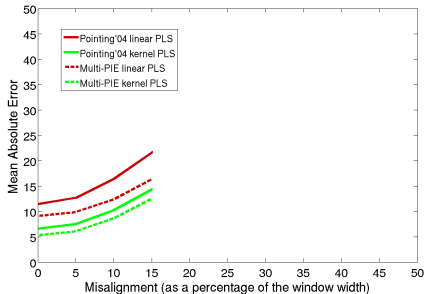
0% shift

pitch 0° and yaw -30°



15% shift

pitch ? and yaw ?



Misalignment Effect



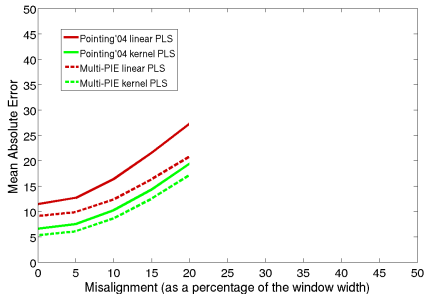
0% shift

pitch 0° and yaw -30°



20% shift

pitch ? and yaw ?



Misalignment Effect



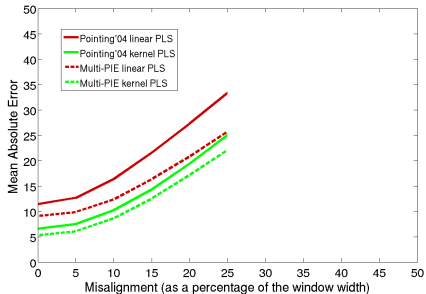
0% shift

pitch 0° and yaw -30°



25% shift

pitch ? and yaw ?



Misalignment Effect



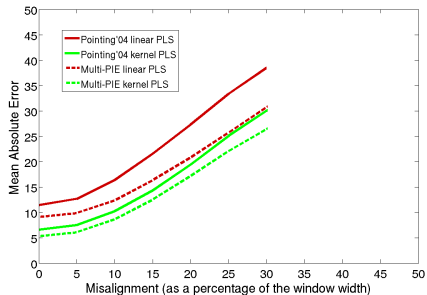
0% shift

pitch 0° and yaw -30°



30% shift

pitch ? and yaw ?



Misalignment Effect



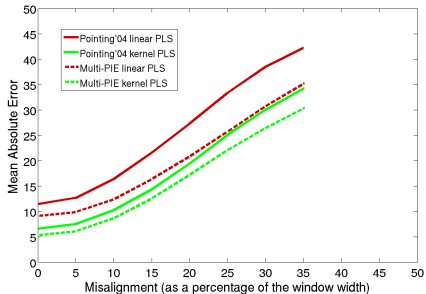
0% shift

pitch 0° and yaw -30°



35% shift

pitch ? and yaw ?



Misalignment Effect



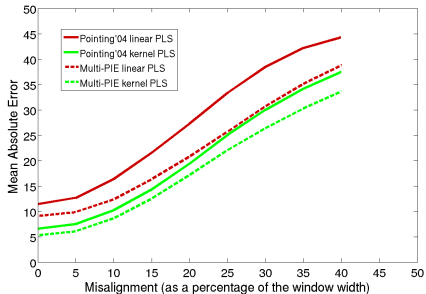
0% shift

pitch 0° and yaw -30°



40% shift

pitch ? and yaw ?



Misalignment Effect



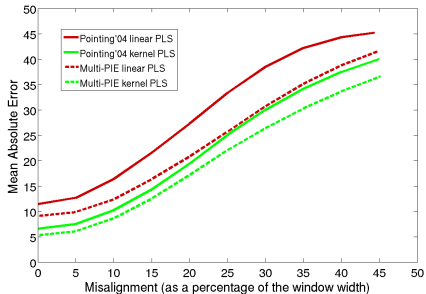
0% shift

pitch 0° and yaw -30°



45% shift

pitch ? and yaw ?



Misalignment Effect



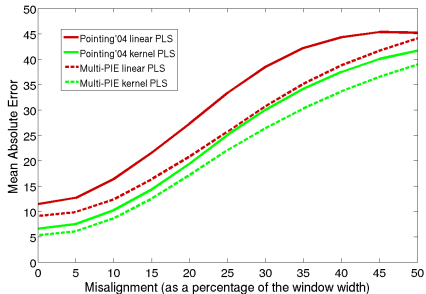
0% shift

pitch 0° and yaw -30°



50% shift

pitch ? and yaw ?



The Alignment Problem

Proposal:

Consider not only the detected face but also a bag of neighboring windows.

The Alignment Problem

Proposal:

Consider not only the detected face but also a bag of neighboring windows.

Given the latent sub-spaces, the instance with the minimum residual is the one with the best response.

Residual estimation

We derived this error as:

linear: $\mathbf{e} = \mathbf{x} - \mathbf{x}\mathbf{X}^T\mathbf{T}(\mathbf{T}^T\mathbf{X}\mathbf{X}^T\mathbf{T})^{-1}\mathbf{T}^T\mathbf{X}.$

kernel: $\mathbf{e} = K(\mathbf{x}, \mathbf{x}) - K(\mathbf{x}, \mathbf{X})\mathbf{T}\mathbf{t}^T - \mathbf{t}\mathbf{T}^TK^T(\mathbf{x}, \mathbf{X}) + \mathbf{t}\mathbf{T}^TK\mathbf{T}\mathbf{t}^T.$

To test the accuracy of minimum residual, compare accuracy of well-aligned samples vs. the selected samples in misaligned bags.

