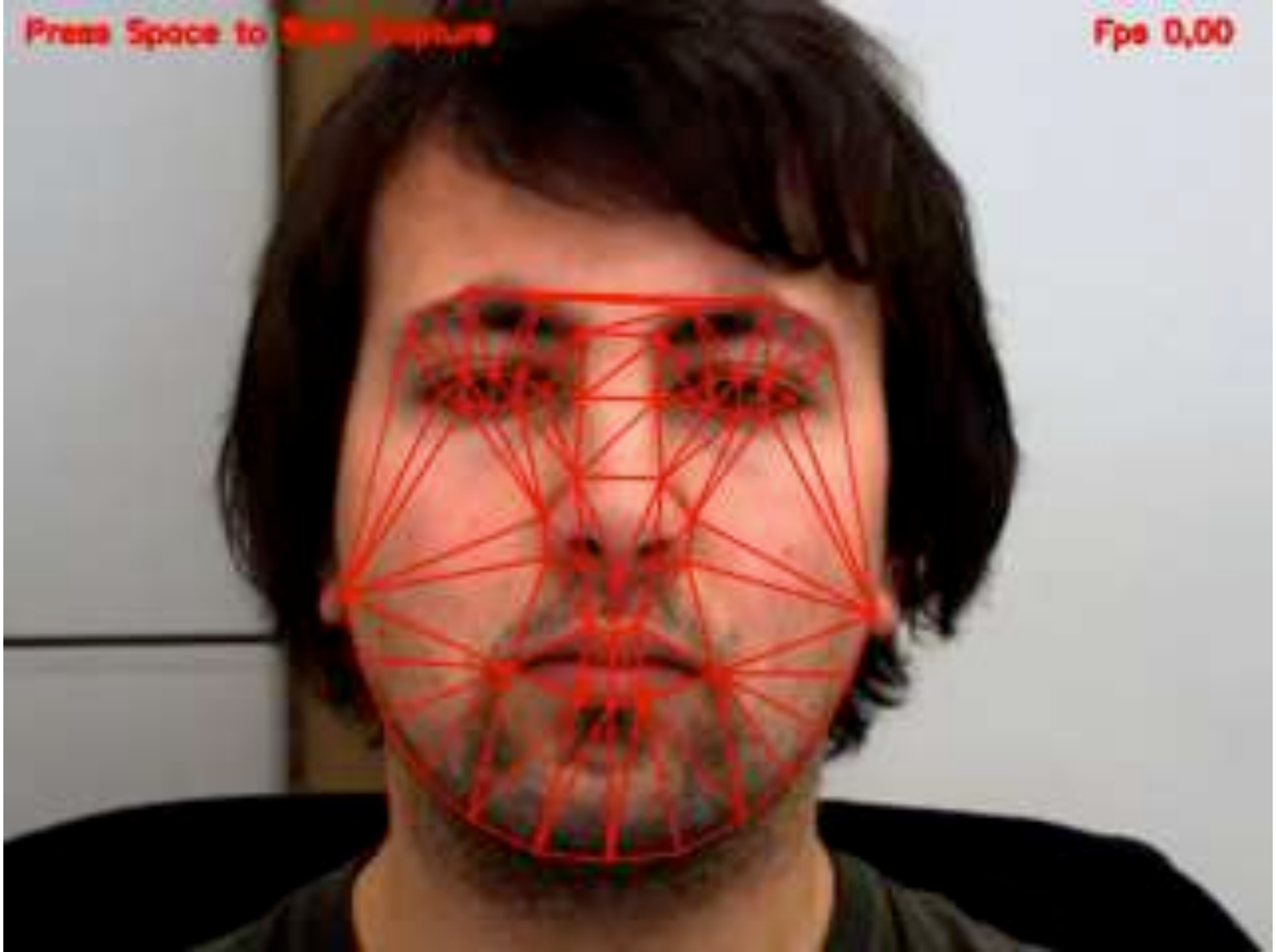
Deformable Face Models in ‘the Wild’

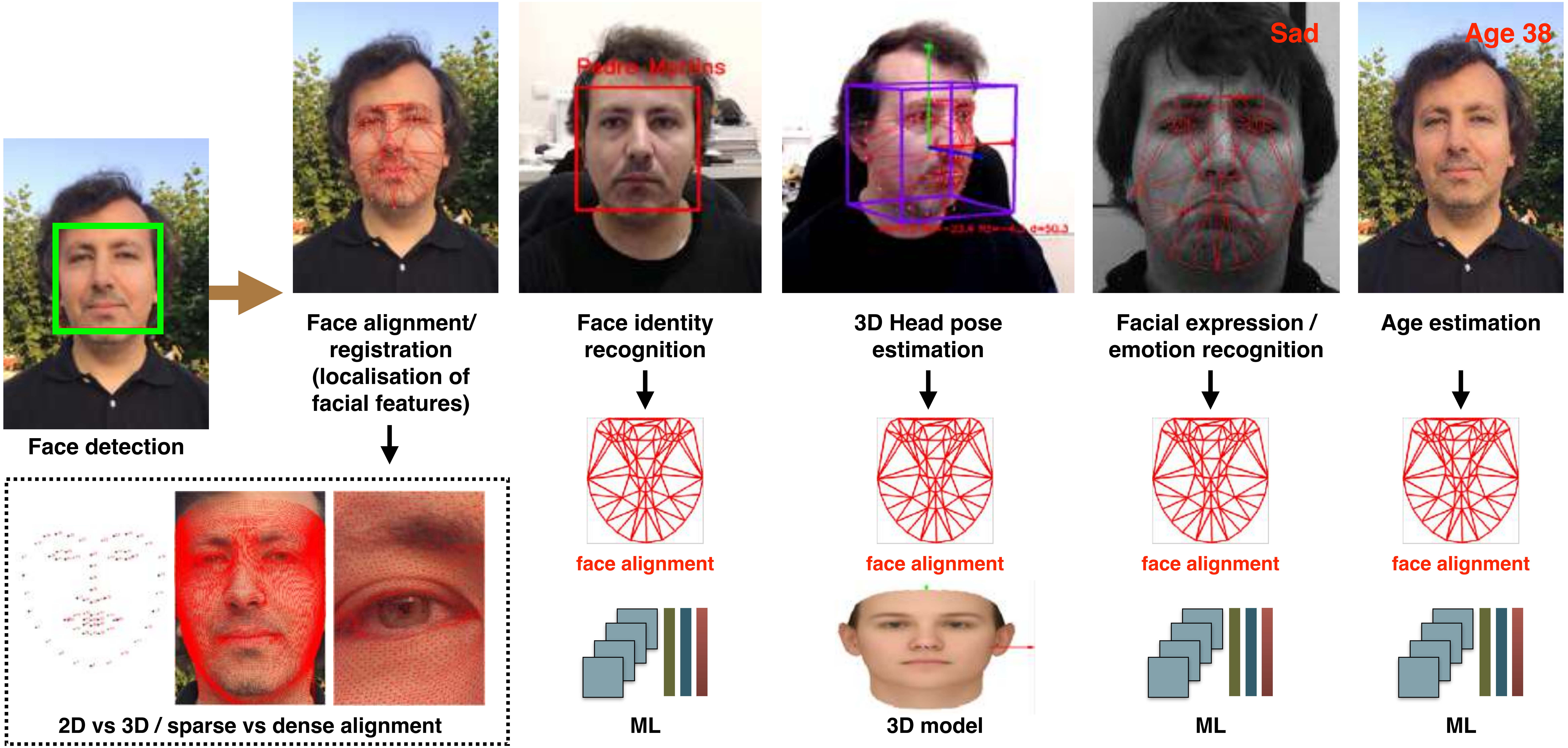
Pedro Martins
https://www.isr.uc.pt/~pedromartins
pedromartins@isr.uc.pt

Computer Vision Lab.
Institute of Systems and Robotics (ISR)
University of Coimbra, Portugal

