Parametric Face Alignment: Generative and Discriminative Approaches

Pedro Alexandre Dias Martins
http://www.isr.uc.pt/~pedromartins

PhD Thesis
Faculty of Sciences and Technology
University of Coimbra
Portugal

· U C ·

Universidade de Coimbra

Acknowledgements

- **Advisor:** Dr. Jorge Batista

- Institute of Systems and Robotics (ISR) - University of Coimbra
  http://www.isr.uc.pt/

- Department of Electrical and Computer Engineering (DEEC) from Faculty of Sciences and Technology of the University of Coimbra (FCTUC)
  http://www.uc.pt/fctuc/deec/

- Research supported by the Portuguese Science Foundation (*Fundação para a Ciência e Tecnologia* - FCT) under the PhD grant SFRH/BD/45178/2008
  http://www.fct.pt/

Introduction

- **Goal:** Non-Rigid Face Registration
- Model based approaches - Parametric Models of Shape and Appearance

Normalized Template

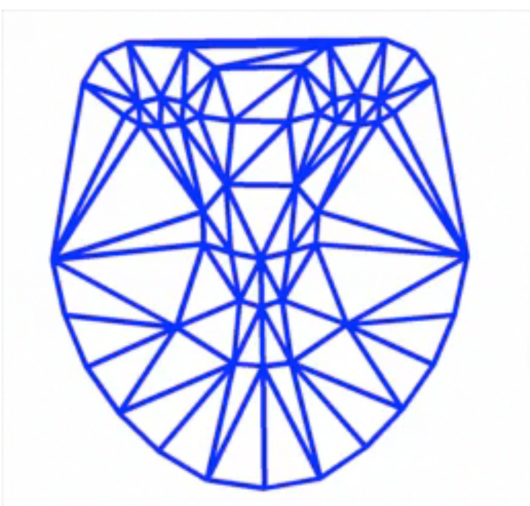
Required for Several Tasks

- Face Recognition
- Facial Expr. Recognition
- Tracking
- Head Pose Estimation
- Gaze Estimation
- Image Compression
- ...

Model's Parameters: $\mathbf{p}, \mathbf{q}, \boldsymbol{\lambda}$

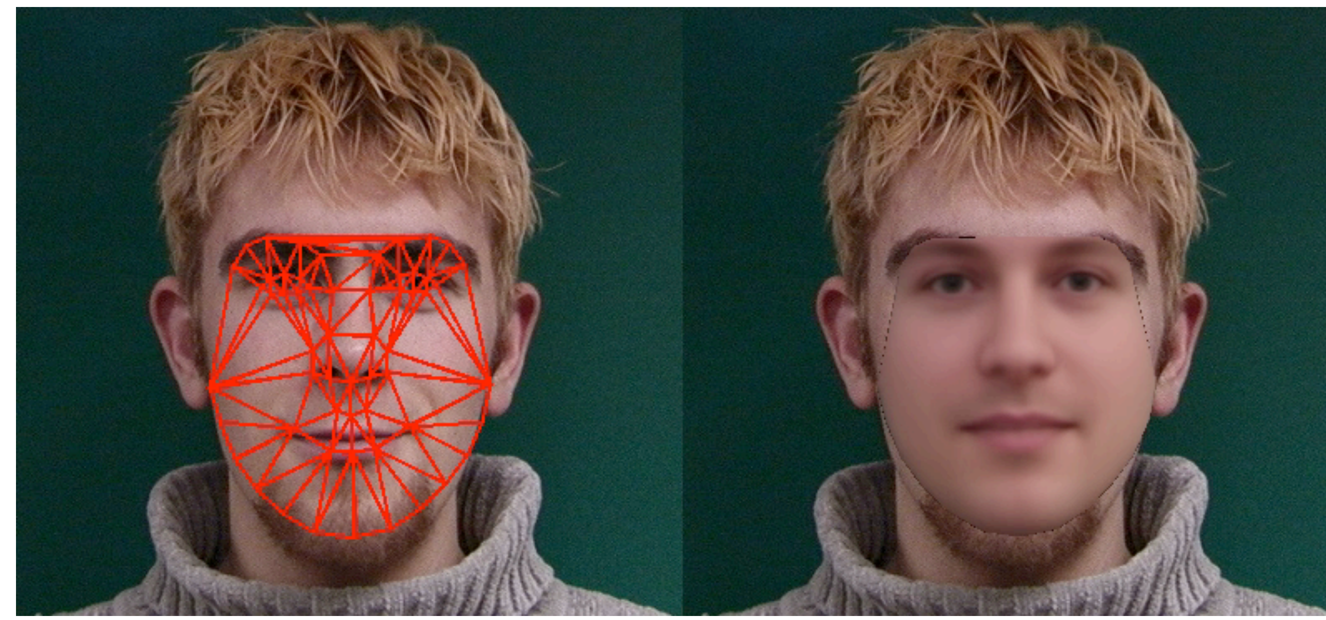
Parametric Image Alignment

Shape Model



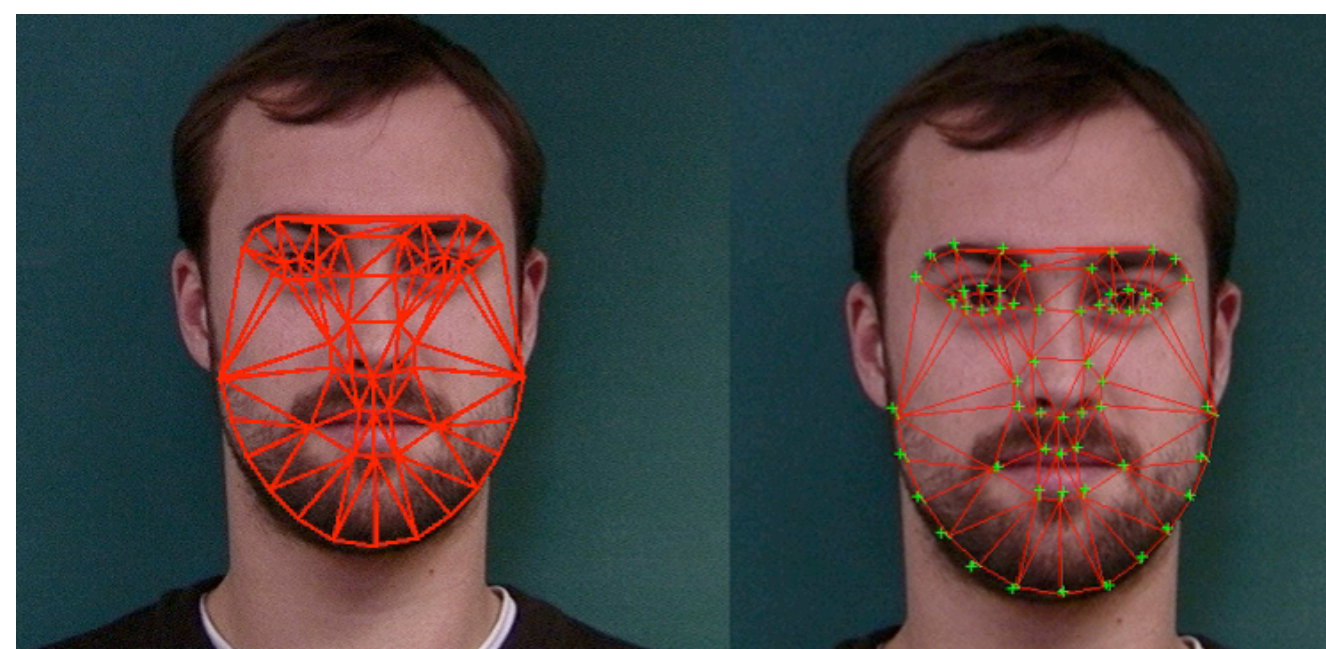
Point Distribution Model (PDM)

- **Generative / Holistic Appearance Model**



e.g. AAM, 3DMM, 2D+3D AAM

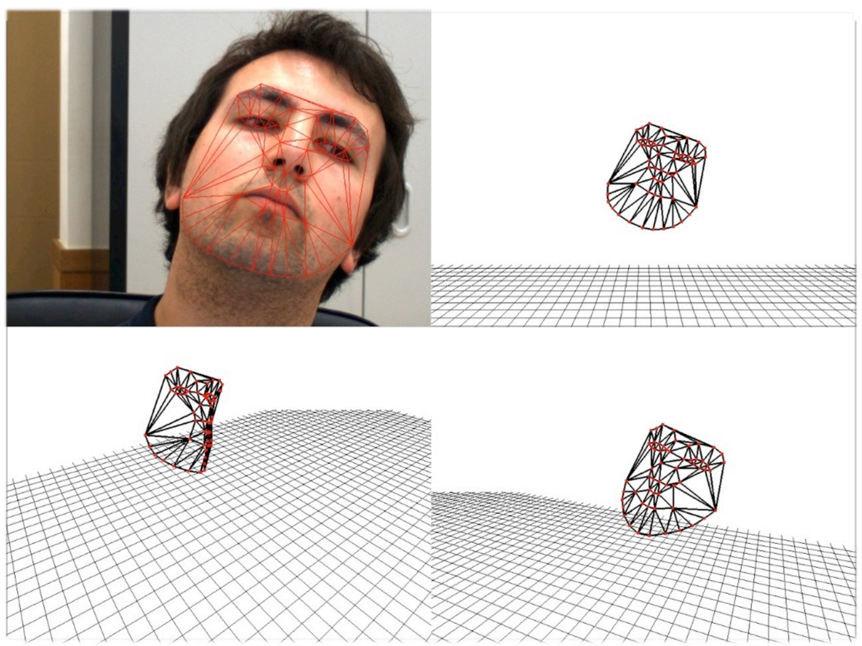
- **Discriminative / Patch Based Appearance Model**



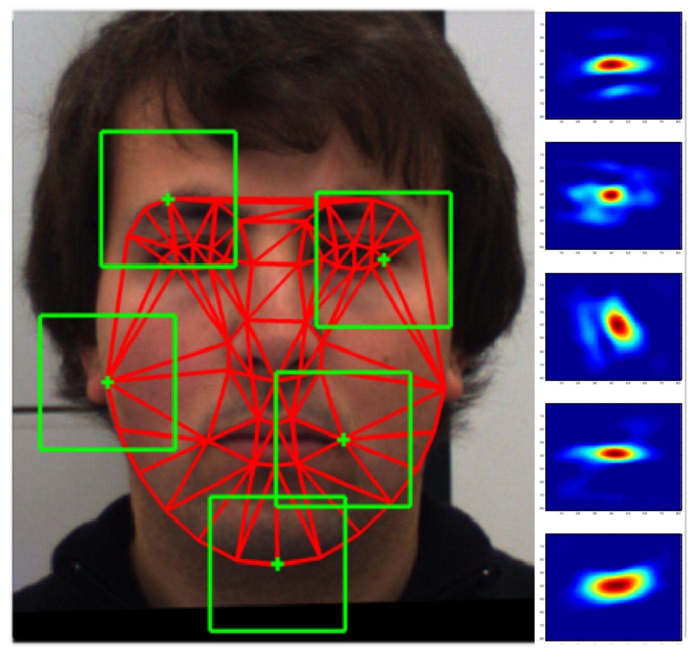
e.g. ASM, CLM, CQF, SCMS

Overview

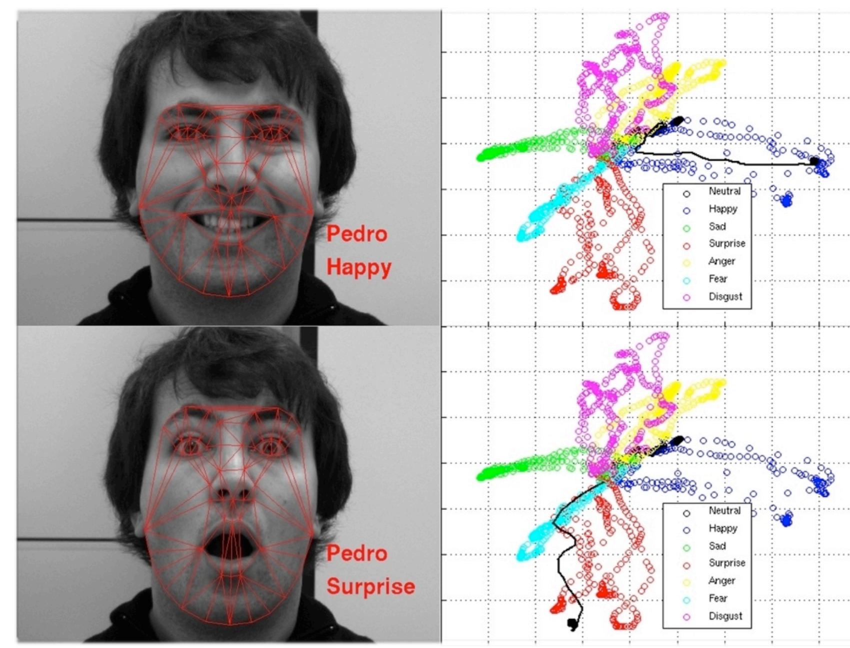
(1) Generative 2.5D AAM



(2) Discriminative ASM



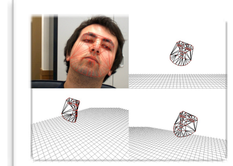
(3) Identity / Facial Expression



- Extension of the original 2D AAM that to deals with a <u>full perspective projection model</u>.

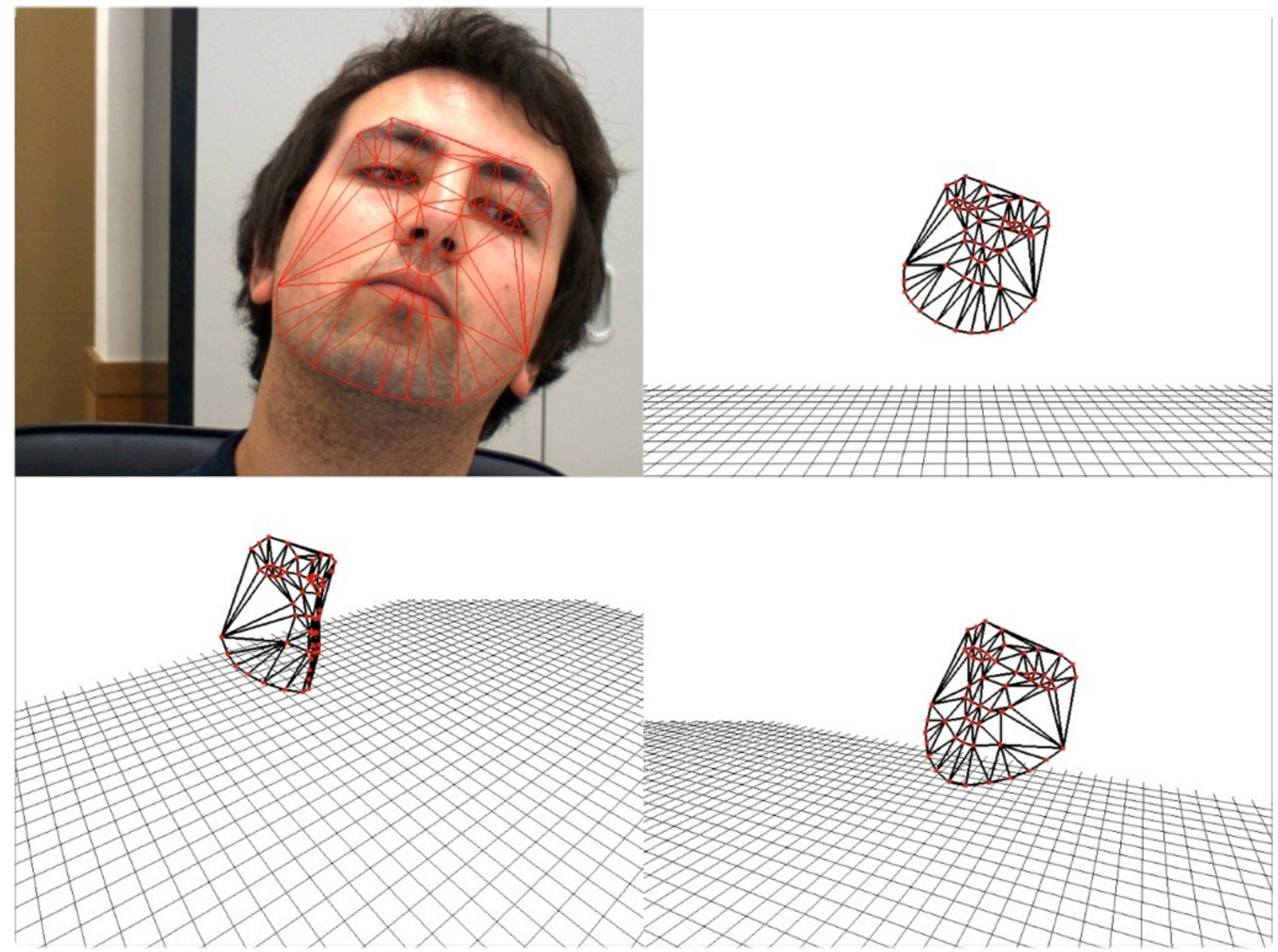
- New Bayesian global optimization strategy that infers the overall alignment using a <u>second order</u> estimate of the PDM parameters.