Misalignment Bags



0%

Misalignment Bags



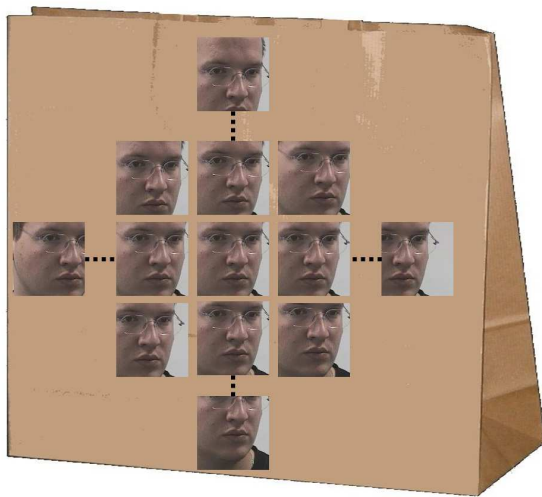
5%

Misalignment Bags



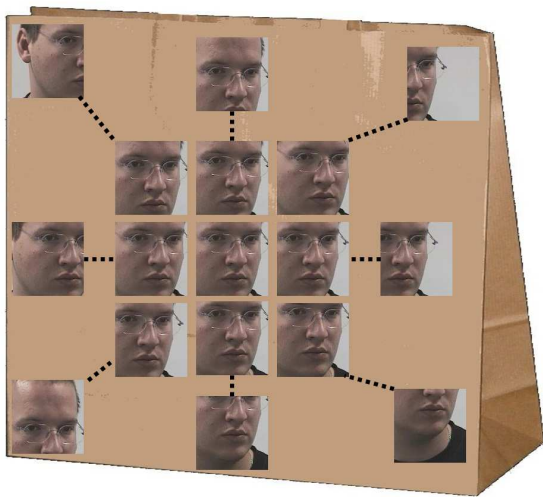
10%

Misalignment Bags



25%

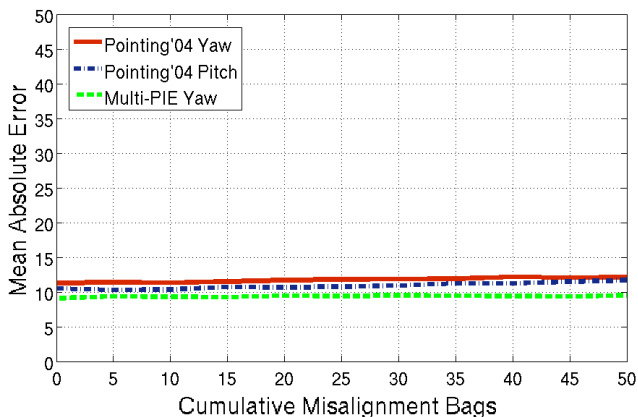
Misalignment Bags



50%

Misalignment Bags

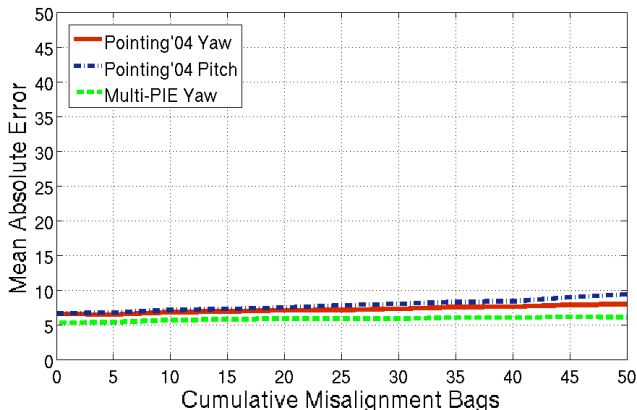
The MAE of applying the regression on the minimum residual sample of each bag is shown below:



linear case

Misalignment Bags

The MAE of applying the regression on the minimum residual sample of each bag is shown below:



kernel case

Our method vs. MIL

- Multiple Instance Learning (MIL) is used to accommodate for misalignment
- On average, our kPLS method outperforms Multi-Instance Multi-Label SVM (MIMLSVM)
- ...despite not having any misaligned sample in the training dataset and being 100x faster.

	Ours (kPLS)	MIMLSVM
MAE Pointing'04 Yaw	7.94°	10.72°
MAE Pointing'04 Pitch	9.35°	12.32°
MAE Multi-PIE Yaw	6.06°	5.40°

Source: Z.-H. Zhou and M.-L. Zhang. "Multi-instance multilabel learning with application to scene classification".

NIPS 2007



Nicolas Gourier, Jérôme Maisonnasse, Daniela Hall, and James L. Crowley.

Head pose estimation on low resolution images.
In In CLEAR 2006, South, 2006.



M.A. Haj, J. Gonzalez, and L.S. Davis.

On partial least squares in head pose estimation: How to simultaneously deal with misalignment.

In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, pages 2602–2609, 2012.



Zhu Li, Yun Fu, Junsong Yuan, T.S. Huang, and Ying Wu.

Query driven localized linear discriminant models for head pose estimation.

In Multimedia and Expo, 2007 IEEE International Conference on, pages 1810–1813, 2007.



Rainer Stiefelhagen.

Estimating head pose with neural networks- results on the pointing04 icpr workshop evaluation data, 2004.



Jilin Tu, Yun Fu, Yuxiao Hu, and Thomas Huang.

Evaluation of head pose estimation for studio data.

In Proceedings of the 1st international evaluation conference on Classification of events, activities and relationships, CLEAR'06, pages 281–290, Berlin, Heidelberg, 2007. Springer-Verlag.



Michael Voit, Kai Nickel, and Rainer Stiefelhagen.

R.: Neural network-based head pose estimation and multi-view fusion.

In Proc. CLEAR Workshop, LNCS, pages 299–304, 2006.