1st Open Day Workshop @ Institute of Systems and Robotics, Coimbra, 14 July 2022


Press Space to
Fps 0,00

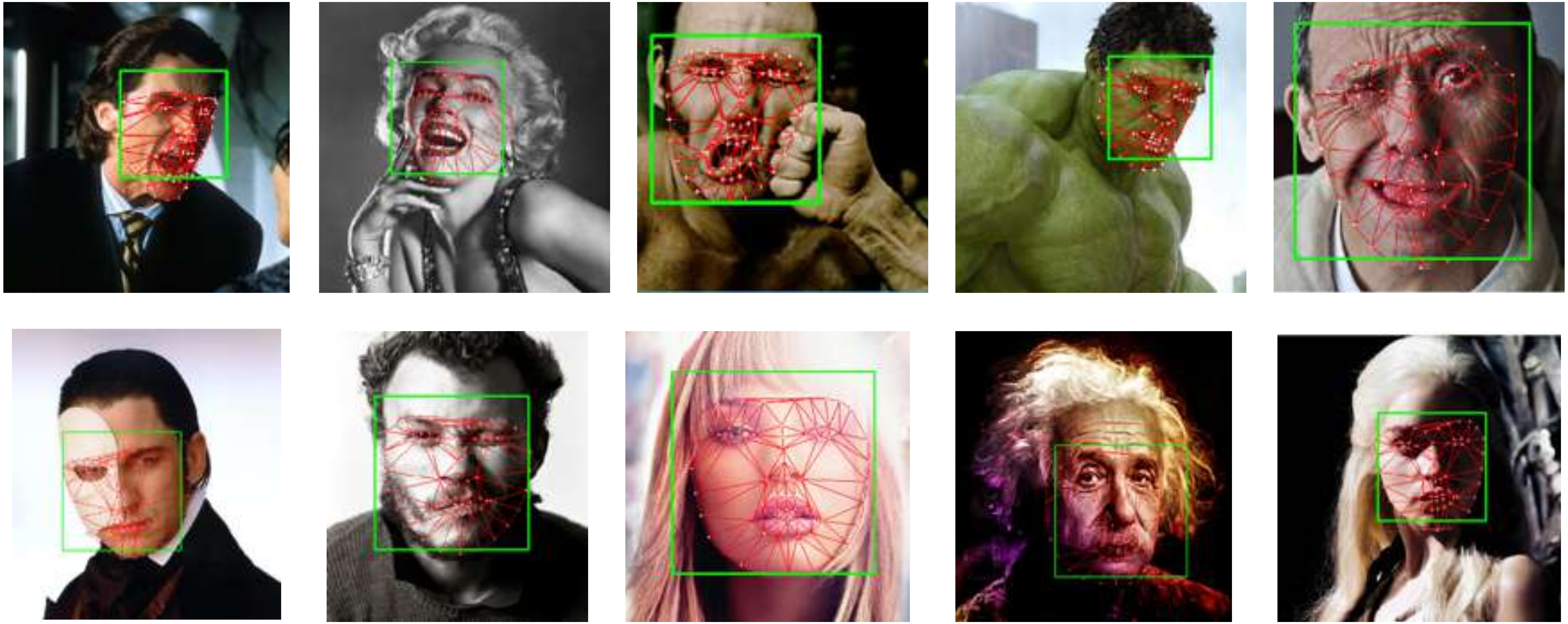
Introduction - Face Tasks



Face Alignment - How Hard Can It Be?

- **Must be able to deal with variations in identity, facial expression, pose, occlusion, illumination, camera parameters, …**

Face Alignment - How Hard Can It Be?

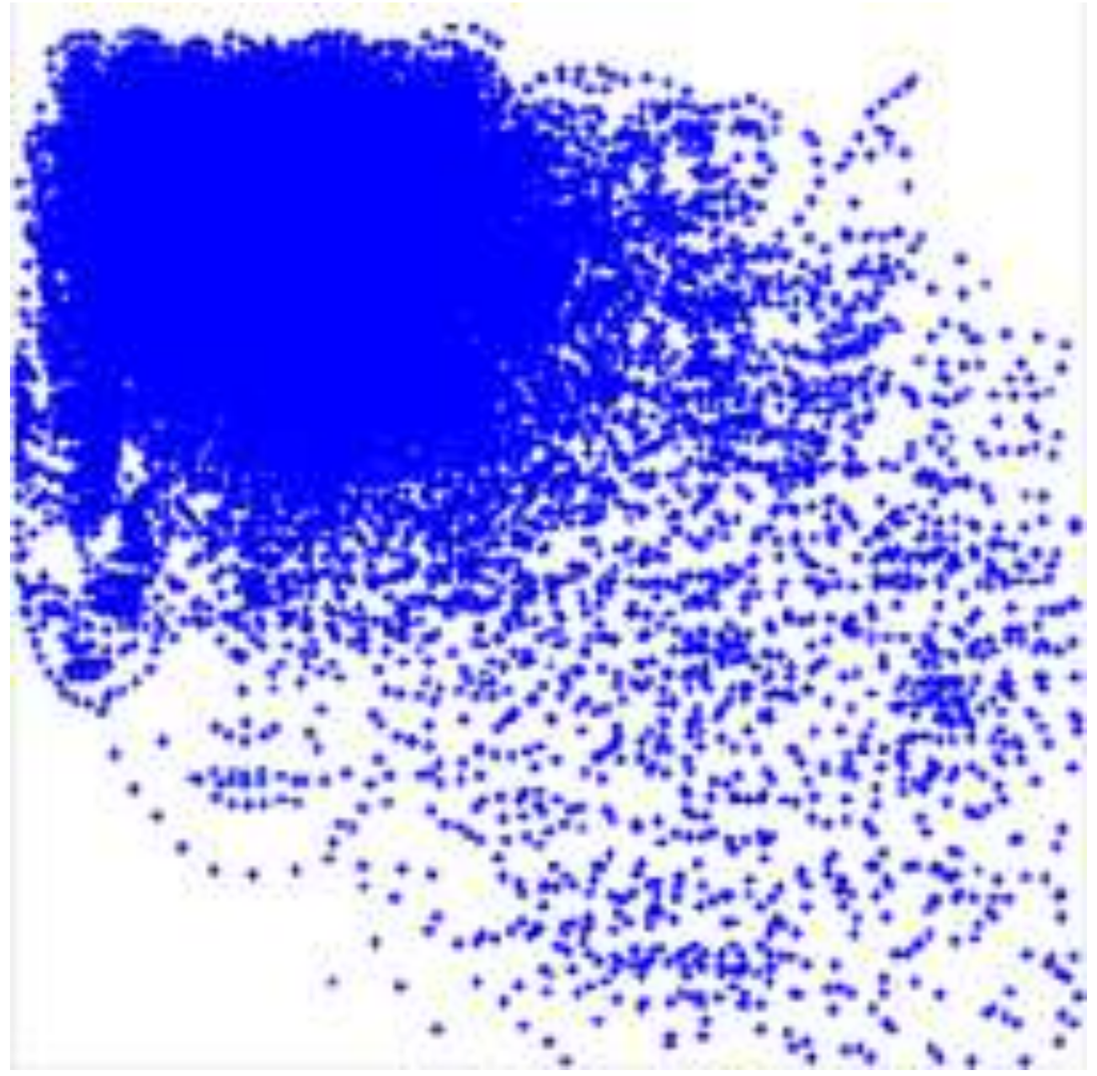


- **Must be able to deal with variations in identity, facial expression, pose, occlusion, illumination, camera parameters, …**