- Identity and facial expression recognition using facial geometry.

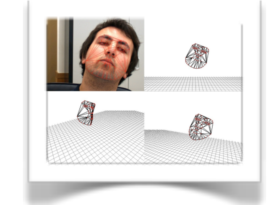


(1) Generative 2.5D Active Appearance Models

- The 2.5D Active Appearance Models (AAM) combines a 3D PDM and a 2D appearance model.
- Extension of the original 2D AAM that deals with a full perspective projection model.
- Model fitting algorithms:
  - Simultaneous Forwards Additive (SFA).
  - Normalization Forwards Additive (NFA).
  - Efficient Approximations (ESFA, ENFA).
  - Robust to partial and self occlusions.
- Larger convergency radius.
- 3D shape from single images.
- (-) Slower than 2D methods.

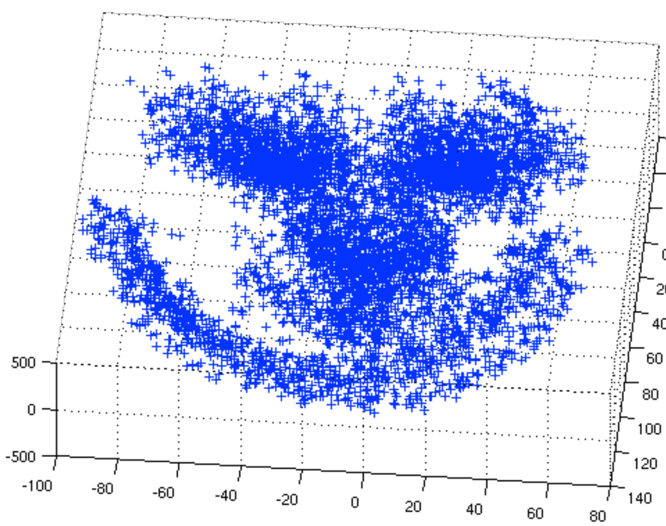


- **Previous Work:**
  - Active Appearance Models (AAM) - ECCV 1998
  - Active Appearance Models Revisited - IJCV 2004
  - Generic vs Person Specific AAMs - BMVC 2004
  - Real Time Combined 2D+3D AAMs - CVPR 2004



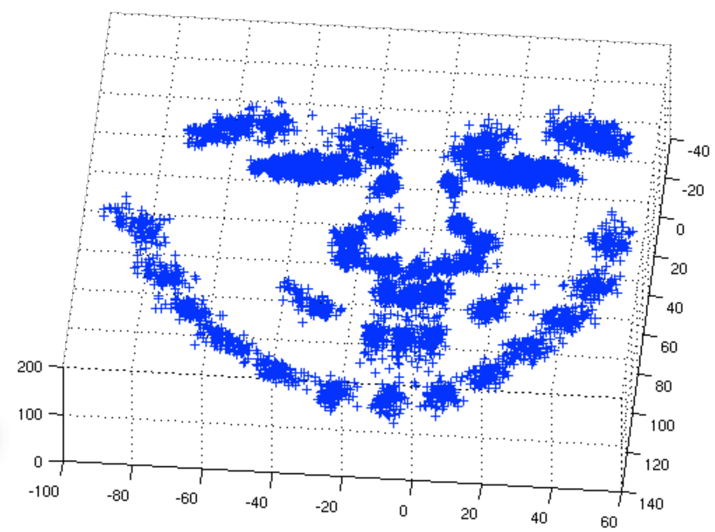
Parametric Models of Shape and Appearance

Raw Data

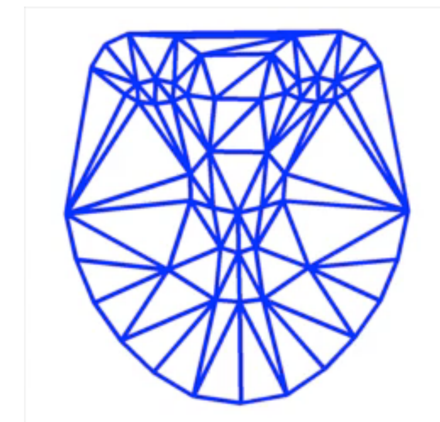


Procrustes Analysis

‘Aligned’ Data

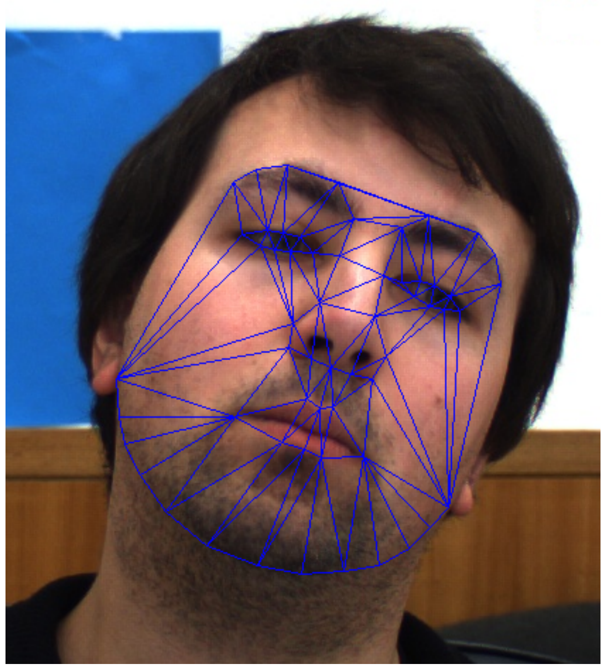


Shape Model



$$s = s_0 + \sum_{i=1}^{n} p_i \phi_i$$

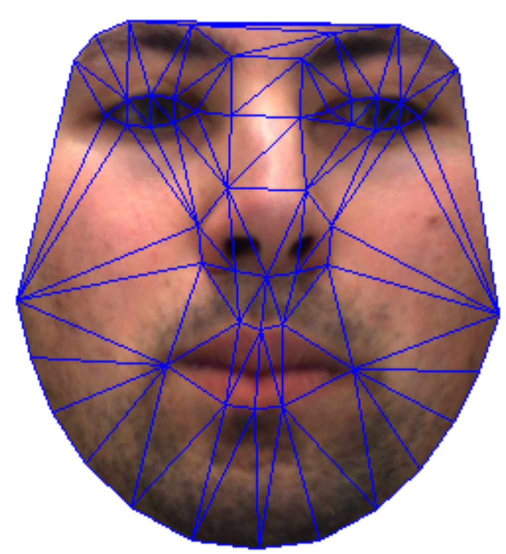
Shape Parameters



$\mathbf{I}(\mathbf{x_p})$

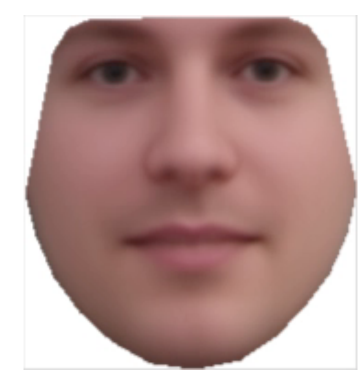
Piecewise Affine Warp

$\mathbf{W}(\mathbf{x_p}, \mathbf{p}, \mathbf{q})$



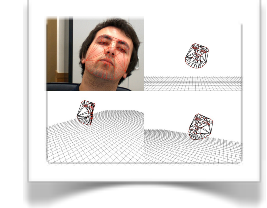
$\mathbf{I}(\mathbf{W}(\mathbf{x_p}, \mathbf{p}, \mathbf{q}))$

Appearance Model

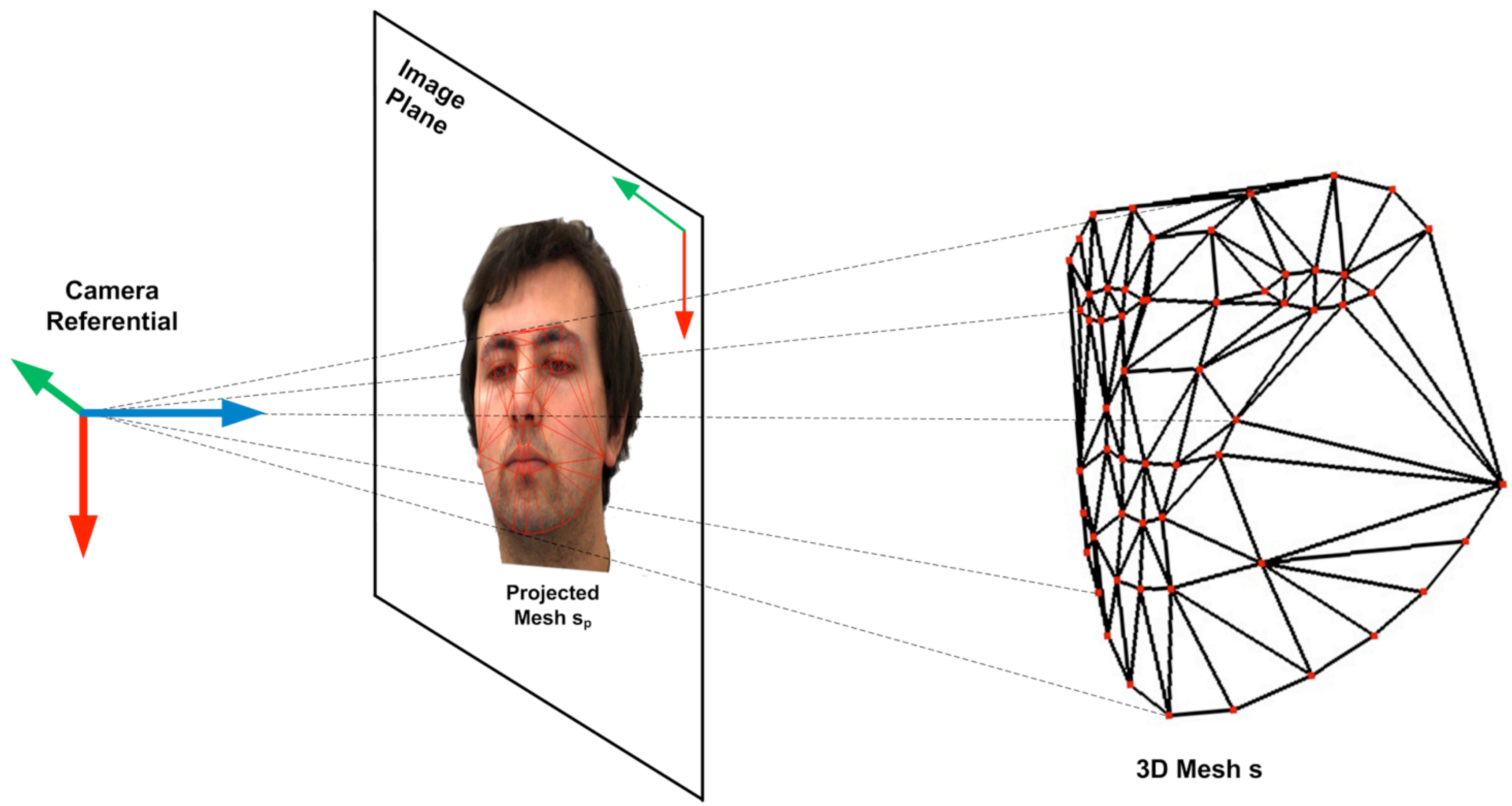


$$\mathbf{A}(\mathbf{x_p}) = \mathbf{A}_0(\mathbf{x_p}) + \sum_{i=1}^{m+2} \lambda_i \mathbf{A}_i(\mathbf{x_p}), \ \ \mathbf{x_p} \in s_{0\mathbf{p}}$$

Appearance Parameters



The Shape Model



Full Perspective Projection

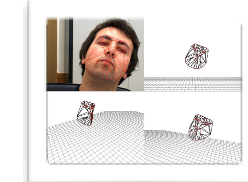
$$\begin{bmatrix} w(x_1 \cdots x_v) \\ w(y_1 \cdots y_v) \\ w \cdots w \end{bmatrix} = \underbrace{\begin{bmatrix} f_x & \alpha_s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}}_{\mathbf{K}} \left[\ \mathbf{R}_0 \mid \mathbf{t}_0\ \right] \begin{bmatrix} s^{x_1} \cdots s^{x_v} \\ s^{y_1} \cdots s^{y_v} \\ s^{z_1} \cdots s^{z_v} \\ 1 \cdots 1 \end{bmatrix} \Leftarrow$$

3D Point Distribution Model (PDM)

$$s = s_0 + \sum_{i=1}^{n} p_i \phi_i + \sum_{j=1}^{6} q_j \psi_j^{(t)} + \underbrace{\int_0^{t-1} \sum_{j=1}^{6} q_j \psi_j^{(t)} \partial t}_{s_\psi}$$

Pose Parameters

Previous pose updates



Model Fitting

$\arg\min_{\mathbf{p},\mathbf{q},\boldsymbol{\lambda}} \sum_{\mathbf{x_p}\in s_{0\mathbf{p}}} \Big[$  $-$  $\Big]^2$

$$\arg\min_{\mathbf{p},\mathbf{q},\boldsymbol{\lambda}} \sum_{\mathbf{x_p}\in s_{0\mathbf{p}}} \left[\mathbf{A}_0(\mathbf{x_p}) + \sum_{i=1}^{m+2} \lambda_i \mathbf{A}_i(\mathbf{x_p}) - \mathbf{I}(\mathbf{W}(\mathbf{x_p}, \mathbf{p}, \mathbf{q}))\right]^2$$

- Additive updates to the parameters

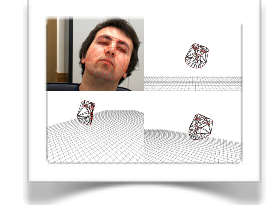
$$\sum_{\mathbf{x}\in s_{0\mathbf{p}}} [\mathbf{A}_0(\mathbf{x_p}) + \sum_{i=1}^{m+2} (\underline{\lambda_i + \Delta\lambda_i})\mathbf{A}_i(\mathbf{x_p}) - \mathbf{I}(\mathbf{W}(\mathbf{x_p}, \underline{\mathbf{p} + \Delta\mathbf{p}}, \underline{\mathbf{q} + \Delta\mathbf{q}}))]^2$$

- Expanding and holding first order terms

$$\sum_{\mathbf{x_p}\in s_{0\mathbf{p}}} \left[\mathbf{A}_0(\mathbf{x_p}) + \sum_{i=1}^{m+2} \lambda_i \mathbf{A}_i(\mathbf{x_p}) - \mathbf{I}(\mathbf{W}(\mathbf{x_p}, \mathbf{p}, \mathbf{q})) - \underline{\mathbf{SD}(\mathbf{x_p})_{\text{sfa}}\Delta\mathbf{r}}\right]^2 \qquad \mathbf{r} = \begin{bmatrix} \mathbf{p} \\ \mathbf{q} \\ \boldsymbol{\lambda} \end{bmatrix}$$

- Parameters updates

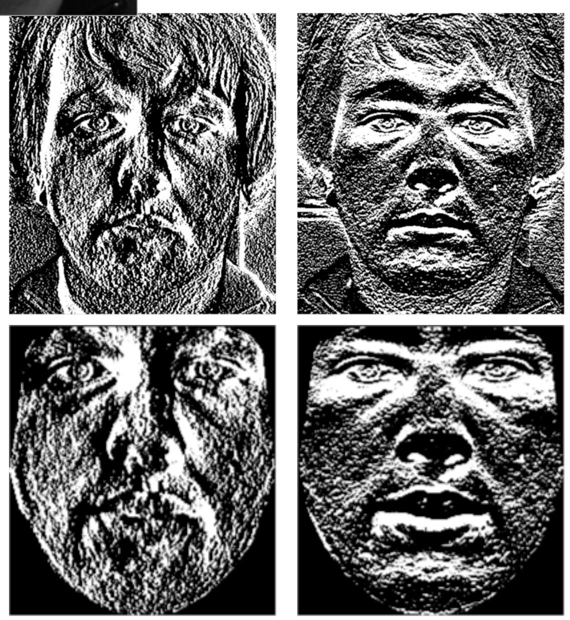
$$\Delta\mathbf{r} = \mathbf{H}_{\text{sfa}}^{-1} \sum_{\mathbf{x_p}\in s_{0\mathbf{p}}} \mathbf{SD}(\mathbf{x_p})_{\text{sfa}}^T \mathbf{E}(\mathbf{x_p})_{\text{sfa}} \qquad \mathbf{r} \leftarrow \mathbf{r} + \Delta\mathbf{r}$$



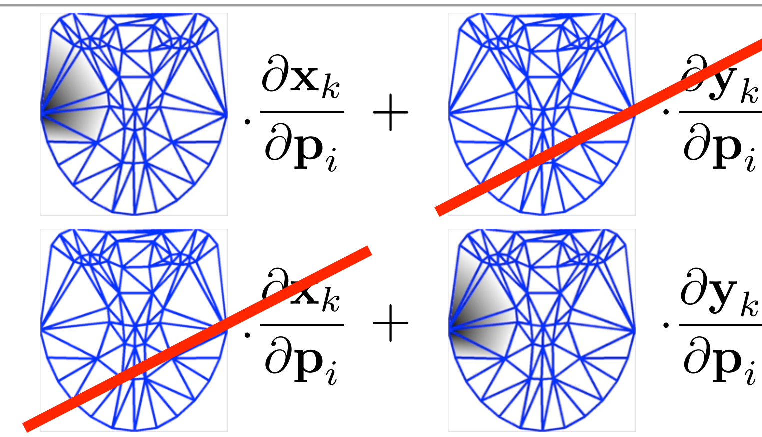
Simultaneous Forwards Additive (SFA)

$\mathbf{I}(\mathbf{x_p})$

$\mathbf{I}(\mathbf{W}(\mathbf{x_p}, \mathbf{p}, \mathbf{q}))$



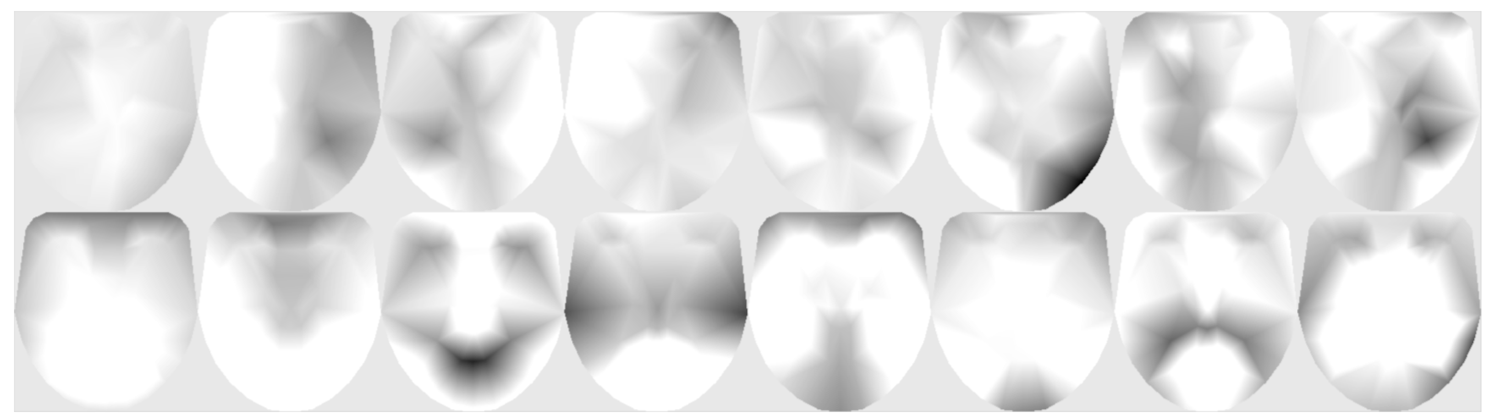
$\frac{\partial \mathbf{I}(\mathbf{W}(\mathbf{x_p}, \mathbf{p}, \mathbf{q}))}{\partial x}$ $\frac{\partial \mathbf{I}(\mathbf{W}(\mathbf{x_p}, \mathbf{p}, \mathbf{q}))}{\partial y}$



The Jacobian of the Warp

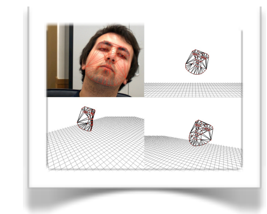
$$\frac{\partial \mathbf{W}(\mathbf{x_p}, \mathbf{p}, \mathbf{q})}{\partial \mathbf{p}} = \sum_{k=1}^{v} \left[\frac{\partial \mathbf{W}(\mathbf{x_p}, \mathbf{p}, \mathbf{q})}{\partial \mathbf{x}_k} \frac{\partial \mathbf{x}_k}{\partial \mathbf{p}} + \frac{\partial \mathbf{W}(\mathbf{x_p}, \mathbf{p}, \mathbf{q})}{\partial \mathbf{y}_k} \frac{\partial \mathbf{y}_k}{\partial \mathbf{p}} \right]$$

$\frac{\partial \mathbf{W}(\mathbf{x_p}, \mathbf{p}, \mathbf{q})}{\partial \mathbf{p}}$

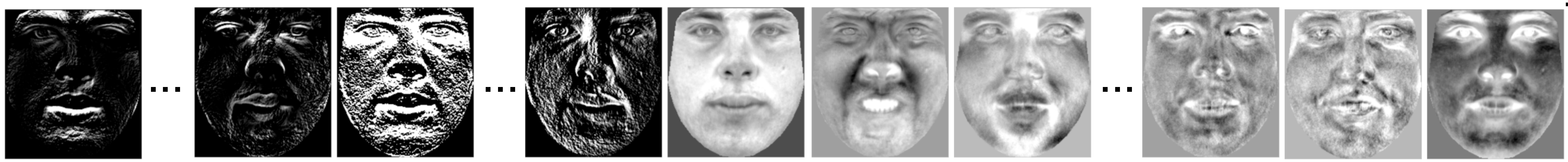


EigenFaces

$$\mathbf{SD}(\mathbf{x_p})_{\text{sfa}} = \left[\boxed{\nabla \mathbf{I}} \boxed{\frac{\partial \mathbf{W}}{\partial \mathbf{p}_1}} \ldots \; \nabla \mathbf{I} \frac{\partial \mathbf{W}}{\partial \mathbf{p}_n} \; \nabla \mathbf{I} \frac{\partial \mathbf{W}}{\partial \mathbf{q}_1} \; \ldots \nabla \mathbf{I} \frac{\partial \mathbf{W}}{\partial \mathbf{q}_6} \; \boxed{-\mathbf{A}_1(\mathbf{x_p}) \; \ldots \; -\mathbf{A}_{m+2}(\mathbf{x_p})} \right]$$



Normalization Forwards Additive (NFA)

$$\mathbf{SD}(\mathbf{x_p})_{\text{sfa}} = \Big[$$  $$\Big]$$

10 ~ 20 shape 'images' | 6 pose 'images' | 60 ~ 80 appearance 'images'

$$\mathbf{SD}(\mathbf{x_p})_{\text{nfa}} = \left[\nabla \mathbf{I} \frac{\partial \mathbf{W}}{\partial \mathbf{p}_1} \;\ldots\; \nabla \mathbf{I} \frac{\partial \mathbf{W}}{\partial \mathbf{p}_n} \; \nabla \mathbf{I} \frac{\partial \mathbf{W}}{\partial \mathbf{q}_1} \ldots \nabla \mathbf{I} \frac{\partial \mathbf{W}}{\partial \mathbf{q}_6} \right]$$

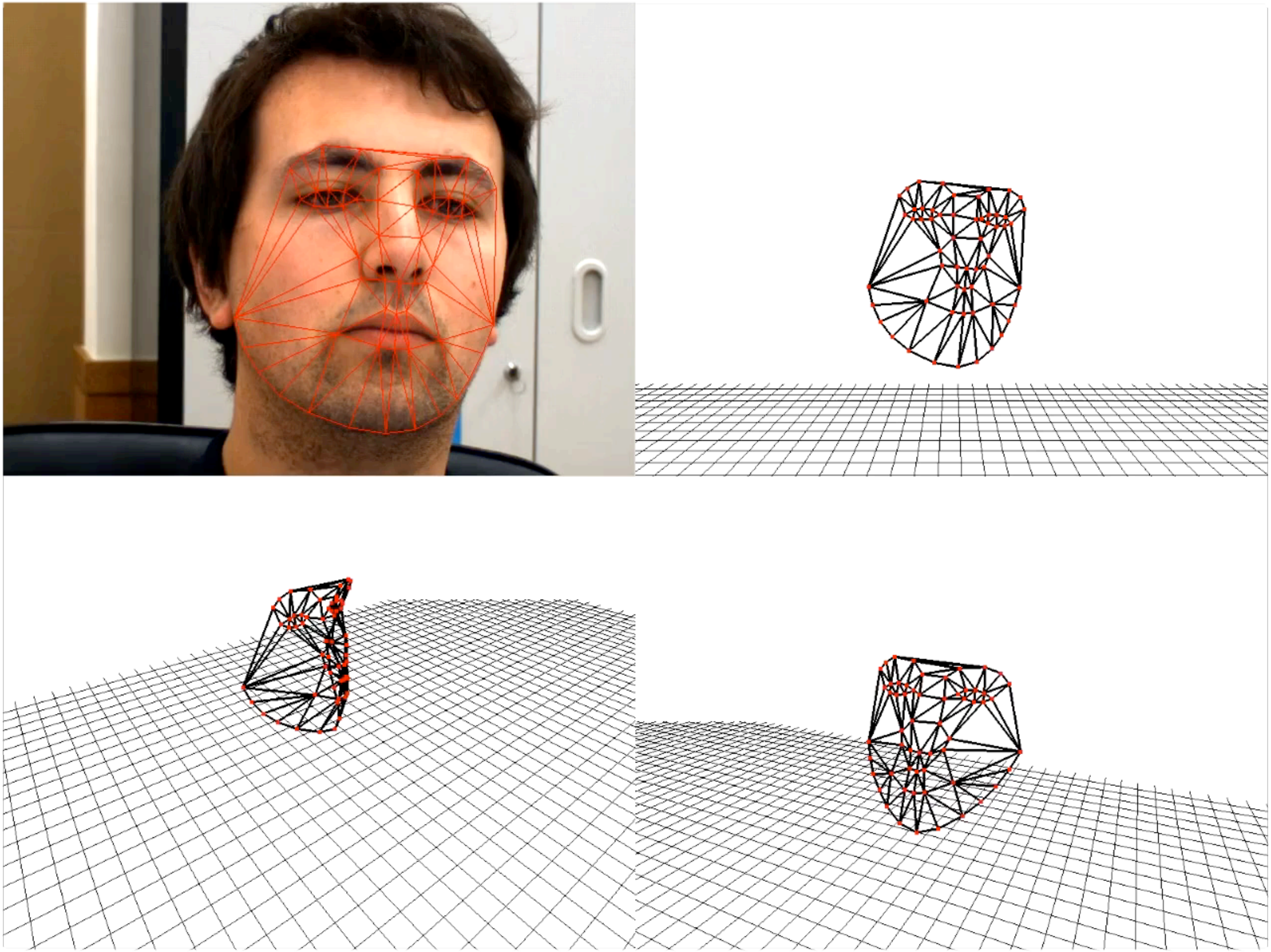
- Project Error Image into Appearance Basis

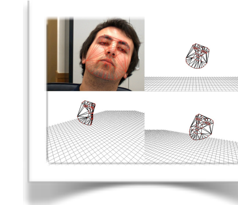
$$\lambda_i = \sum_{\mathbf{x_p} \in s_{0\mathbf{p}}} \mathbf{A}_i(\mathbf{x_p}) \underbrace{\left(\mathbf{I}(\mathbf{W}(\mathbf{x_p}, \mathbf{p}, \mathbf{q})) - \mathbf{A}_0(\mathbf{x_p})\right)}_{\mathbf{E}(\mathbf{x_p})_{\text{lk}}}$$

Error Image

- Normalize the Error Image

$$\mathbf{E}_{\text{nfa}}(\mathbf{x_p}) = \mathbf{E}(\mathbf{x_p})_{\text{lk}} - \sum_{i=1}^{m+2} \lambda_i \mathbf{A}_i(\mathbf{x_p})$$



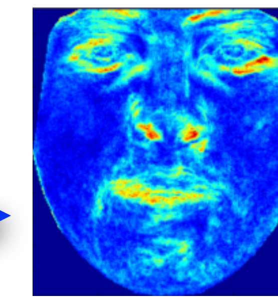


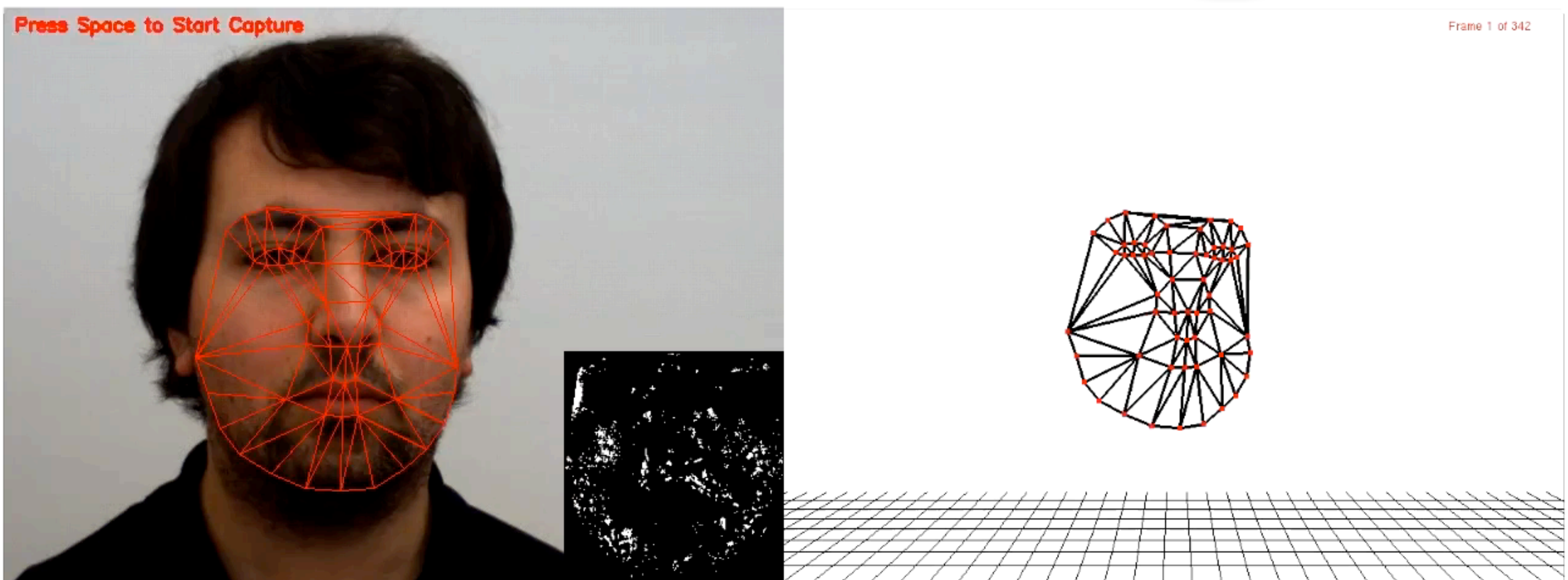
Robust Fitting

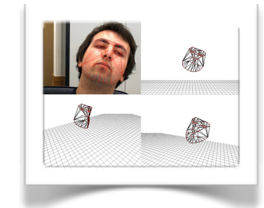
$\underset{\mathbf{p},\mathbf{q},\boldsymbol{\lambda}}{\arg\min} \sum_{\mathbf{x_p} \in s_{0\mathbf{p}}}$  $\Big[$  $-$  $\Big]^2$

$$\underset{\mathbf{p},\mathbf{q},\boldsymbol{\lambda}}{\arg\min} \sum_{\mathbf{x_p} \in s_{0\mathbf{p}}} \rho\left(\left[\mathbf{A}_0(\mathbf{x_p}) + \sum_{i=1}^{m+2} \lambda_i \mathbf{A}_i(\mathbf{x_p}) - \mathbf{I}(\mathbf{W}(\mathbf{x_p},\mathbf{p},\mathbf{q}))\right]^2, \underline{\sigma_{\mathbf{x_p}}}\right) \rightarrow$$