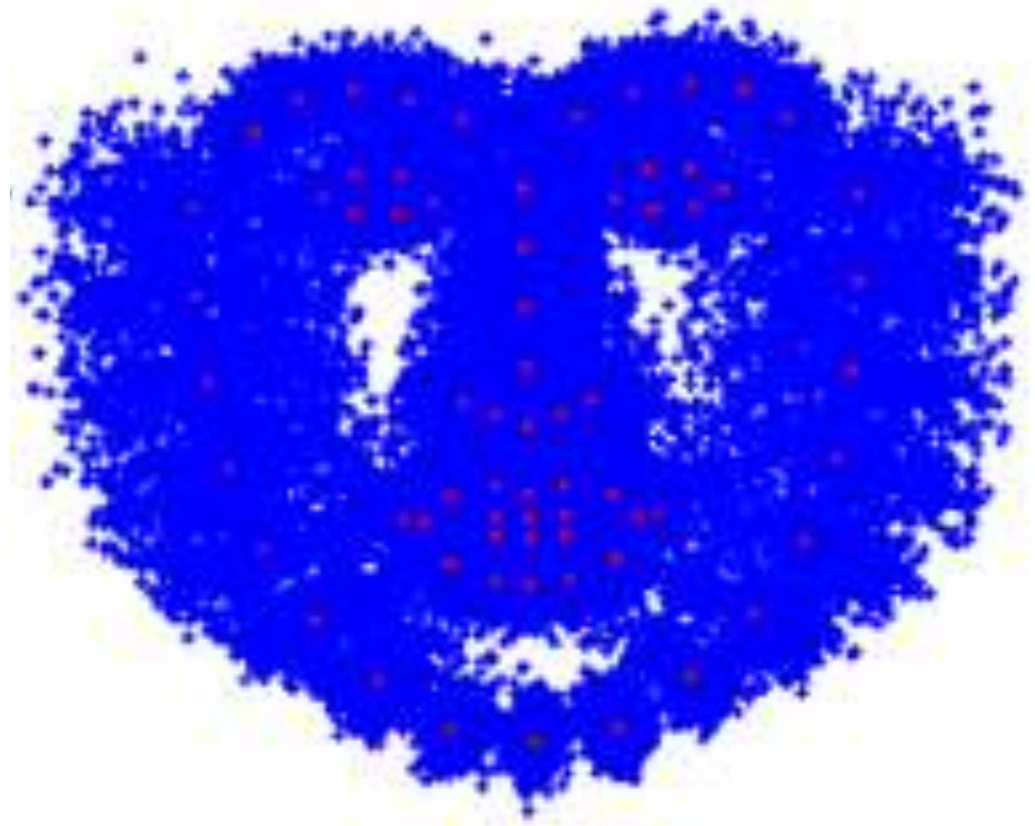
Linear Shape Model

'In the Wild' Image Database

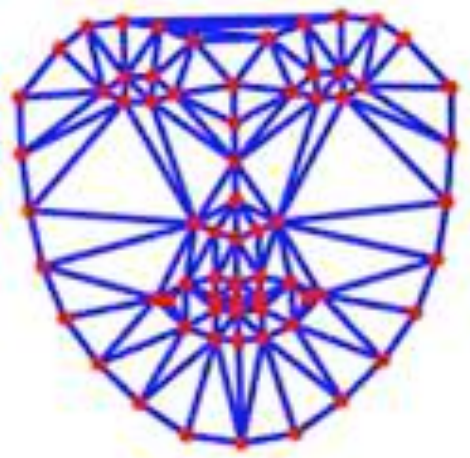
RAW Shape Data



Procrustes Alignment



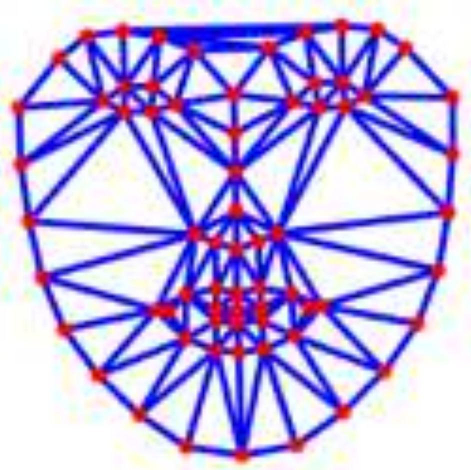
Shape Model



$$\mathcal{B}(\mathbf{s};\mathbf{b}) = \mathbf{s}_0 + \sum_{i=1}^{n} \phi_i b_i$$

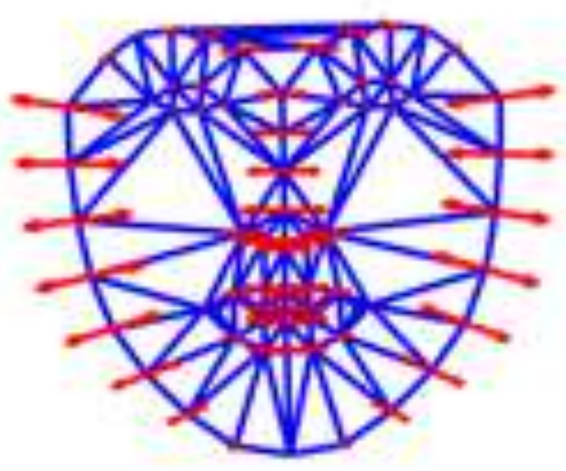
shape parameters

Mean Shape

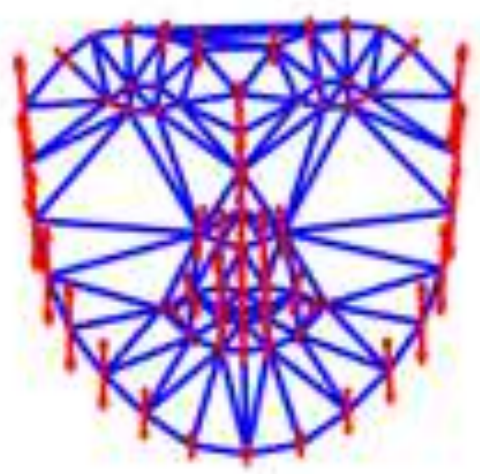


$\mathbf{s}_0$

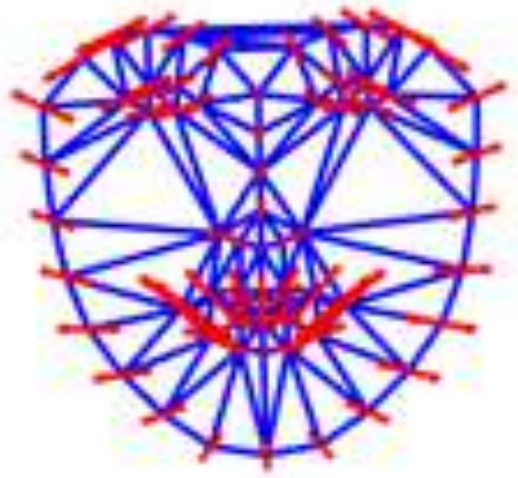
Shape Basis



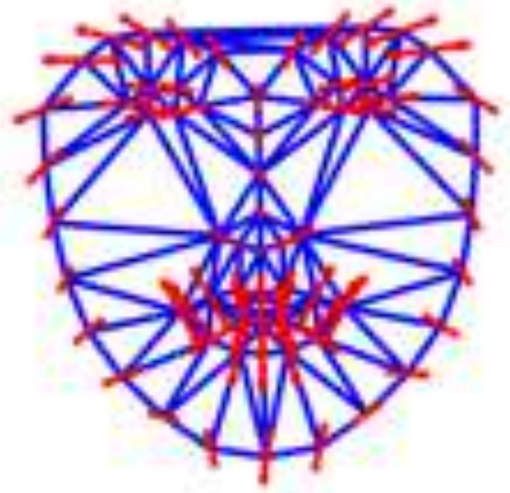
ϕ_1



ϕ_2



ϕ_3



ϕ_4

Similarity Transform

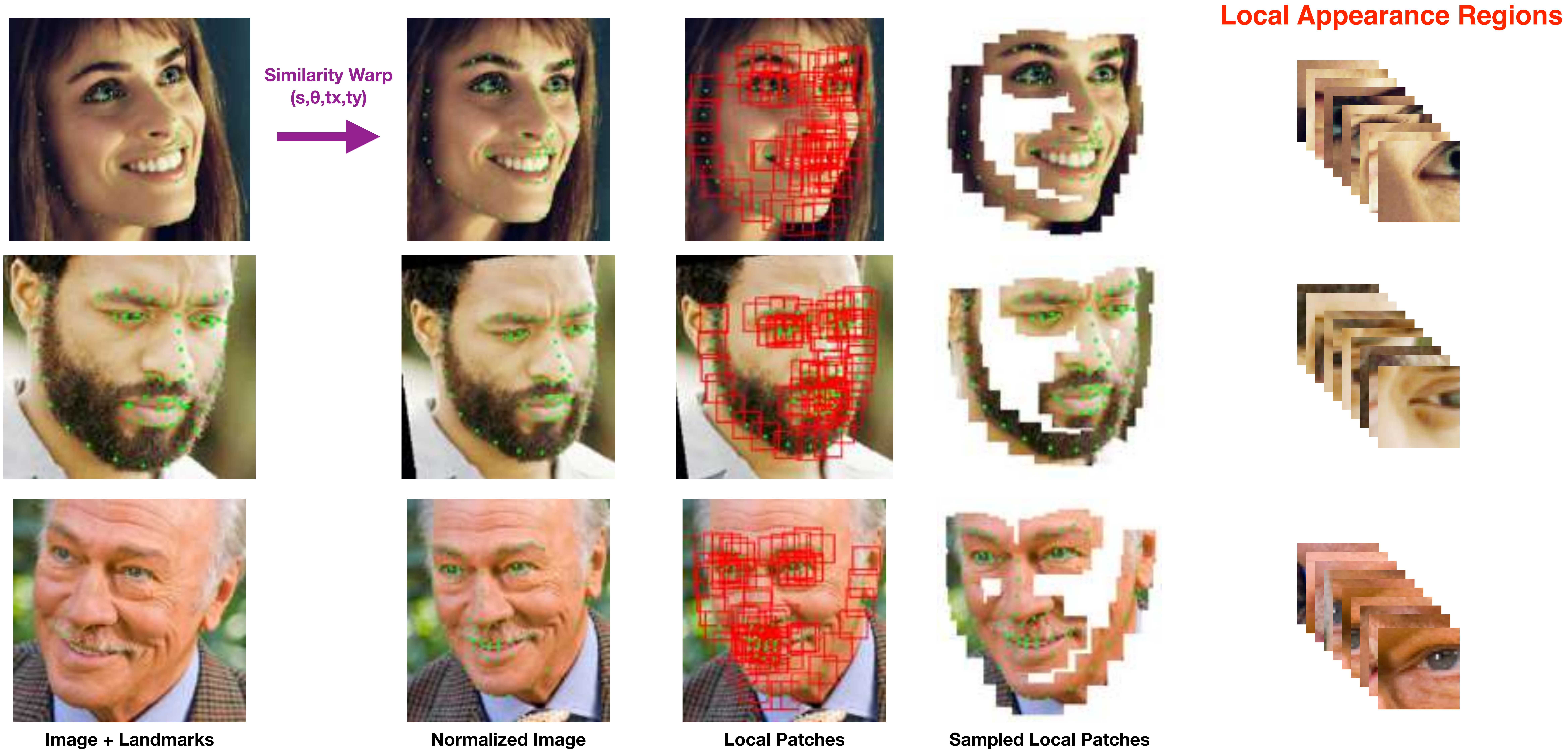
$$\mathcal{S}(\mathbf{s};\mathbf{q}) = \mathbf{s} + \sum_{j=1}^{4} \psi_j q_j$$