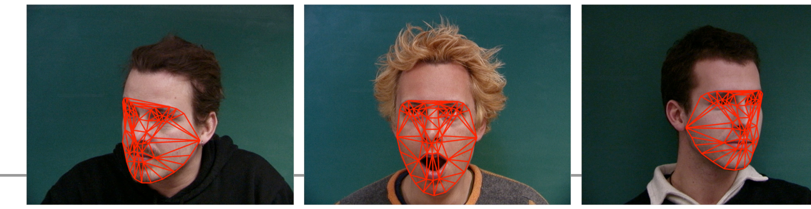
MAD



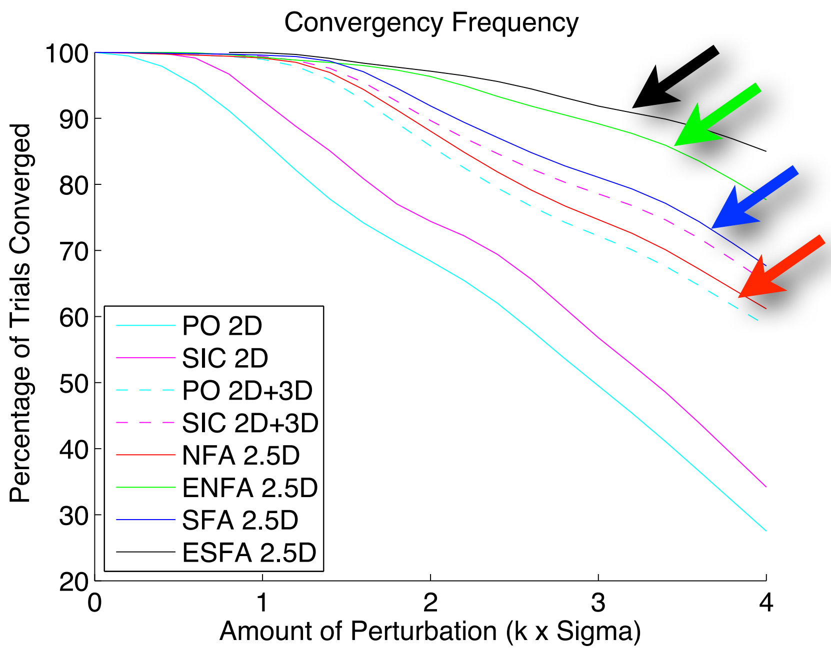




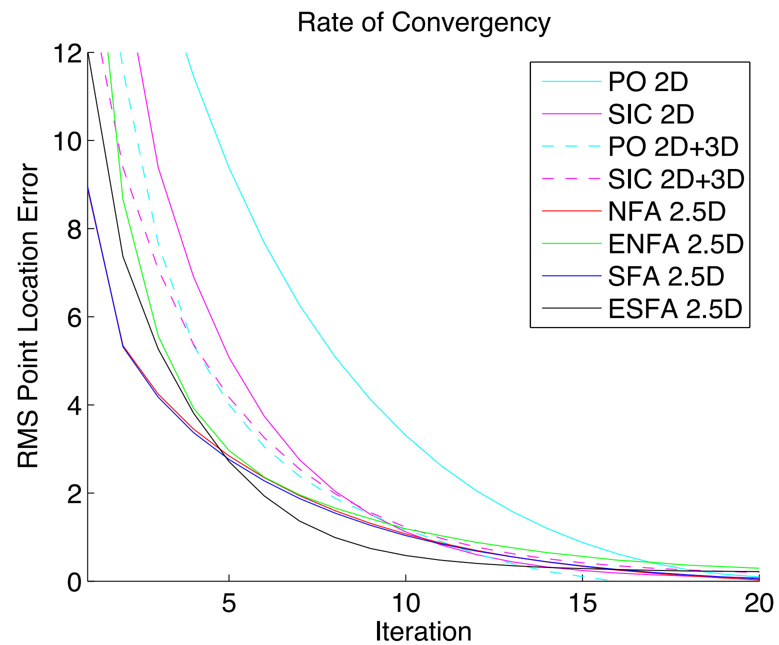
Evaluation Results



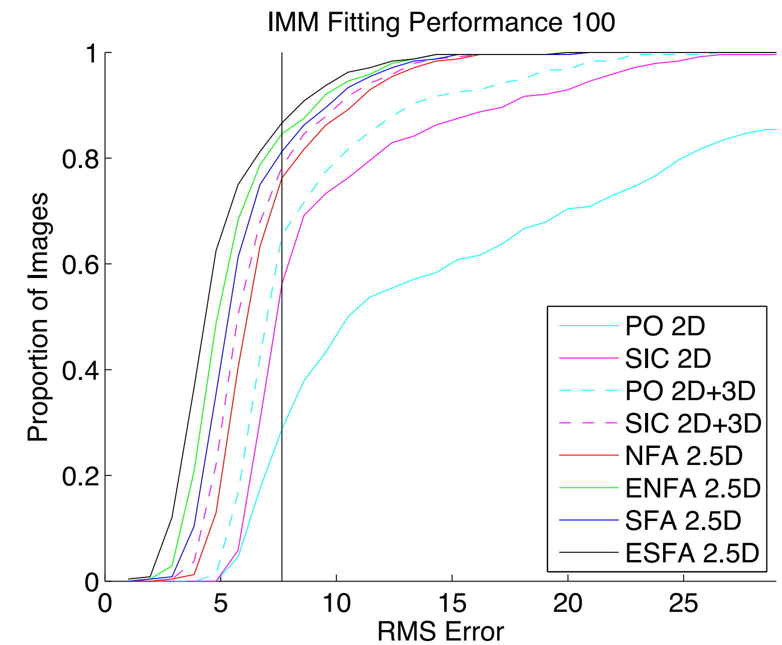
Convergency Frequency



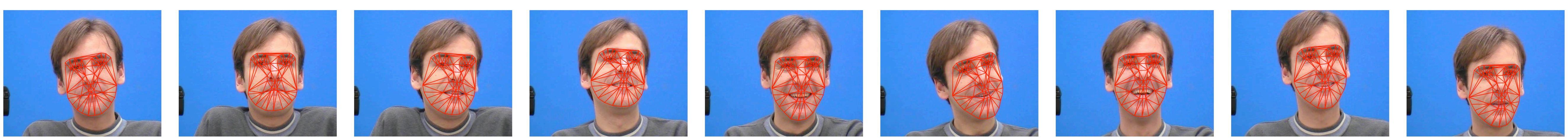
Rate of Convergency



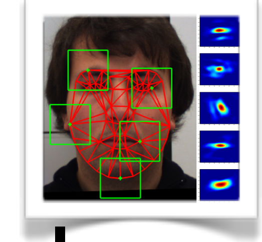
Fitting Performance Curve



Tracking Performance - FGNET Talking Face

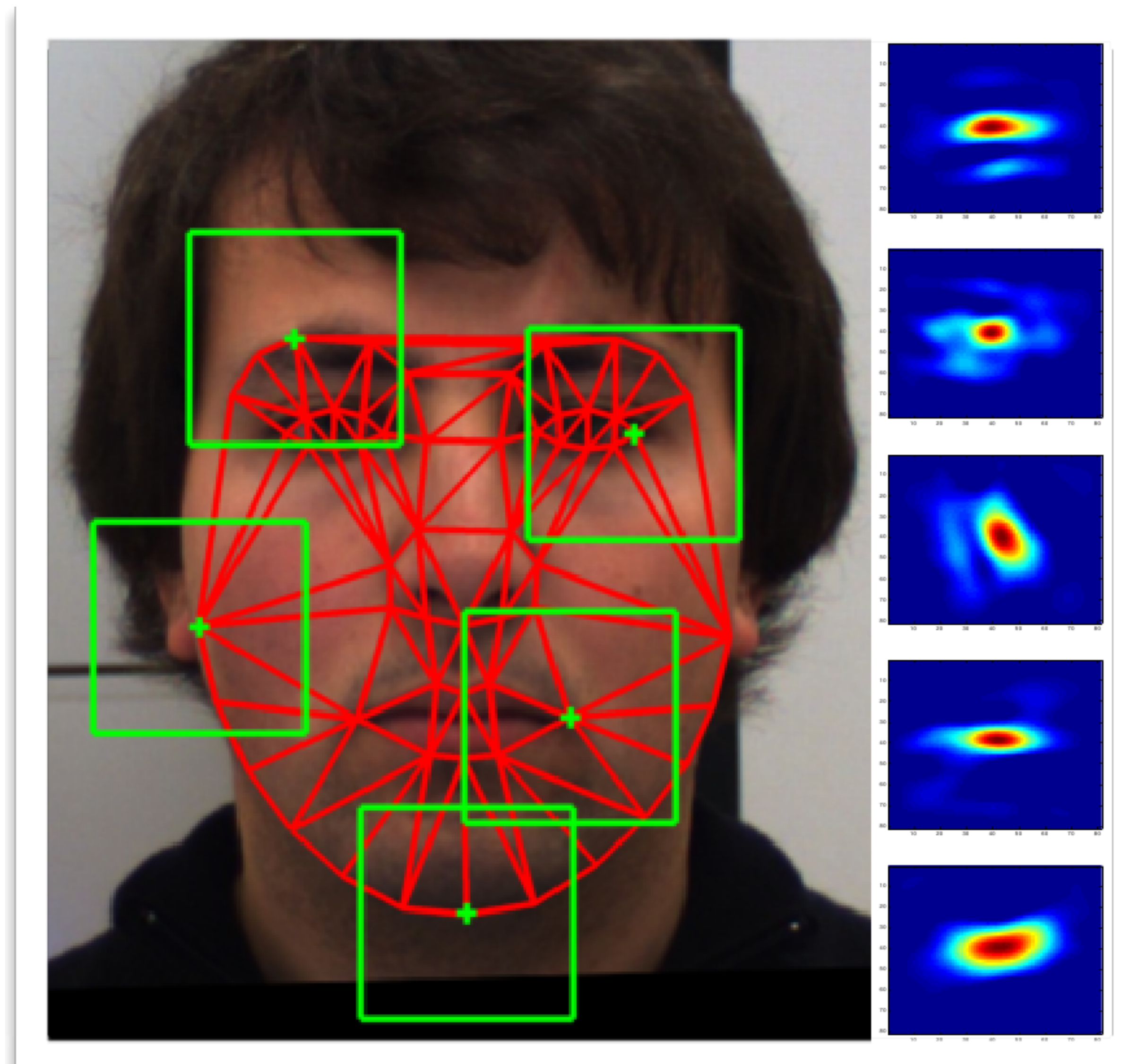


RMS Error	PO	PO 2D+3D	NFA 2.5D	ENFA 2.5D	SIC 2D	SIC 2D+3D	SFA 2.5D	ESFA 2.5D
Mean	7.4	7.0	6.6	6.2	7.1	6.6	6.4	6.0
Standard Deviation	3.4	2.5	2.1	1.3	3.3	3.2	1.4	1.2

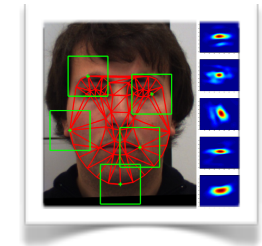


(2) Discriminative Bayesian Active Shape Models

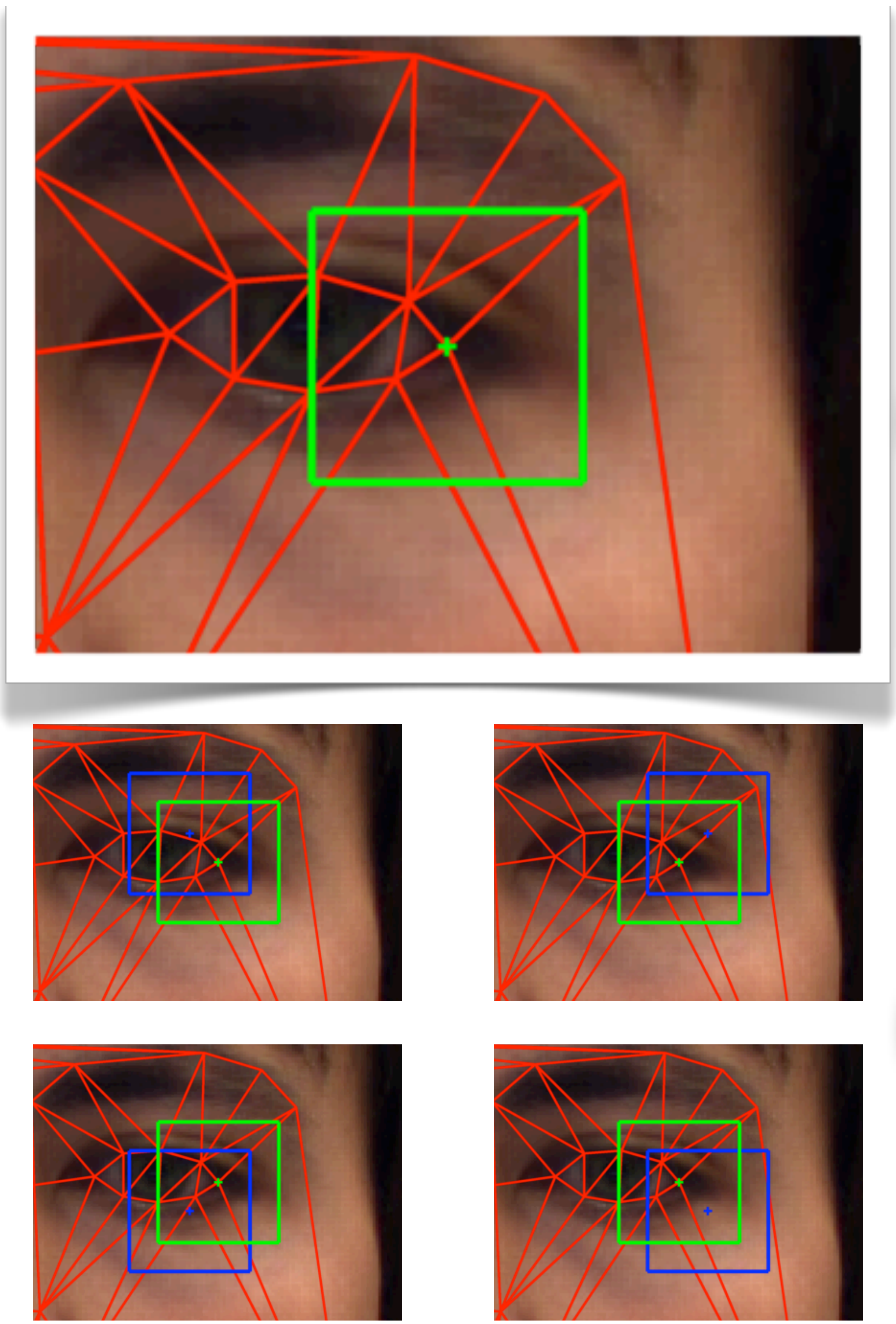
- Related to CLM and/or ASM, where a set of local detectors is constrained to lie in the subspace spanned by a PDM.
- Two step model fitting approach:
    - (1) Local search using the detectors.
    - (2) Global optimization strategy that finds the PDM parameters that jointly maximize all the detections.
- New Bayesian global optimization strategy that infers the overall alignment using a second order estimate of the PDM parameters.
- Extension that models the prior distribution.
- Performance in unseen data
- Efficient and simple approach.
- Fusion of multiple detectors.



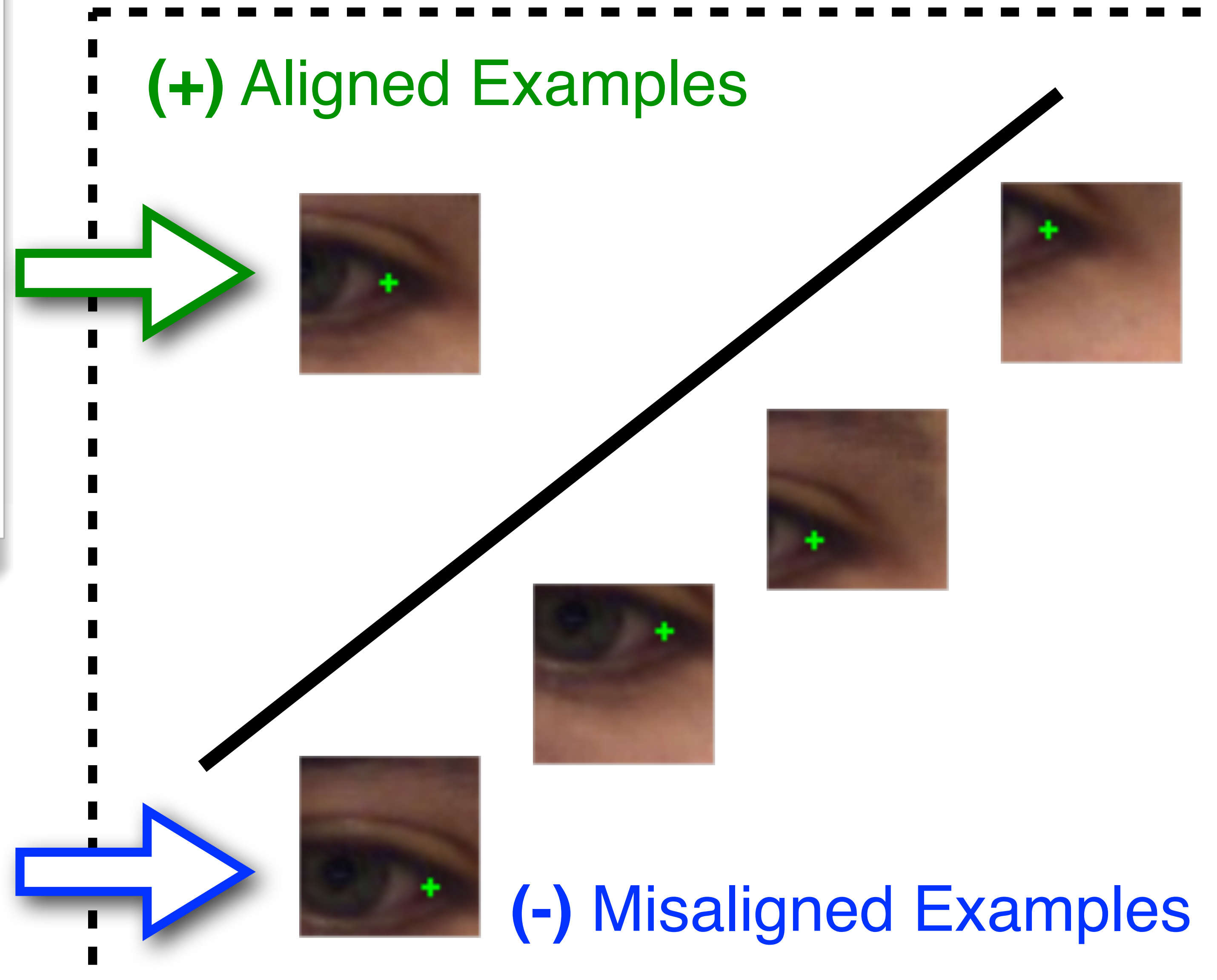
- **Previous Work:**
    - Active Shape Models (ASM) - CVIU 1995
    - Constrained Local Model (CLM) - BMVC 2006
    - Convex Quadratic Fitting (CQF) - CVPR 2008
    - Subspace Constrained Mean-Shifts (SCMS) - ICCV 2009



Local Landmark Detectors

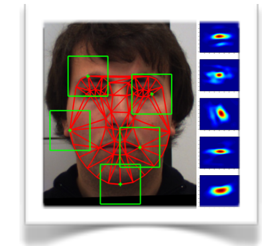


Linear SVM



$$\mathcal{D}_i^{\text{linear}}(\mathbf{I}(\mathbf{y}_i)) = \mathbf{w}_i^T \mathbf{I}(\mathbf{y}_i) + b_i$$

$$i = 1, \ldots, v \text{ landmarks}$$



The Alignment Goal

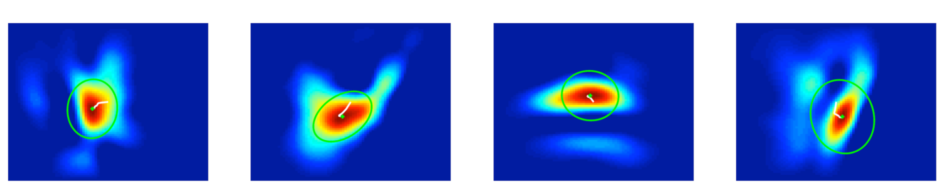
- Given a shape observation vector **(y)**, find the optimal set of shape (and pose) parameters **(b)** that maximize the posterior probability

$$\mathbf{b}^* = \arg\max_{\mathbf{b}} p(\mathbf{b}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{b})p(\mathbf{b})$$

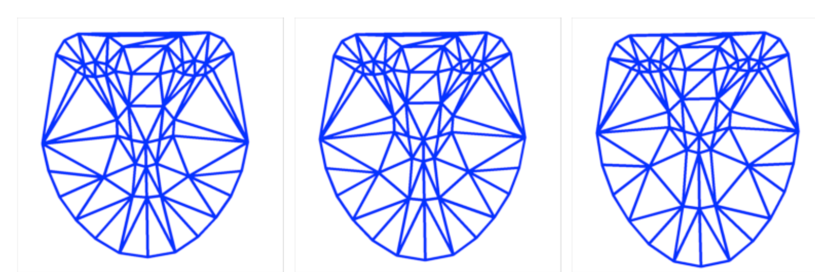
- Assuming:
  - Conditional independence between landmarks
  - Close to a solution

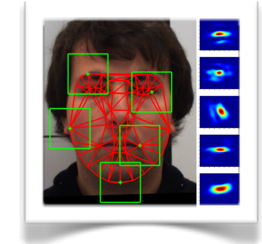
$$p(\mathbf{b}|\mathbf{y}) \propto \left(\prod_{i=1}^{v} p(\mathbf{y}_i|\mathbf{b})\right) p(\mathbf{b}|\mathbf{b}^*_{k-1})$$

Likelihood from the local detectors



Prior on how parameters change



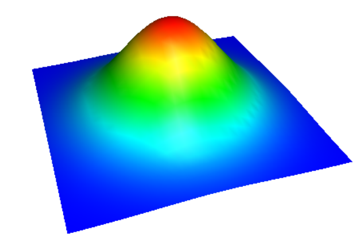


The Likelihood Term

- Convex energy function:

Observed shape **(y)**

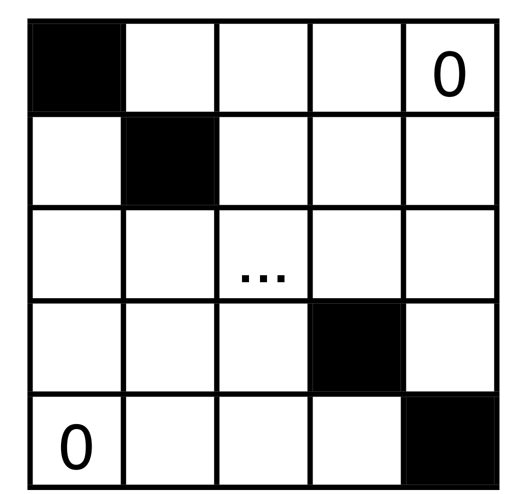
$$p(\mathbf{y}|\mathbf{b}) \propto \exp\left(-\frac{1}{2}(\underbrace{\mathbf{y} - (\mathbf{s}_0}_{\Delta\mathbf{y}} + \Phi\mathbf{b}))^T \Sigma_{\mathbf{y}}^{-1} (\mathbf{y} - (\mathbf{s}_0 + \Phi\mathbf{b})) \right)$$



Uncertainty covariance

Difference between the
observed and the **mean shape**

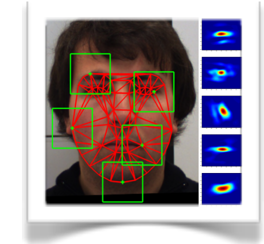
$\Sigma_{\mathbf{y}}$



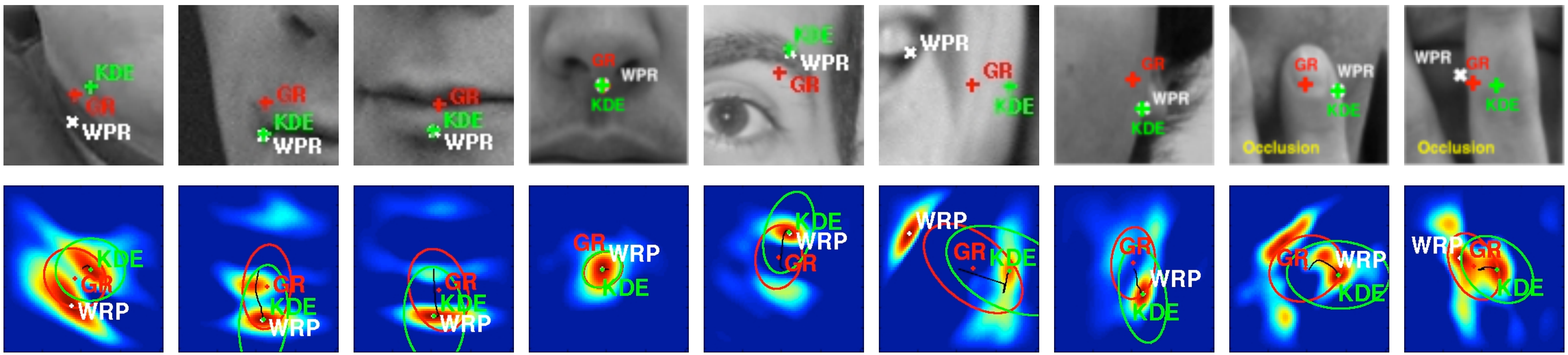
$2v \times 2v$ Block diagonal

The likelihood follow a Gaussian distribution

$$p(\mathbf{y}|\mathbf{b}) \propto \mathcal{N}(\Delta\mathbf{y}|\Phi\mathbf{b}, \Sigma_{\mathbf{y}})$$



Local Optimization Strategies



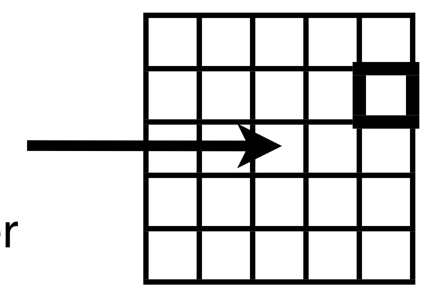
Patches under occlusion

$p_i(\mathbf{z}_i)$

Prob. $\mathbf{z}_i$ is aligned

$\mathbf{y}_i^c$

Center



$\mathbf{z}_i = (x_i, y_i)$

Pixel candidate to i[th] landmark location

$\boldsymbol{\Omega}_{\mathbf{y}_i^c}$ Local search grid

Weighted Peak Response (WPR)

$$\mathbf{y}_i^{\mathrm{WPR}} = \max_{\mathbf{z}_i \in \boldsymbol{\Omega}_{\mathbf{y}_i^c}} (p_i(\mathbf{z}_i))$$

$$\Sigma_{\mathbf{y}_i}^{\mathrm{WPR}} = diag(p_i(\mathbf{y}_i^{\mathrm{WPR}})^{-1})$$

ASM [BMVC92]

Gaussian Response (GR)

$$\mathbf{y}_i^{\mathrm{GR}} = \frac{1}{d} \sum_{\mathbf{z}_i \in \boldsymbol{\Omega}_{\mathbf{y}_i^c}} p_i(\mathbf{z}_i)\mathbf{z}_i \qquad d = \sum_{\mathbf{z}_i \in \boldsymbol{\Omega}_{\mathbf{y}_i^c}} p_i(\mathbf{z}_i)$$

$$\Sigma_{\mathbf{y}_i}^{\mathrm{GR}} = \frac{1}{d-1} \sum_{\mathbf{z}_i \in \boldsymbol{\Omega}_{\mathbf{y}_i^c}} p_i(\mathbf{z}_i)(\mathbf{z}_i - \mathbf{y}_i^{\mathrm{GR}})(\mathbf{z}_i - \mathbf{y}_i^{\mathrm{GR}})^T$$

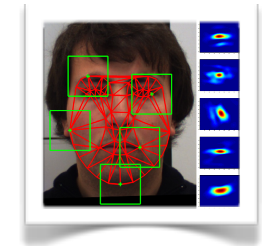
CQF, BCLM [CVPR2008, CVPR2009]

Kernel Density Estimator (KDE)

$$\mathbf{y}_i^{\mathrm{KDE}(\tau+1)} \leftarrow \frac{\sum_{\mathbf{z}_i \in \boldsymbol{\Omega}_{\mathbf{y}_i^c}} \mathbf{z}_i \; p_i(\mathbf{z}_i) \; \mathcal{N}(\mathbf{y}_i^{\mathrm{KDE}(\tau)} | \mathbf{z}_i, \sigma_{h_j}^2 \mathbf{I}_2)}{\sum_{\mathbf{z}_i \in \boldsymbol{\Omega}_{\mathbf{y}_i^c}} p_i(\mathbf{z}_i) \; \mathcal{N}(\mathbf{y}_i^{\mathrm{KDE}(\tau)} | \mathbf{z}_i, \sigma_{h_j}^2 \mathbf{I}_2)}$$

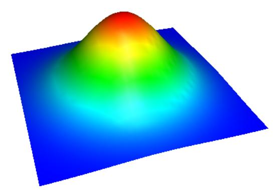
$$\Sigma_{\mathbf{y}_i}^{\mathrm{KDE}} = \frac{1}{d-1} \sum_{\mathbf{z}_i \in \boldsymbol{\Omega}_{\mathbf{y}_i^c}} p_i(\mathbf{z}_i)(\mathbf{z}_i - \mathbf{y}_i^{\mathrm{KDE}})(\mathbf{z}_i - \mathbf{y}_i^{\mathrm{KDE}})^T$$

SCMS [ICCV2009]



MAP Global Alignment (DBASM)

Likelihood

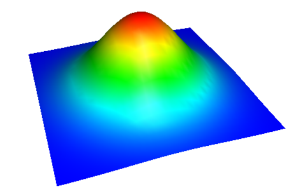


$$p(\mathbf{y}|\mathbf{b}) \propto \mathcal{N}(\Delta\mathbf{y}|\Phi\mathbf{b}, \Sigma_{\mathbf{y}})$$

Mean that is function of **b**
Covariance independent of **b**

Bayes's Theorem for Gaussian variables

Posterior

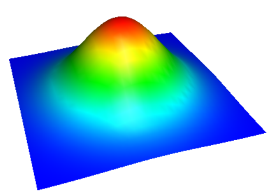


$$p(\mathbf{b}_k|\mathbf{y}) \propto \mathcal{N}(\mathbf{b}_k|\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