pose parameters

Full Shape Model

$$\mathbf{s} = \mathcal{S}(\mathcal{B}(\mathbf{b});\mathbf{q})$$

Local (Patch) Appearance Regions



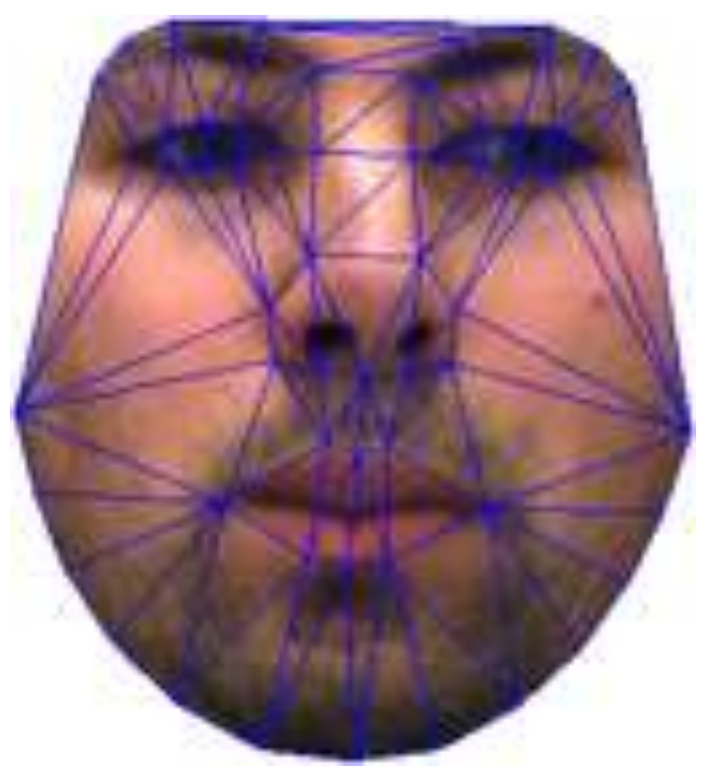
Piecewise Affine Warp

$$\mathbf{W}(\mathbf{x},\mathbf{p}) = \mathbf{x}_i + \alpha\left(\mathbf{x}_j - \mathbf{x}_i\right) + \beta\left(\mathbf{x}_k - \mathbf{x}_i\right), \quad \{\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k\} \sim \mathbf{s}$$

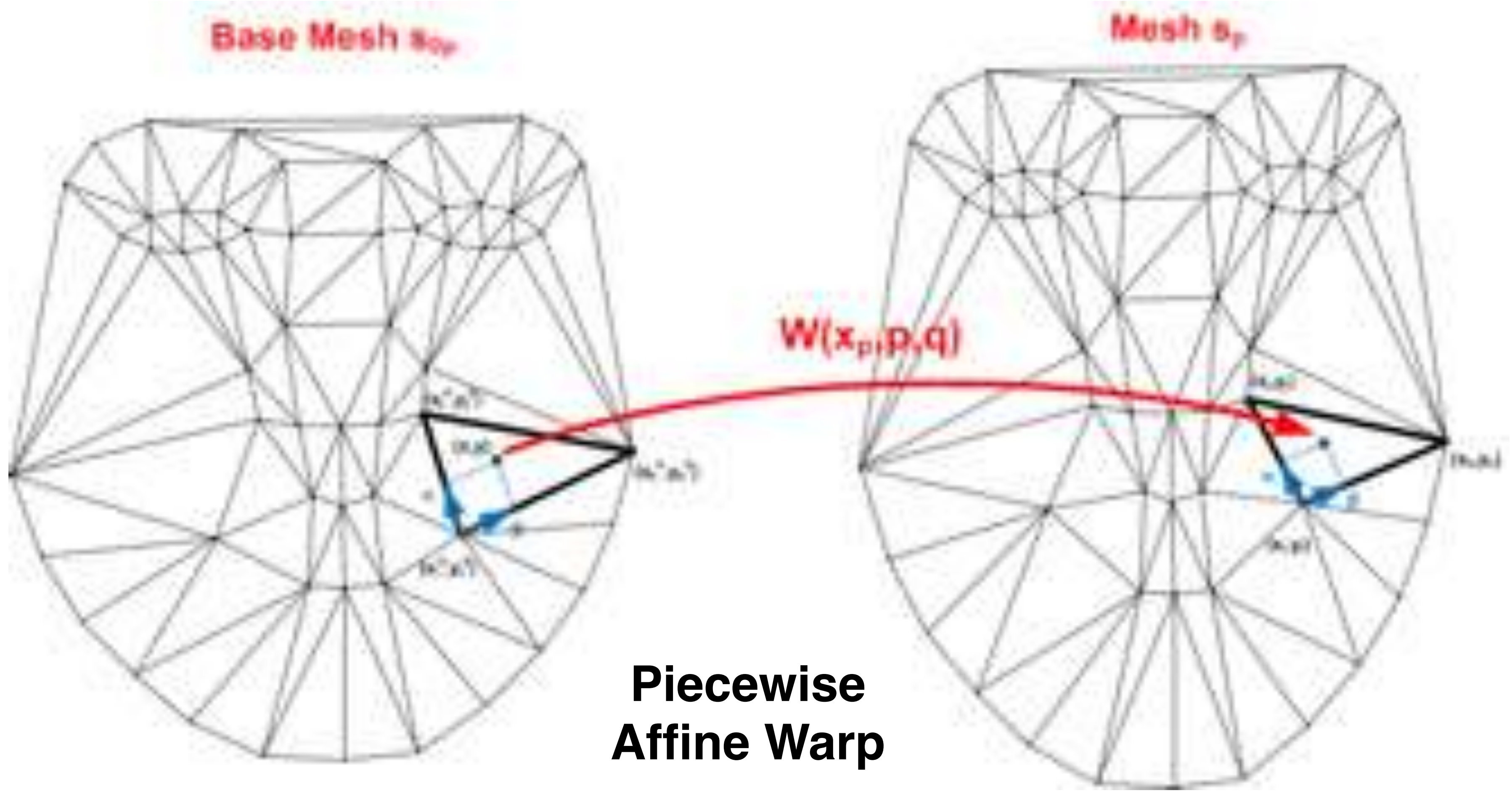
Warped Image

$$\mathbf{s} = \left(x_1 \ \ldots \ x_v, \ y_1 \ \ldots \ y_v\right)^T$$

Source Image



$\mathbf{I}(\mathbf{W}(\mathbf{x},\mathbf{p}))$

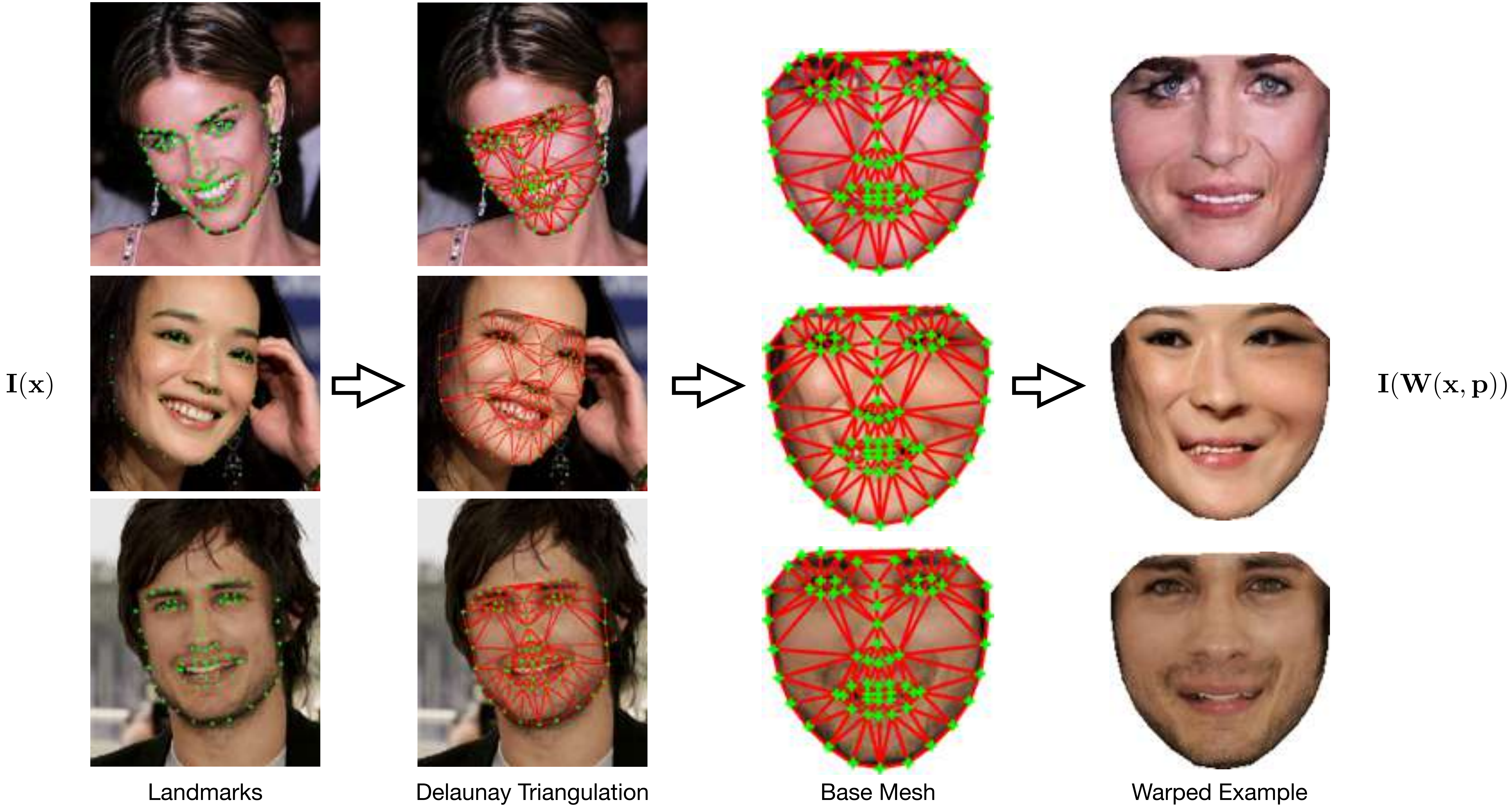


$\mathbf{W}(\mathbf{x},\mathbf{p})$

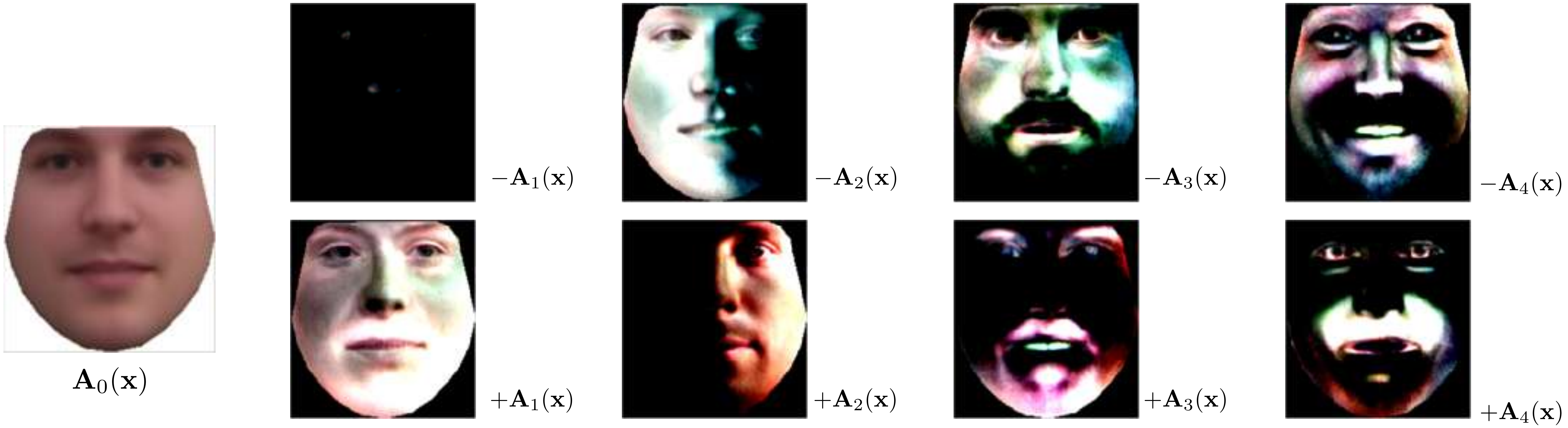
$\mathbf{I}(\mathbf{x})$

$$\mathbf{s} = \mathbf{s}_0 + \sum_{i=1}^{n} p_i \phi_i$$

Holistic Regions (Piecewise Affine Warp)

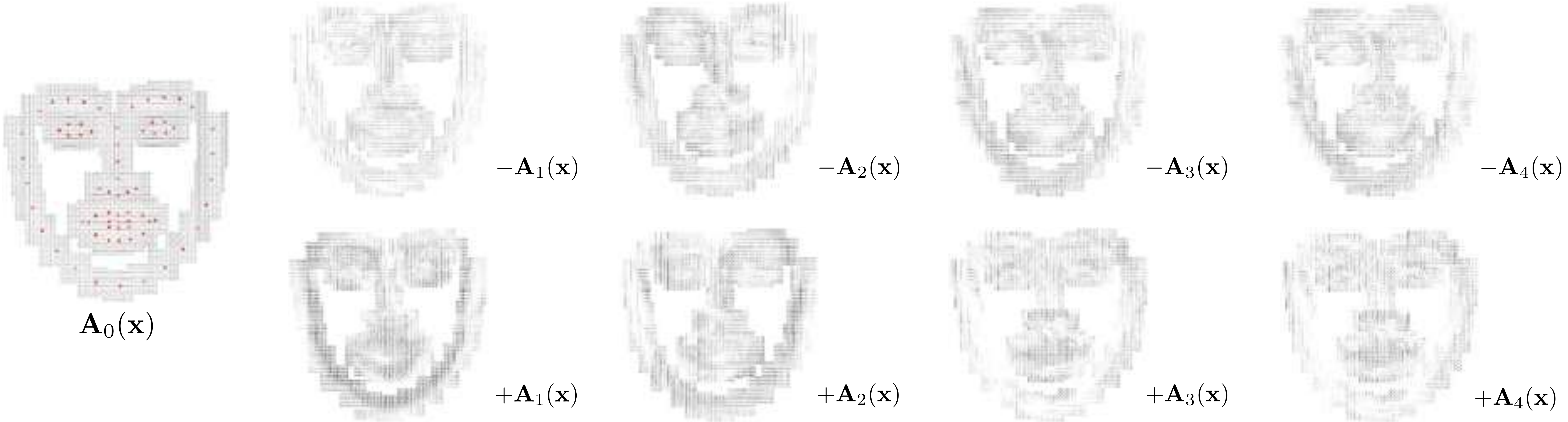


Linear Appearance Model



Appearance Model (RGB)

$$\mathcal{A}(\mathbf{x};\boldsymbol{\lambda}) = \mathbf{A}_0(\mathbf{x}) + \sum_{i=1}^{m} \mathbf{A}_i(\mathbf{x})\lambda_i$$