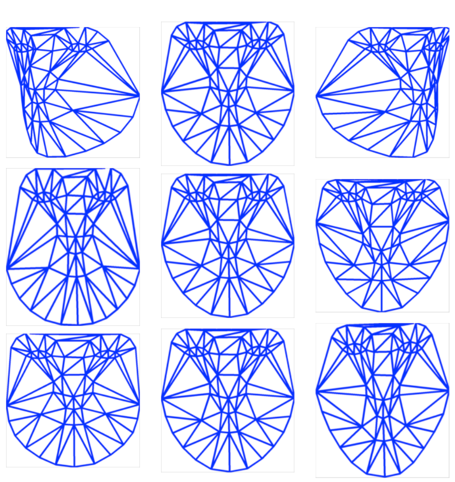
$$\boldsymbol{\Sigma} = (\Sigma_{\mathbf{b}}^{-1} + \Phi^T\Sigma_{\mathbf{y}}^{-1}\Phi)^{-1}$$

$$\boldsymbol{\mu} = \boldsymbol{\Sigma}(\Phi^T\Sigma_{\mathbf{y}}^{-1}\mathbf{y} + \Sigma_{\mathbf{b}}^{-1}\mu_{\mathbf{b}})$$

Prior



$$p(\mathbf{b}_k|\mathbf{b}_{k-1}) \propto \mathcal{N}(\mathbf{b}_k|\mu_{\mathbf{b}}, \Sigma_{\mathbf{b}})$$



$$\Sigma_{\mathbf{b}} = \Lambda + \Xi$$

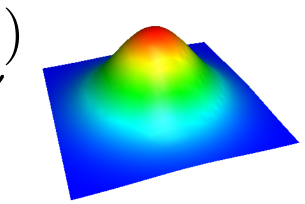
PCA eigenvalues + additive dynamic noise

Model the Covariance of b (2nd Order Estimate)

Linear Dynamic System (LDS)

$$\begin{aligned} \mathbf{b}_k &= \mathbf{A}\mathbf{b}_{k-1} + q \qquad & q &\sim \mathcal{N}(\mathbf{0}, \Sigma_{\mathbf{b}}) \\ \Delta\mathbf{y} &= \Phi\mathbf{b}_k + r \qquad & r &\sim \mathcal{N}(\mathbf{0}, \Sigma_{\mathbf{y}}) \end{aligned}$$

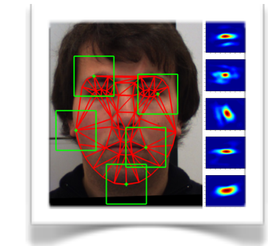
$$\mathcal{N}(\mathbf{A}\boldsymbol{\mu}_k^{\mathbf{F}}, \underbrace{(\Lambda + \Xi) + \mathbf{A}\boldsymbol{\Sigma}_{k-1}^{\mathbf{F}}\mathbf{A}^T}_{\mathbf{P}_{k-1}})$$



Posterior

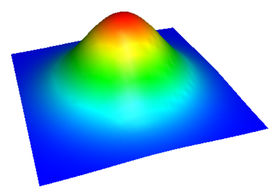
$$p(\mathbf{b}_k|\mathbf{y}_k, \ldots, \mathbf{y}_0) \propto \mathcal{N}(\mathbf{b}_k|\boldsymbol{\mu}_k^{\mathbf{F}}, \boldsymbol{\Sigma}_k^{\mathbf{F}})$$

$$\begin{aligned} \boldsymbol{\mu}_k^{\mathbf{F}} &= \mathbf{A}\boldsymbol{\mu}_{k-1}^{\mathbf{F}} + \mathbf{K}(\mathbf{y} - \Phi\mathbf{A}\boldsymbol{\mu}_{k-1}^{\mathbf{F}}) \\ \boldsymbol{\Sigma}_k^{\mathbf{F}} &= (\mathbf{I}_n - \mathbf{K}\Phi)\mathbf{P}_{k-1} \end{aligned}$$



MAP Global Alignment (BASM)

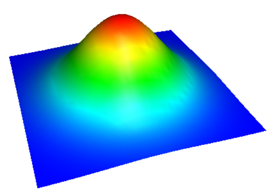
Likelihood



$$p(\mathbf{y}|\mathbf{b}) \propto \mathcal{N}(\Delta\mathbf{y}|\Phi\mathbf{b}, \Sigma_{\mathbf{y}})$$

- 2nd Order Estimate
- Bayesian Fusion of Multiple Detectors

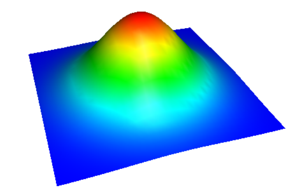
Prior



$$p(\mathbf{b}_k|\mathbf{b}_{k-1}) \propto \mathcal{N}(\mathbf{b}_k|\boxed{\mu_{\mathbf{b}}, \Sigma_{\mathbf{b}}})$$

Unknown

Posterior



$$p(\mathbf{b}_k|\mathbf{y}_k, \ldots, \mathbf{y}_0) \propto \mathcal{N}(\mathbf{b}_k|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

$$\boldsymbol{\Sigma}_k = \left((\Sigma_{\mathbf{b}_k} + \Sigma_{k-1})^{-1} + \Phi^T \sum_{m=1}^{M} \left(\Sigma^{-1}_{\mathbf{y}_{(m)}} \right) \Phi \right)^{-1}$$

$$\boldsymbol{\mu}_k = \boldsymbol{\Sigma}_k \left(\Phi^T \sum_{m=1}^{M} \left(\Sigma^{-1}_{\mathbf{y}_{(m)}} \Delta\mathbf{y}_{(m)} \right) + (\Sigma_{\mathbf{b}_k} + \Sigma_{k-1})^{-1} \mu_{\mathbf{b}_k} \right)$$

Observable vector **b**

$$p(\mu_{\mathbf{b}}, \Sigma_{\mathbf{b}}|\mathbf{b}) \propto p(\mathbf{b}|\mu_{\mathbf{b}}, \Sigma_{\mathbf{b}})\ p(\mu_{\mathbf{b}}, \Sigma_{\mathbf{b}})$$

Joint Posterior
Normal Inverse-Wishart

Joint Prior
Normal Inverse-Wishart

Conjugate Prior for a Gaussian with unknown mean and covariance is a Normal Inverse-Wishart distribution

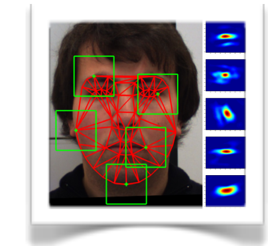
$p(\mu_{\mathbf{b}}|\mathbf{b}) \propto$ Multivariate Student t

$$\mu_{\mathbf{b}_k} = E(\mu_{\mathbf{b}}|\mathbf{b}) = \theta_k$$

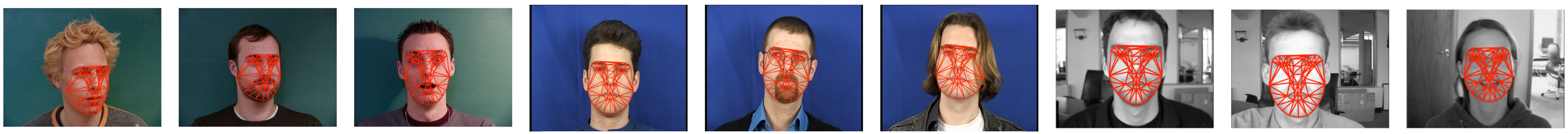
$p(\Sigma_{\mathbf{b}}|\mathbf{b}) \propto$ Inv-Wishart

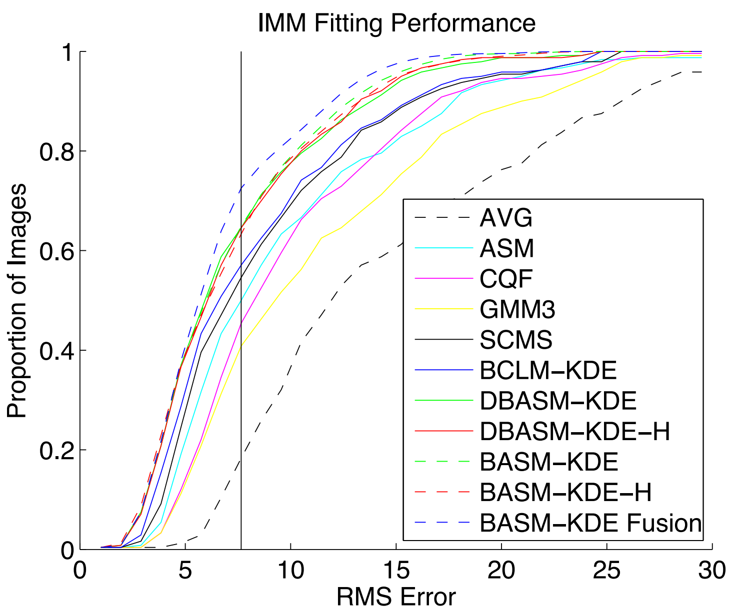
$$\Sigma_{\mathbf{b}_k} = E(\Sigma_{\mathbf{b}}|\mathbf{b}) = (v_k - n - 1)^{-1} \Lambda_k$$

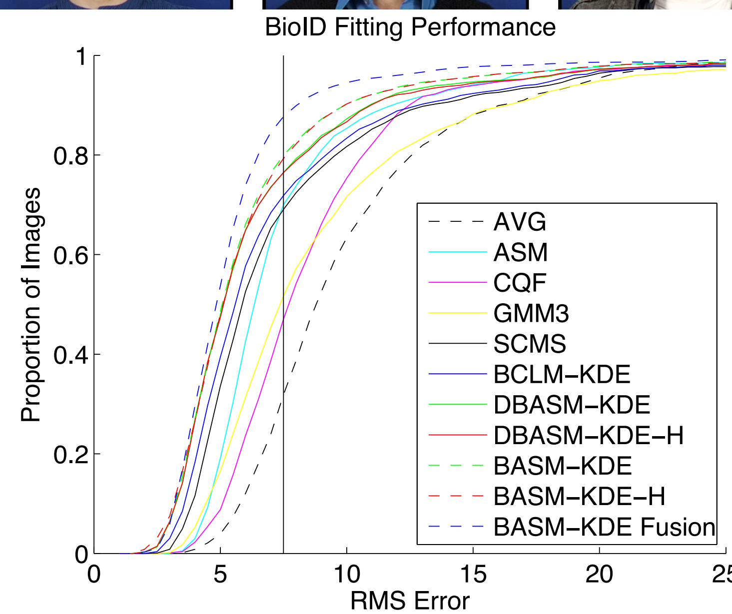
The Prior distribution is **continuously kept up to date**

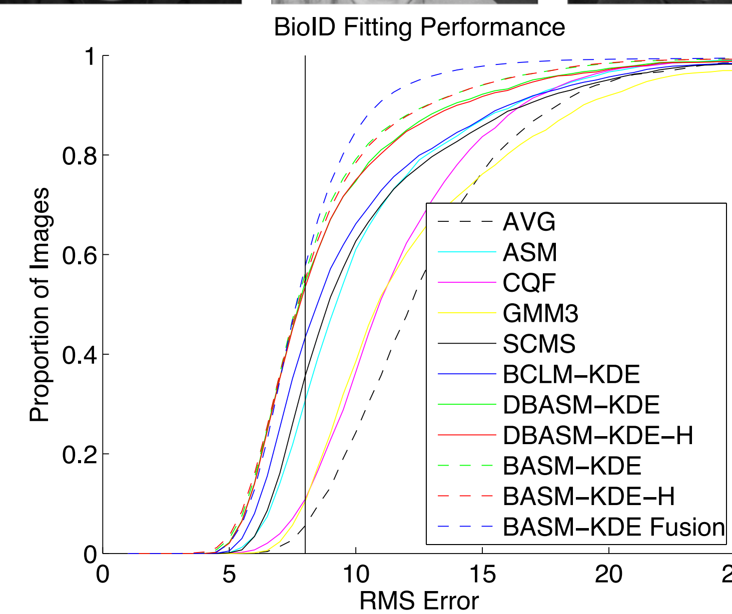


Evaluation Results

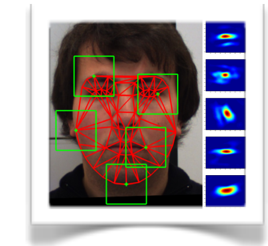




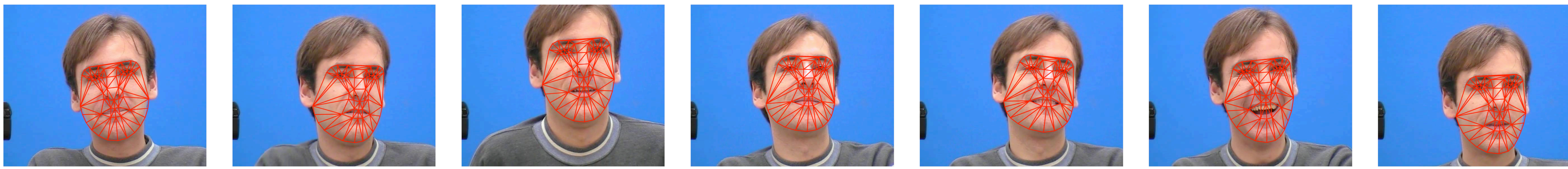


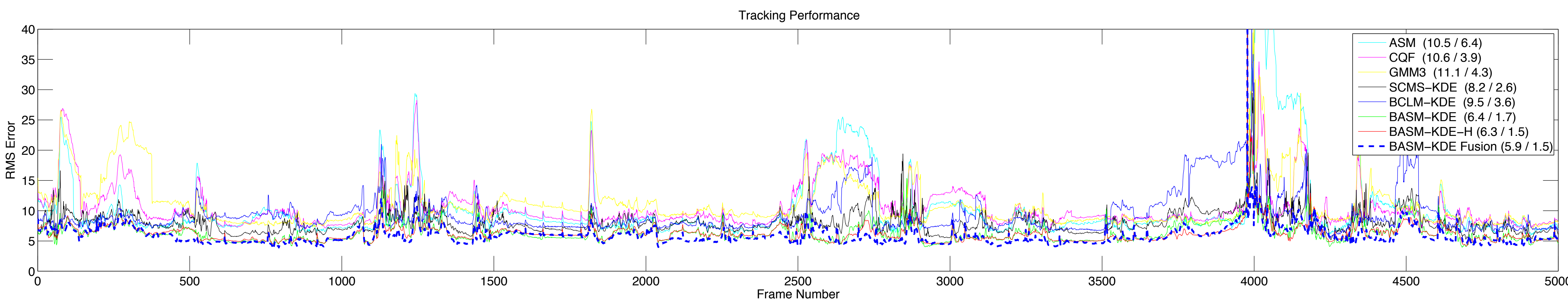


Reference 7.5 RMS	IMM (240 images)		XM2VTS (2360 images)		BioID (1521 images)	
ASM	50.0		30.7		70.0	
DBASM-WPR	56.7	(+6.7)	45.1	(+14.4)	75.4	(+5.4)
BASM-WPR	**58.4**	(+8.4)	**47.4**	(+16.7)	**77.1**	(+7.1)
CQF	45.4		10.9		47.0	
GMM3	40.8	(-4.6)	10.4	(-0.5)	51.7	(+4.7)
BCLM-GR	48.3	(+2.9)	15.9	(+5.0)	54.2	(+7.2)
DBASM-GR	50.4	(+5.0)	18.0	(+7.1)	62.2	(+15.2)
BASM-GR	**51.8**	(+6.4)	**19.7**	(+8.8)	**63.5**	(+16.5)
SCMS-KDE	54.6		35.7		69.0	
BCLM-KDE	57.1	(+2.5)	43.4	(+7.7)	71.9	(+2.9)
DBASM-KDE	64.6	(+10.0)	54.5	(+18.8)	76.5	(+7.5)
DBASM-KDE-H	64.6	(+10.0)	53.5	(+17.8)	76.5	(+7.5)
BASM-KDE	**65.4**	(+10.8)	**57.0**	(+21.3)	**80.3**	(+11.3)
BASM-KDE-H	64.0	(+9.4)	56.6	(+20.9)	79.9	(+10.9)
BASM-KDE Fusion of 2 Detectors	**72.5**	(+17.9)	**58.7**	(+23.0)	**88.2**	(+19.2)