Appearance Model (HoG)

$$\mathcal{A}(\mathbf{x};\boldsymbol{\lambda}) = \mathbf{A}_0(\mathbf{x}) + \sum_{i=1}^{m} \mathbf{A}_i(\mathbf{x})\lambda_i$$

Active Appearance Models (AAMs) - 2D Fitting

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

Active Appearance Models (AAMs)

Fitting Goal

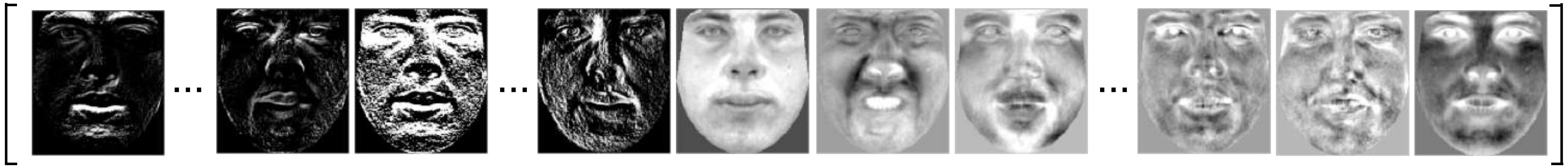
$$\arg\min_{\mathbf{p},\boldsymbol{\lambda}} \sum_{\mathbf{x}\in\mathbf{s}_0} \left[\mathbf{A}_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i \mathbf{A}_i(\mathbf{x}) - \mathbf{I}(\mathbf{W}(\mathbf{x},\mathbf{p})) \right]^2$$

Error Image

Solution

$$\begin{bmatrix} \Delta\mathbf{p} \\ \Delta\boldsymbol{\lambda} \end{bmatrix} = \mathbf{H}^{-1} \sum_{\mathbf{x}\in\mathbf{s}_0} \mathbf{J}(\mathbf{x},\mathbf{p},\boldsymbol{\lambda})^T \left(\mathbf{A}_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i \mathbf{A}_i(\mathbf{x}) - \mathbf{I}(\mathbf{W}(\mathbf{x},\mathbf{p})) \right)$$

$$\mathbf{J}(\mathbf{x},\mathbf{p},\boldsymbol{\lambda}) =$$



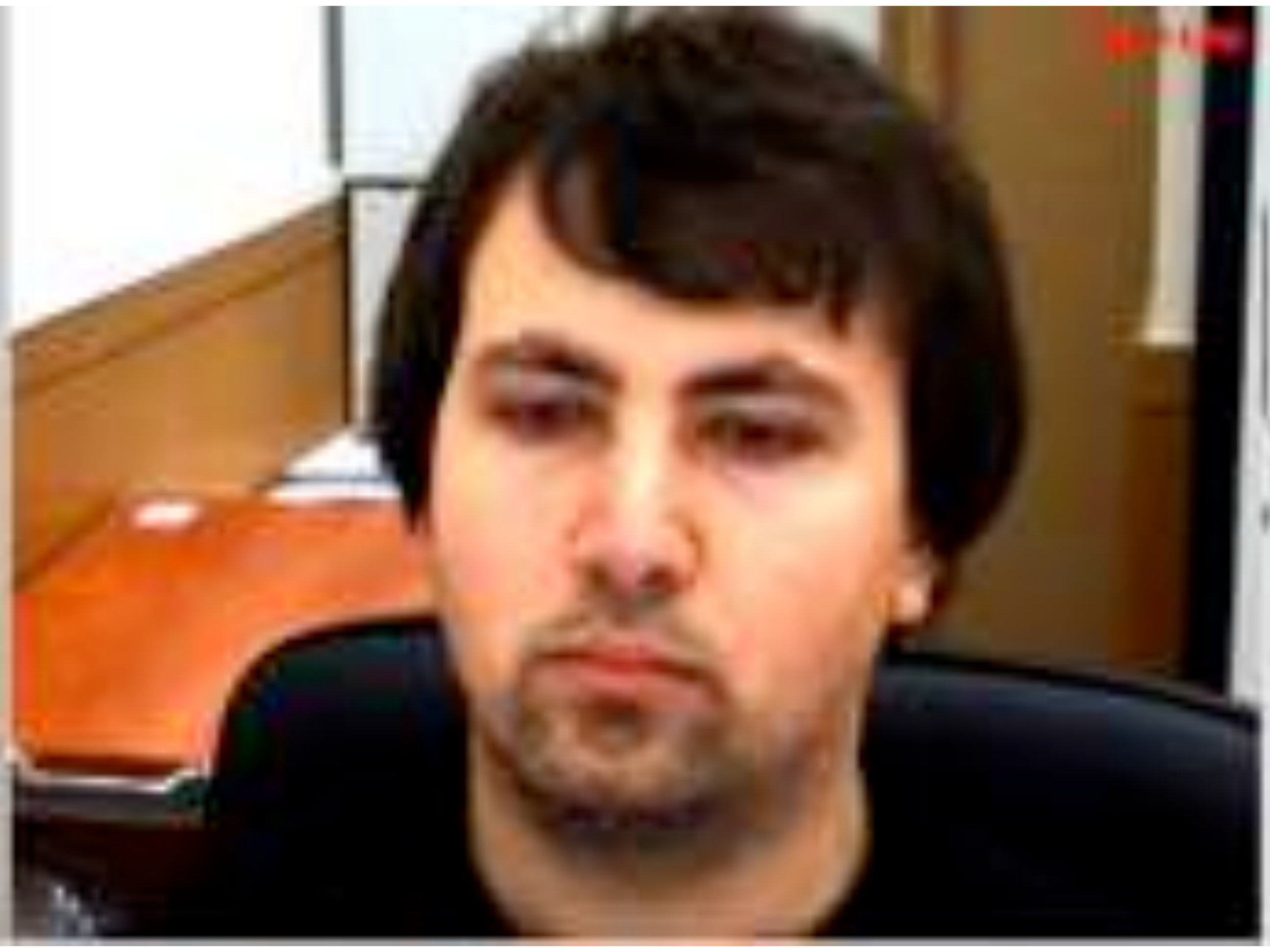
10 ~ 20 shape 'images'

4 pose 'images'

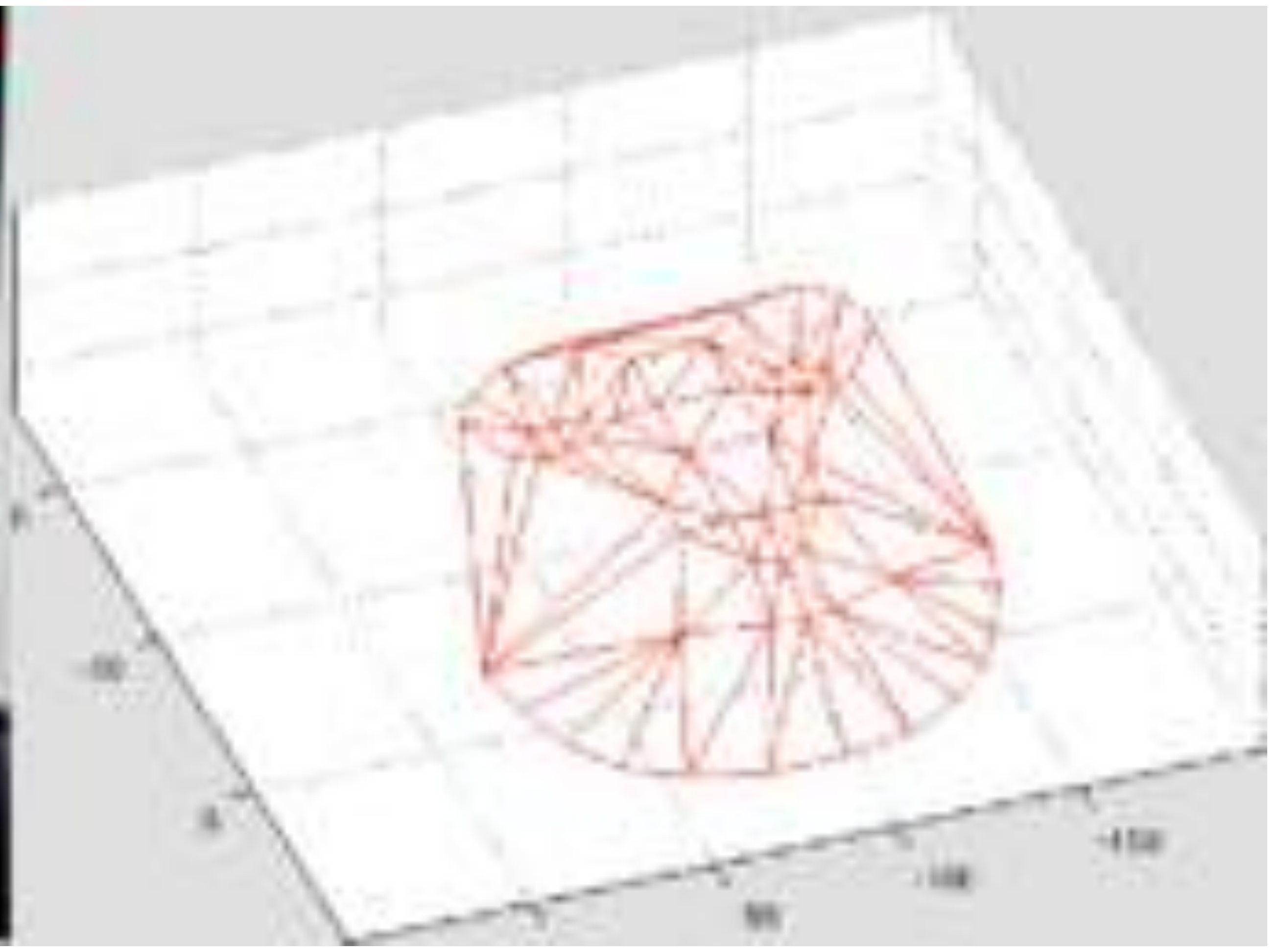
60 ~ 80 appearance 'images'