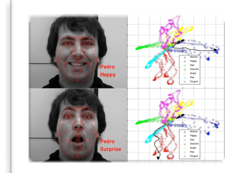


Tracking Performance - FGNET Talking Face



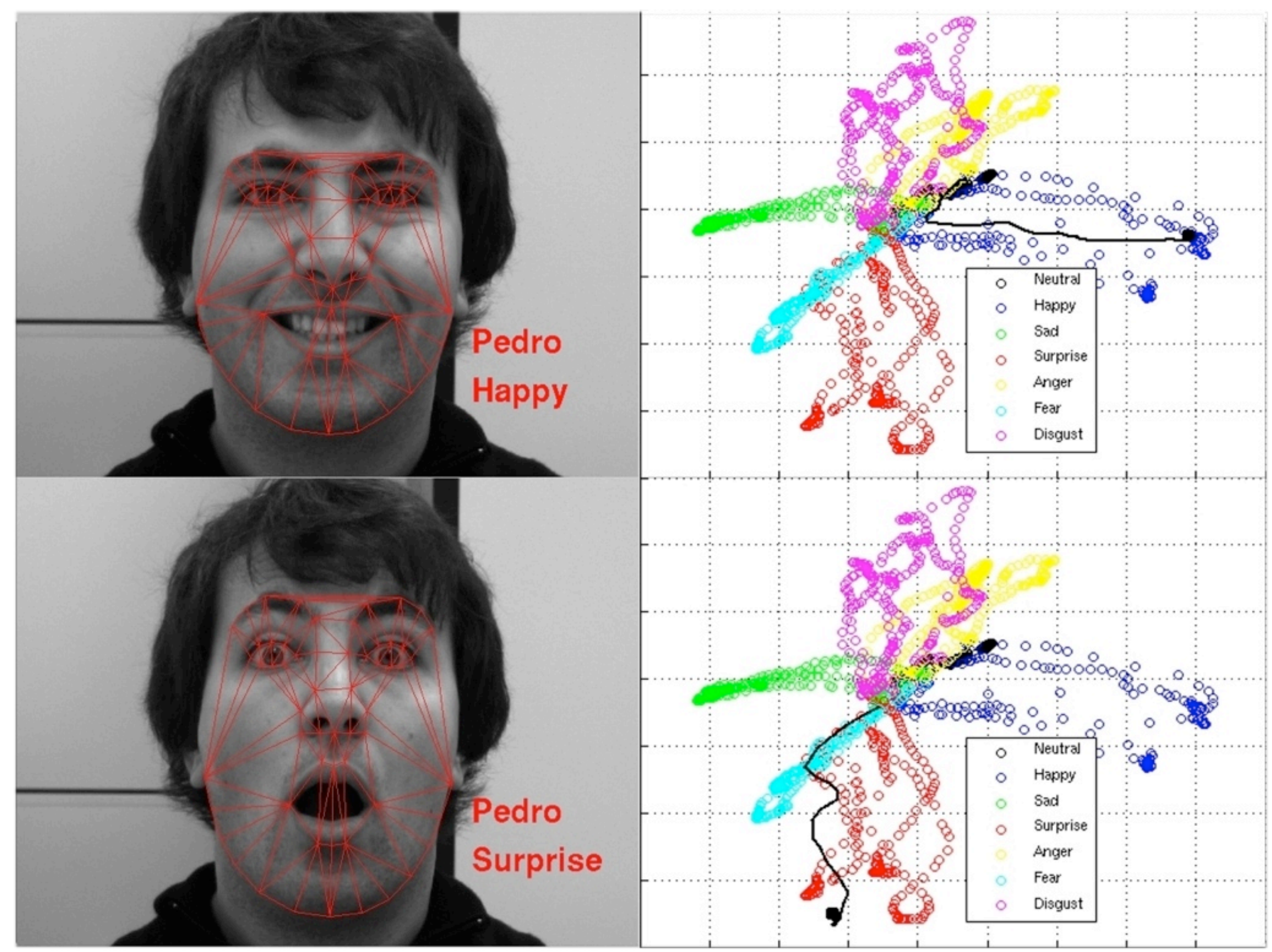


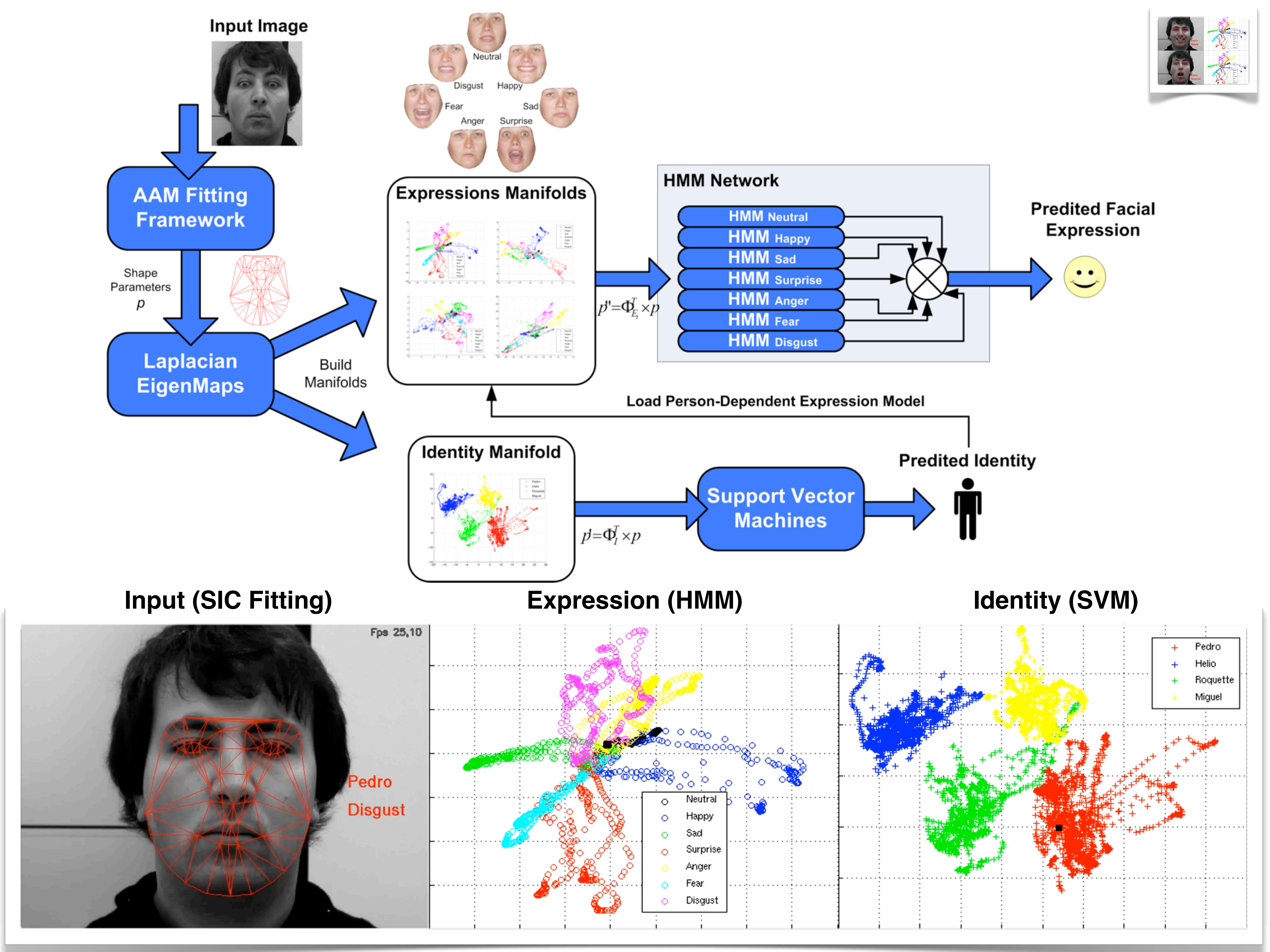
RMS Error	ASM	CQF	GMM3	SCMS-KDE	BCLM-KDE	DBASM-KDE	DBASM-KDE-H	BASM-KDE	BASM-KDE-H	BASM-KDE Fusion
Mean	10.5	10.6	11.1	8.2	9.5	7.0	7.2	6.4	6.3	5.9
Standard Deviation	6.4	3.9	4.3	2.6	3.6	2.1	2.2	1.7	1.5	1.5



(3) Identity and Facial Expression Recognition

- Six basic emotions (happiness, sadness, surprise, anger, fear, disgust) plus the neutral expression.
- Identity and facial expression recognition using facial geometry (captured by the AAM).
- Low dimensional manifolds derived using Laplacian EigenMaps
  - Identity
  - Person-dependent expression
- Two step recognition approach.





Input Image
Neutral
Disgust
Happy
Fear
Sad
Anger
Surprise
AAM Fitting Framework
Shape Parameters p
Laplacian EigenMaps
Build Manifolds
Expressions Manifolds
$p'=\Phi_{E_i}^T \times p$
HMM Network
HMM Neutral
HMM Happy
HMM Sad
HMM Surprise
HMM Anger
HMM Fear
HMM Disgust
Predited Facial Expression
Load Person-Dependent Expression Model
Identity Manifold
$p'=\Phi_I^T \times p$
Support Vector Machines
Predited Identity
Input (SIC Fitting)
Expression (HMM)
Identity (SVM)
Fps 25,10
Pedro
Disgust
Neutral
Happy
Sad
Surprise
Anger
Fear
Disgust
Pedro
Helio
Roquette
Miguel

Conclusions

- **(1) Generative Face Alignment (2.5D Active Appearance Models)**
  - Extension of the standard 2D Active Appearance Models to deal with a full perspective projection model.
  - Two model fitting algorithms (SFA, NFA) and their efficient approximations.
  - Robust solutions account for partial and self occlusions.
- **(2) Discriminative Face Alignment (Bayesian Active Shape Models)**
  - New Bayesian formulation for aligning faces in unseen images.
  - New global optimization strategy that infers both shape and pose parameters, in MAP sense, using second order statistics.
  - Extension that models the prior distribution.
- **(3) Identity and Facial Expression Recognition**
  - Two step recognition approach (identity then expression) using low dimensional representations of face geometry.

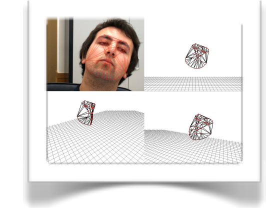
Future Work

- **Unconstrained Non-Rigid Registration**
    - 3D Point Distribution Model.
    - Extend the likelihood term to a non-parametric distribution.
    - Shape representation (non-parametric shape model).
    - Constrain model fitting using 3D depth data.
    - 3D dynamic recognition (identity and facial expression).

The End

Additional Slides

- Linear Pose Updates
- Efficient Approximations (ESFA, ENFA)
- Combined 2D+3D AAM
- MOSSE Filters
- KDE Landmark Updates
- BASM - The Algorithm
- Hierarchical Search (KDE-H)
- Tracking Performance - FGNET Talking Face (Video)
- The Recognition Approach (Overview)



Linear Pose Updates

$$\mathbf{R}(\mathbf{w},\theta) = e^{\hat{\mathbf{w}}\theta} = \sum_{n=0}^{\infty} \frac{(\hat{\mathbf{w}}\theta)^n}{n!} = \mathbf{I}_3 + \hat{\mathbf{w}}\theta + \frac{(\hat{\mathbf{w}}\theta)^2}{2!} + \frac{(\hat{\mathbf{w}}\theta)^3}{3!} + \cdots$$

$$\hat{\mathbf{w}} = \begin{bmatrix} 0 & -w_z & w_y \\ w_z & 0 & -w_x \\ -w_y & w_x & 0 \end{bmatrix}$$

$$\mathbf{P}'_i = \underbrace{\begin{bmatrix} 1 & -w_z & w_y \\ w_z & 1 & -w_x \\ -w_z & w_x & 1 \end{bmatrix}}_{\mathbf{R}(\mathbf{w})} \underbrace{\begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix}}_{\mathbf{P}_i} + \underbrace{\begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}}_{\mathbf{T}_i}$$

Pose Parameters

$$\mathbf{q} = [w_x \; w_y \; w_z \; t_x \; t_y \; t_z]^T$$

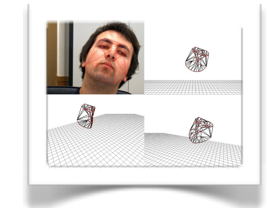
$$\mathbf{P}'_i = \begin{bmatrix} 0 & Z_i & -Y_i & 1 & 0 & 0 \\ -Z_i & 0 & X_i & 0 & 1 & 0 \\ Y_i & -X_i & 0 & 0 & 0 & 1 \end{bmatrix} \mathbf{q}$$

$$\sum_{j=1}^{6} q_j \psi_j^{(t)}$$

$$\underbrace{\left[\begin{array}{cccccc} 0 & s_0^{z_1} & -s_0^{y_1} & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & s_0^{z_v} & -s_0^{y_v} & 1 & 0 & 0 \\ \hline -s_0^{z_1} & 0 & s_0^{x_1} & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -s_0^{z_v} & 0 & s_0^{x_v} & 0 & 1 & 0 \\ \hline s_0^{y_1} & -s_0^{x_1} & 0 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ s_0^{y_v} & -s_0^{x_v} & 0 & 0 & 0 & 1 \end{array}\right]}_{\psi_1,\ldots,\psi_6} \mathbf{q}$$

$$\left[\begin{array}{cccccc} 0 & s_0^{z_1} + s_\psi^{z_1} & -s_0^{y_1} - s_\psi^{y_1} & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & s_0^{z_v} + s_\psi^{z_v} & -s_0^{y_v} - s_\psi^{y_v} & 1 & 0 & 0 \\ \hline -s_0^{z_1} - s_\psi^{z_1} & 0 & s_0^{x_1} + s_\psi^{x_1} & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -s_0^{z_v} - s_\psi^{z_v} & 0 & s_0^{x_v} + s_\psi^{x_v} & 0 & 1 & 0 \\ \hline s_0^{y_1} + s_\psi^{y_1} & -s_0^{x_1} - s_\psi^{x_1} & 0 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ s_0^{y_v} + s_\psi^{y_v} & -s_0^{x_v} - s_\psi^{x_v} & 0 & 0 & 0 & 1 \end{array}\right] \mathbf{q}$$

$$s_\psi = \int_0^{t-1} \sum_{j=1}^{6} q_j \psi_j^{(t)} \partial t$$



Efficient Approximations (ESFA, ENFA)

$$\mathbf{A}_0(\mathbf{x_p}) + \sum_{i=1}^{m+2} \lambda_i \mathbf{A}_i(\mathbf{x_p})$$

$$\left(\mathbf{A}_0(\mathbf{x_p}) + \sum_{i=1}^{m+2} \lambda_i \mathbf{A}_i(\mathbf{x_p})\right) \approx \mathbf{I}(\mathbf{W}(\mathbf{x_p}, \mathbf{p}, \mathbf{q}))$$

$$\Downarrow$$

$$\underbrace{\left(\nabla \mathbf{A}_0(\mathbf{x_p}) + \sum_{i=1}^{m+2} \lambda_i \nabla \mathbf{A}_i(\mathbf{x_p})\right)}_{\nabla \mathbf{A}_i(\mathbf{x_p}, \boldsymbol{\lambda})} \approx \nabla \mathbf{I}(\mathbf{W}(\mathbf{x_p}, \mathbf{p}, \mathbf{q}))$$

$$\mathbf{I}(\mathbf{W}(\mathbf{x_p}, \mathbf{p}, \mathbf{q}))$$

$$\mathbf{SD}(\mathbf{x_p})_{\text{enfa}} = \left[\nabla \mathbf{A}_0(\mathbf{x_p}) \frac{\partial \mathbf{W}}{\partial \mathbf{p}_1} \;\ldots\; \nabla \mathbf{A}_0(\mathbf{x_p}) \frac{\partial \mathbf{W}}{\partial \mathbf{p}_n} \; \nabla \mathbf{A}_0(\mathbf{x_p}) \frac{\partial \mathbf{W}}{\partial \mathbf{q}_1} \ldots \nabla \mathbf{A}_0(\mathbf{x_p}) \frac{\partial \mathbf{W}}{\partial \mathbf{q}_6}\right]$$

$$\mathbf{SD}(\mathbf{x_p})_{\text{esfa}} = \left[\nabla \mathbf{A}_i(\mathbf{x_p}, \boldsymbol{\lambda}) \frac{\partial \mathbf{W}}{\partial \mathbf{p}_1} \;\ldots\; \nabla \mathbf{A}_i(\mathbf{x_p}, \boldsymbol{\lambda}) \frac{\partial \mathbf{W}}{\partial \mathbf{p}_n} \; \nabla \mathbf{A}_i(\mathbf{x_p}, \boldsymbol{\lambda}) \frac{\partial \mathbf{W}}{\partial \mathbf{q}_1} \ldots \nabla \mathbf{A}_i(\mathbf{x_p}, \boldsymbol{\lambda}) \frac{\partial \mathbf{W}}{\partial \mathbf{q}_6} \;\; -\mathbf{A}_1(\mathbf{x_p}) \;\ldots\; -\mathbf{A}_{m+2}(\mathbf{x_p})\right]$$

Combined 2D+3D AAMs

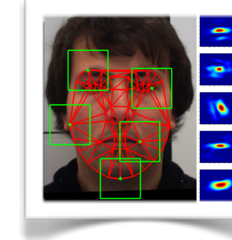
$$\sum_{\mathbf{x}\in\mathbf{s}_0}\left[\mathbf{A}_0(\mathbf{x})+\sum_{i=1}^{m}\lambda_i\mathbf{A}_i(\mathbf{x})-\mathbf{I}(\mathcal{S}(\mathbf{W}(\mathbf{x},\mathbf{p}),\mathbf{q})))\right]^2+$$

$$K.\left\|\mathbf{P}(\mathbf{s}_0^{3d}+\sum_{i=1}^{n3D}p_i^{3d}\phi_i^{3d})+\begin{pmatrix} o_x & \dots & o_x \\ o_y & \dots & o_y \end{pmatrix}-\mathcal{S}(s_0+\sum_{i=1}^{n}p_i\phi_i,\mathbf{q})\right\|^2$$

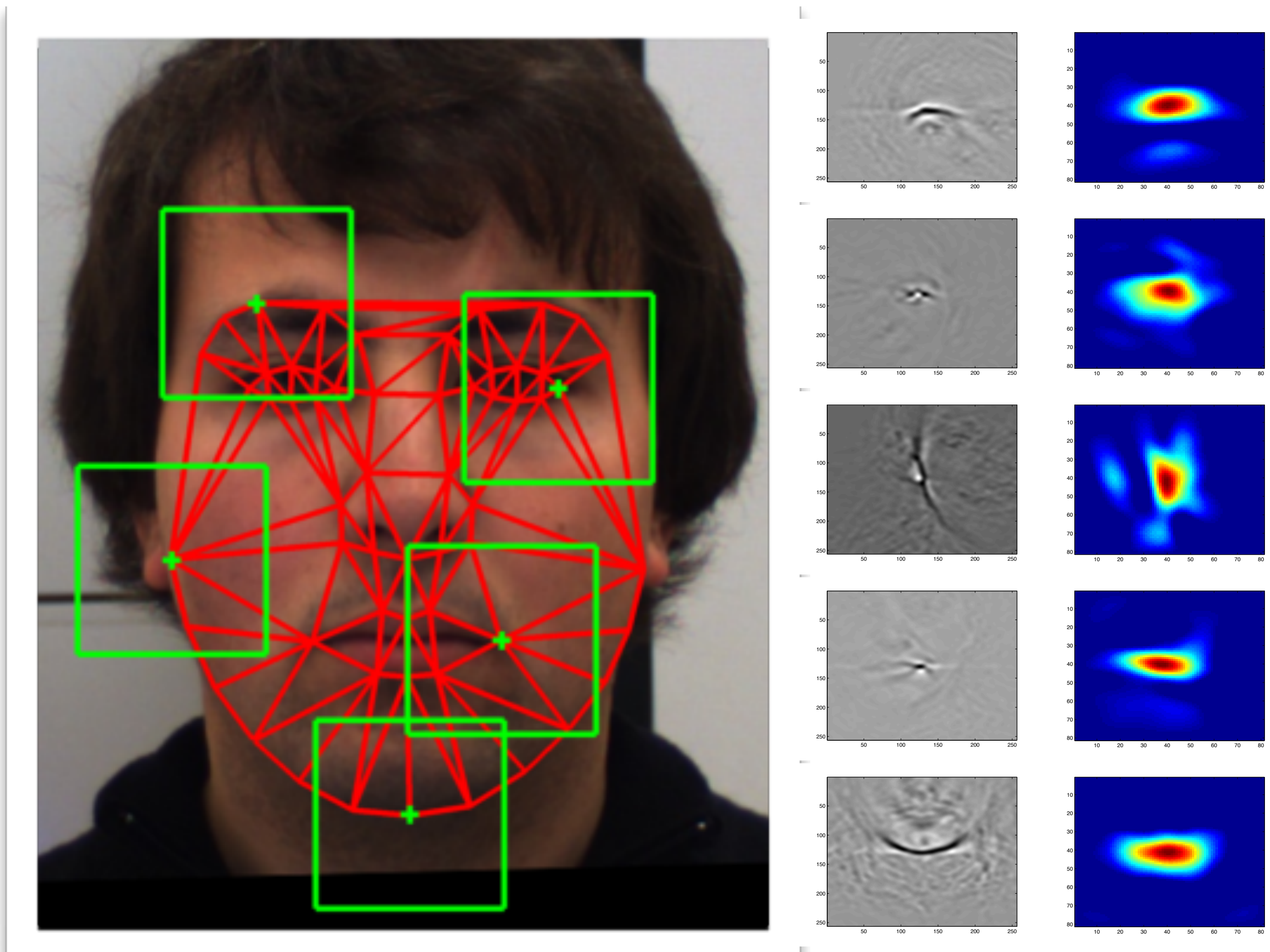
$$\mathbf{x}=\underbrace{\begin{pmatrix} i_x & i_y & i_z \\ j_x & j_y & j_z \end{pmatrix}}_{\mathbf{P}}\begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix}+\begin{pmatrix} o_x \\ o_y \end{pmatrix}$$

$$\begin{pmatrix} \Delta\mathbf{p} \\ \Delta\mathbf{q} \\ \Delta\mathbf{p}^{3d} \\ \Delta\mathbf{P} \\ \Delta o_x \\ \Delta o_y \end{pmatrix}=-\mathbf{H}_{3D}^{-1}\left(\begin{pmatrix} \Delta\mathbf{p}_{SD} \\ \Delta\mathbf{q}_{SD} \\ \mathbf{0} \\ \mathbf{0} \\ 0 \\ 0 \end{pmatrix}+K.\ \mathbf{SD}_F^T\ F(\mathbf{p},\mathbf{q},\mathbf{p}^{3D},\mathbf{P},o_x,o_y)\right)$$

$$\mathbf{SD}_F=\begin{pmatrix} \frac{\partial F}{\partial\mathbf{p}}\mathbf{J_p} & \frac{\partial F}{\partial\mathbf{q}}\mathbf{J_q} & \frac{\partial F}{\partial\mathbf{p}^{3d}} & \frac{\partial F}{\partial\sigma} & \frac{\partial F}{\partial\Delta\theta_x} & \frac{\partial F}{\partial\Delta\theta_y} & \frac{\partial F}{\partial\Delta\theta_z} & \frac{\partial F}{\partial\Delta o_x} & \frac{\partial F}{\partial\Delta o_y} \end{pmatrix}$$



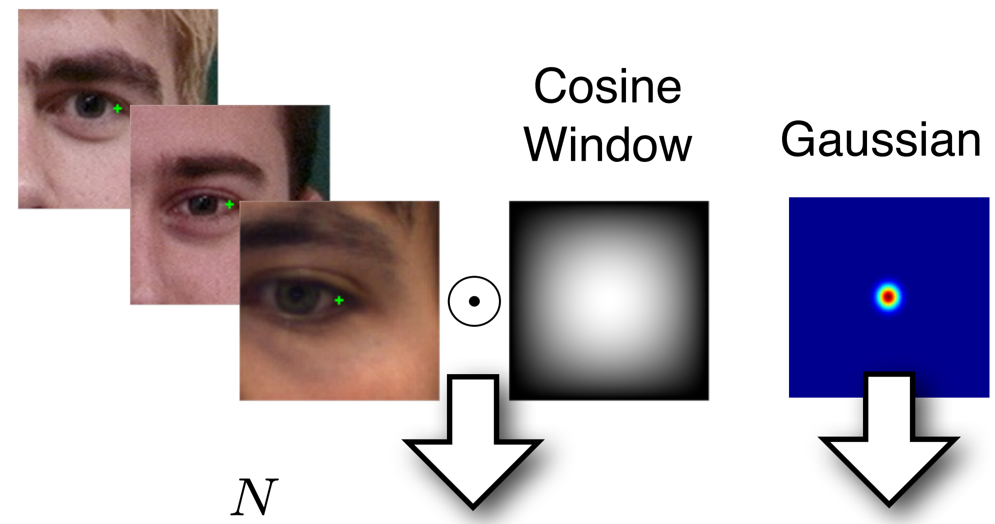
Local Landmark Detectors - MOSSE Filters



$\mathcal{F}^{-1}\{\mathbf{H}_i^*\}$ $\mathcal{D}_i^{\mathrm{MOSSE}}(\mathbf{I}(\mathbf{y}_i))$

- Correlation in Fourier Domain

$$\mathbf{G} = \mathcal{F}\{\mathbf{I}\} \odot \mathbf{H}^*$$



$$\min_{\mathbf{H}^*} \sum_{j=1}^{N} (\mathcal{F}\{\mathbf{I}_j\} \odot \mathbf{H}^* - \mathbf{G}_j)^2$$

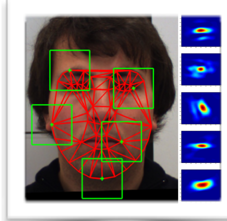
MOSSE Filter

$$\mathbf{H}^* = \frac{\sum_{j=1}^{N} \mathbf{G}_j \odot \mathcal{F}\{\mathbf{I}_j\}^*}{\sum_{j=1}^{N} \mathcal{F}\{\mathbf{I}_j\} \odot \mathcal{F}\{\mathbf{I}_j\}^*}$$

$$\mathcal{D}_i^{\mathrm{MOSSE}}(\mathbf{I}(\mathbf{y}_i)) = \mathcal{F}^{-1}\{\mathcal{F}\{\mathbf{I}(\mathbf{y}_i)\} \odot \mathbf{H}_i^*\}$$

- Visual object tracking using adaptive correlation filters
  D.Bolme, J.Beveridge, B.Draper, Y.Lui, CVPR 2010

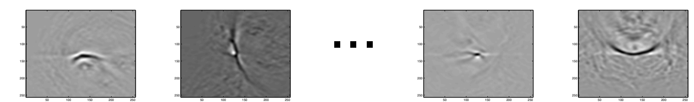
KDE Landmark Updates



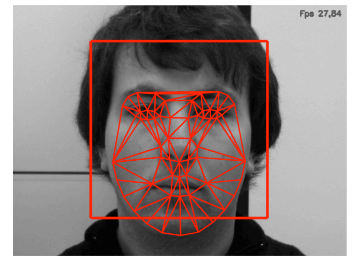
BASM - The Algorithm

Precompute: PDM: $\mathbf{s}_0, \Phi, \Psi, \Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$

MOSSE Filters: $\mathbf{H}_i^* \quad {}_{i=1,\ldots,v}$



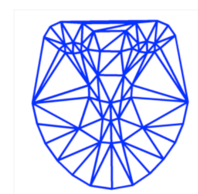
Initial estimate $(\mathbf{b}_0, \Sigma_0), (\mathbf{q}_0, \Sigma_0^q)$



`for k=1:1:MaxIterations`

Warp Image to the base mesh, using the current pose parameters

Generate current shape $\mathbf{s} = \mathcal{S}\left(\mathbf{s}_0 + \Phi \mathbf{b}_k; \mathbf{q}_k\right)$

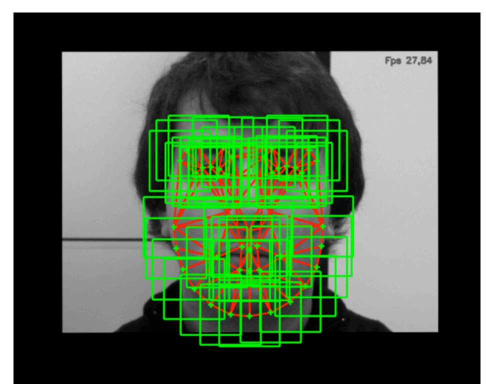


`for i=1:1:Landmarks`

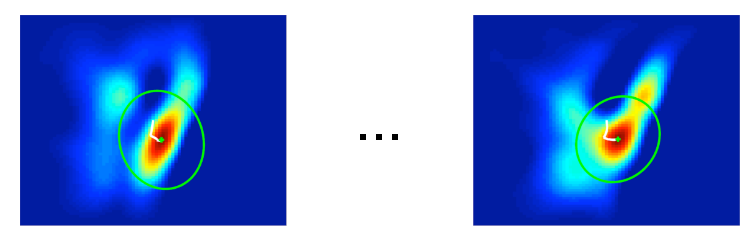
Evaluate detectors response

Find the likelihood parameters $\mathbf{y}_i, \Sigma_{\mathbf{y}_i}$

`end`



WPR, **GR** or **KDE** Local Strategy



Estimate the shape/pose parameters:

Update the parameters of Normal Inv-Wishart distribution $\rightarrow \upsilon_k, \kappa_k, \theta_k, \varLambda_k$

Expectation of the prior shape parameters $\longrightarrow \mu_{\mathbf{b}_k} = \theta_k, \quad \Sigma_{\mathbf{b}_k} = (\upsilon_k - n - 1)^{-1} \varLambda_k$

Evaluate the **global** shape parameters and the covariance $\rightarrow \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k$

`end`

Hierarchical Search (KDE-H)

- When response maps are approximated by KDE representations.

Mean-Shift Landmark Update

$$\mathbf{y}_i^{\mathrm{KDE}(\tau+1)} \leftarrow \frac{\sum_{\mathbf{z}_i \in \mathbf{\Omega}_{\mathbf{y}_i^c}} \mathbf{z}_i \; p_i(\mathbf{z}_i) \; \mathcal{N}(\mathbf{y}_i^{\mathrm{KDE}(\tau)} | \mathbf{z}_i, \sigma_{h_j}^2 \mathbf{I}_2)}{\sum_{\mathbf{z}_i \in \mathbf{\Omega}_{\mathbf{y}_i^c}} p_i(\mathbf{z}_i) \; \mathcal{N}(\mathbf{y}_i^{\mathrm{KDE}(\tau)} | \mathbf{z}_i, \sigma_{h_j}^2 \mathbf{I}_2)}$$

Bandwidth schedule

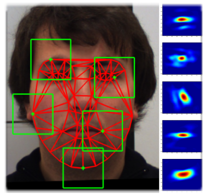
$$\sigma_h^2 = (15, 10, 5, 2)$$

Standard Search

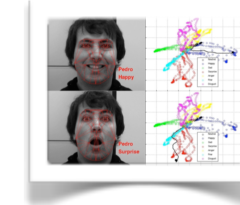
```
for k=1:1:MaxIterations
    for i=1:1:v (LandMarks)
        Evaluate de Detectors Response
        Mean-Shift Landmark Update  σ_h^2 = (15, 10, 5, 2)
    end
    Global Optimization
end
```

Hierarchical Search

```
for k=1:1:MaxIterations
    for σ_h^2 = (15, 10, 5, 2)
        for i=1:1:v (LandMarks)
            Evaluate de Detectors Response
            Mean-Shift Landmark Update σ_{h_j}^2
        end
        Global Optimization
    end
end
```



Tracking Performance - FNET Talking Face



The Recognition Approach