3D Shape Data

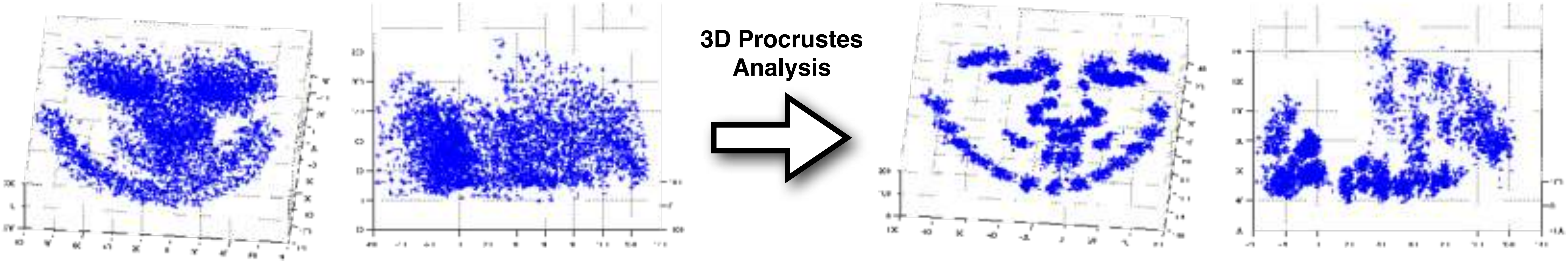
Left Camera



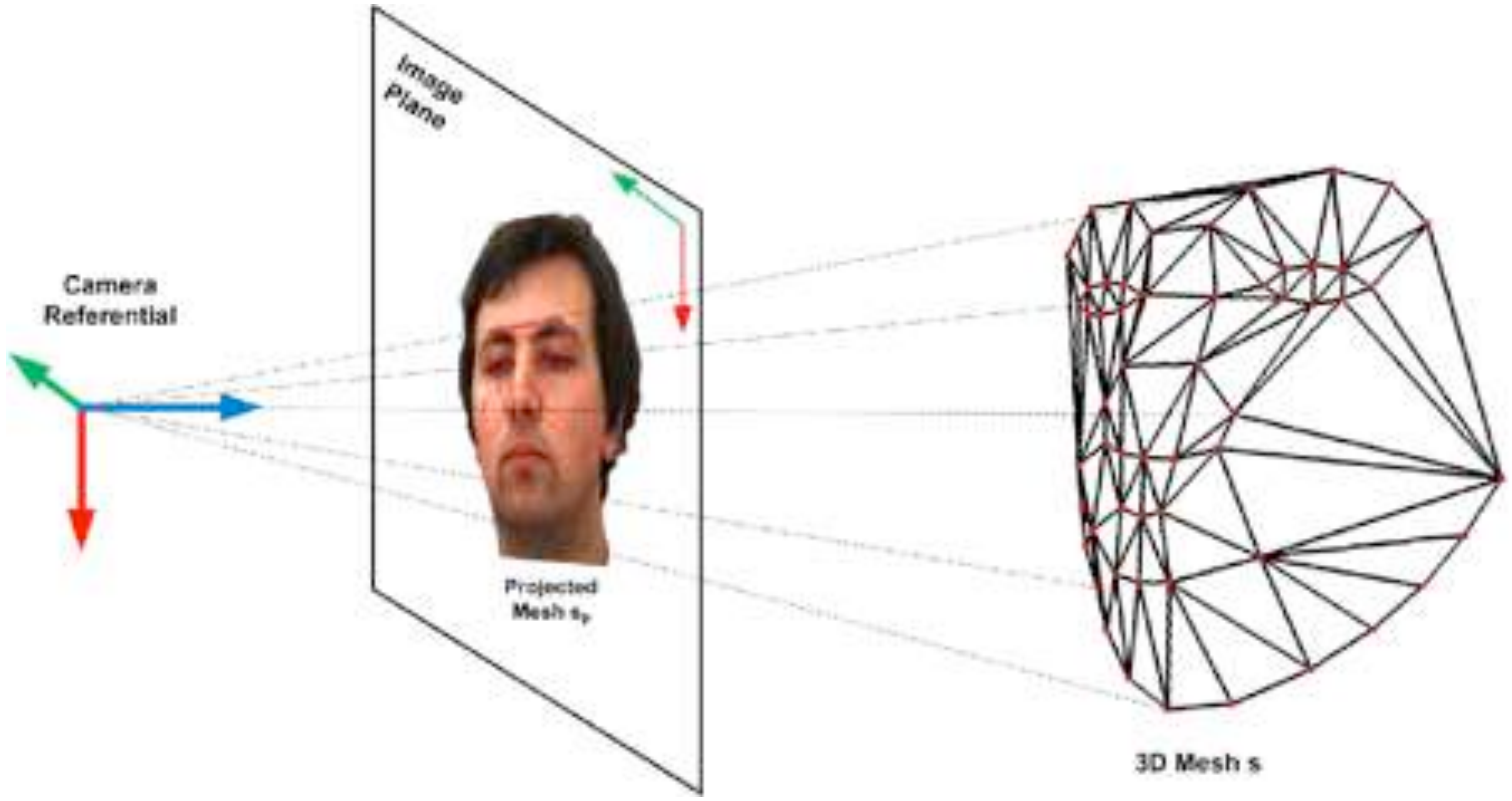
Right Camera



3D Shape



3D Shape Model - Perspective Projection



Full Perspective Projection

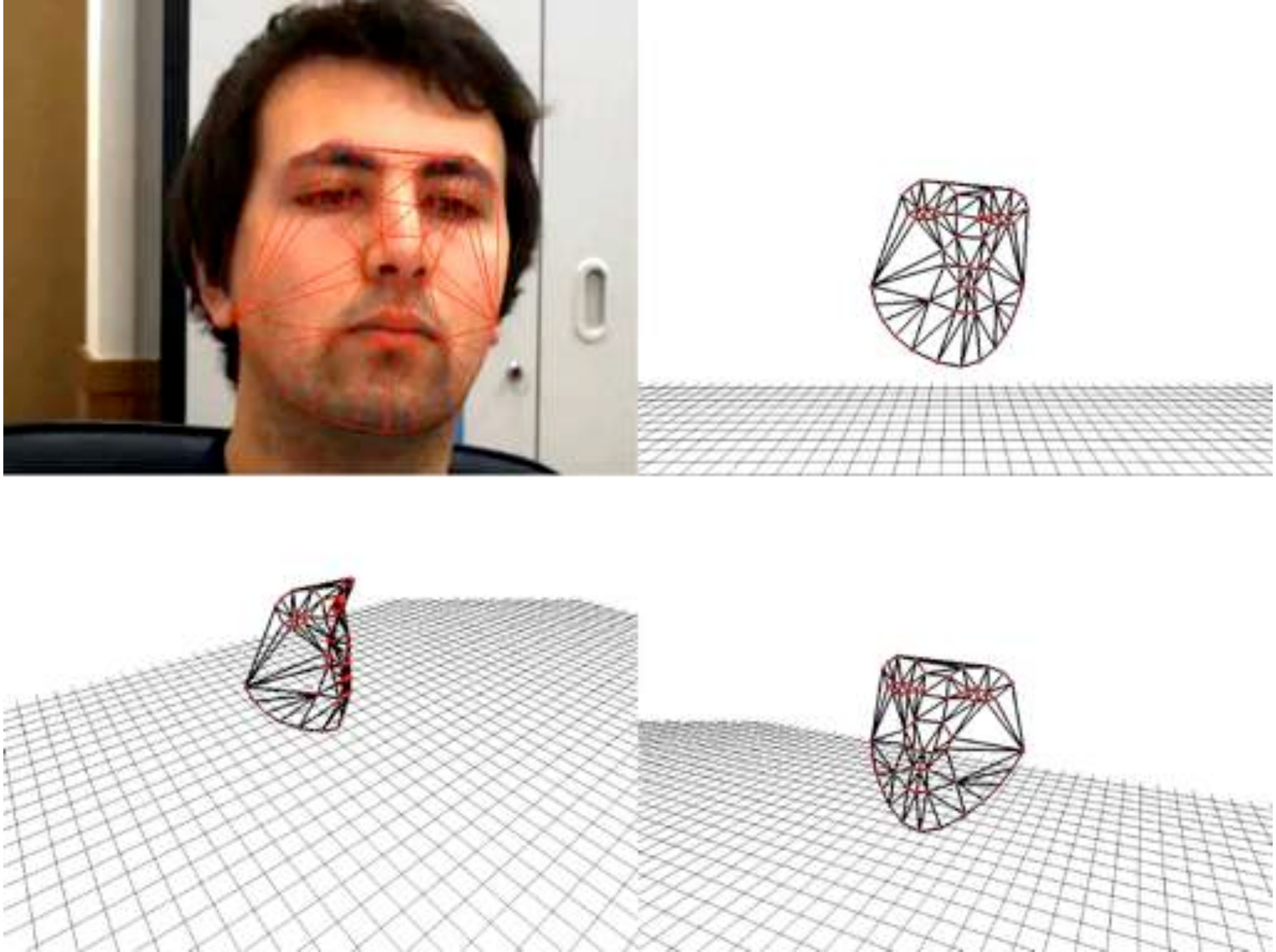
$$\begin{bmatrix} w(x_1 \cdots x_v) \\ w(y_1 \cdots y_v) \\ w \cdots w \end{bmatrix} = \underbrace{\begin{bmatrix} f_x & \alpha_s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}}_{\mathbf{K}} \left[\ \mathbf{R}_0 \mid \mathbf{t}_0\ \right] \begin{bmatrix} s^{x_1} \ldots s^{x_v} \\ s^{y_1} \ldots s^{y_v}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

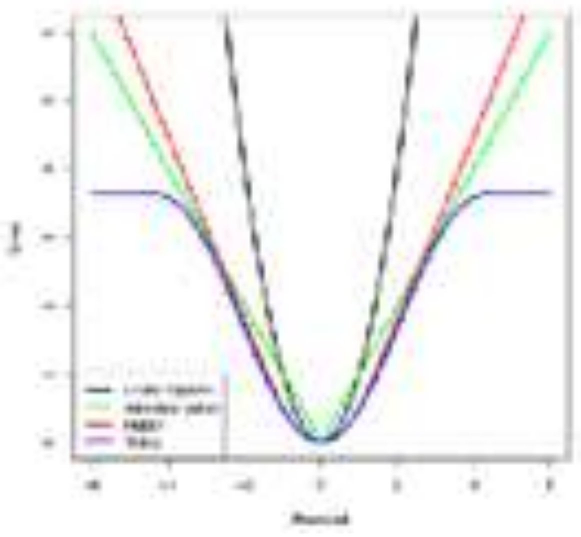
3D Point Distribution Model (PDM)

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

Pose
Parameters

Previous pose
updates





Robust 2.5D Model Fitting

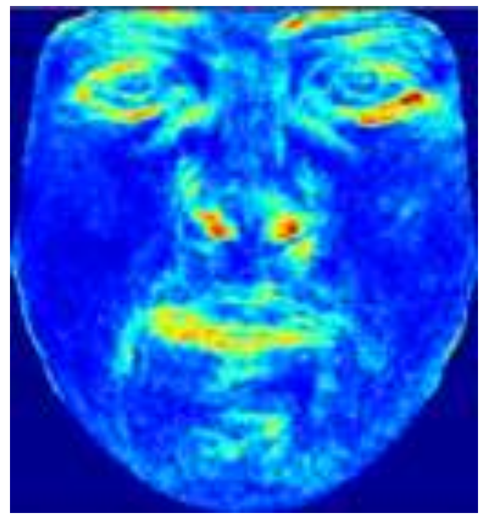
$$\arg\min_{\mathbf{p},\boldsymbol{\lambda}} \sum_{\mathbf{x}\in\mathbf{s}_0}$$

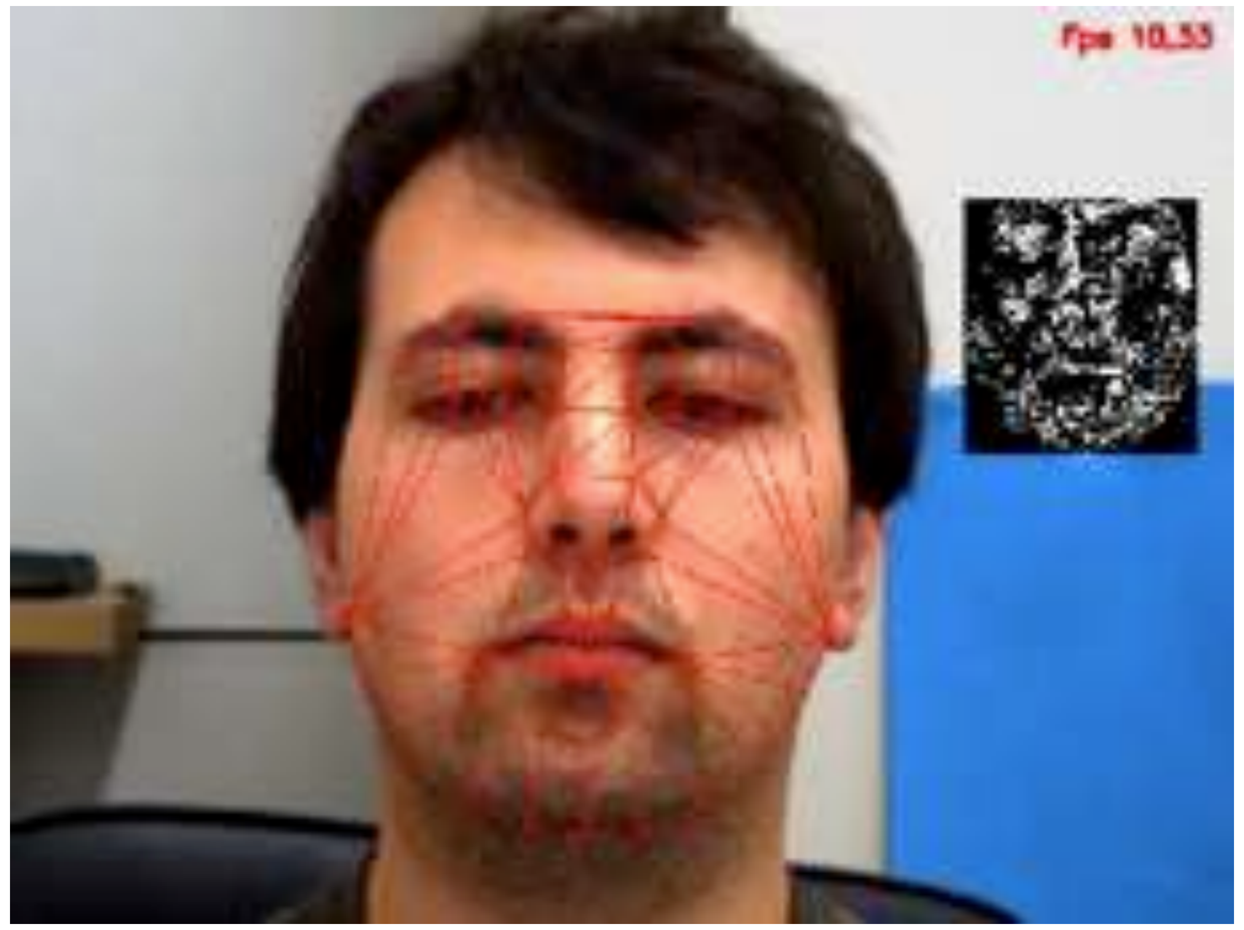
$\Big[$

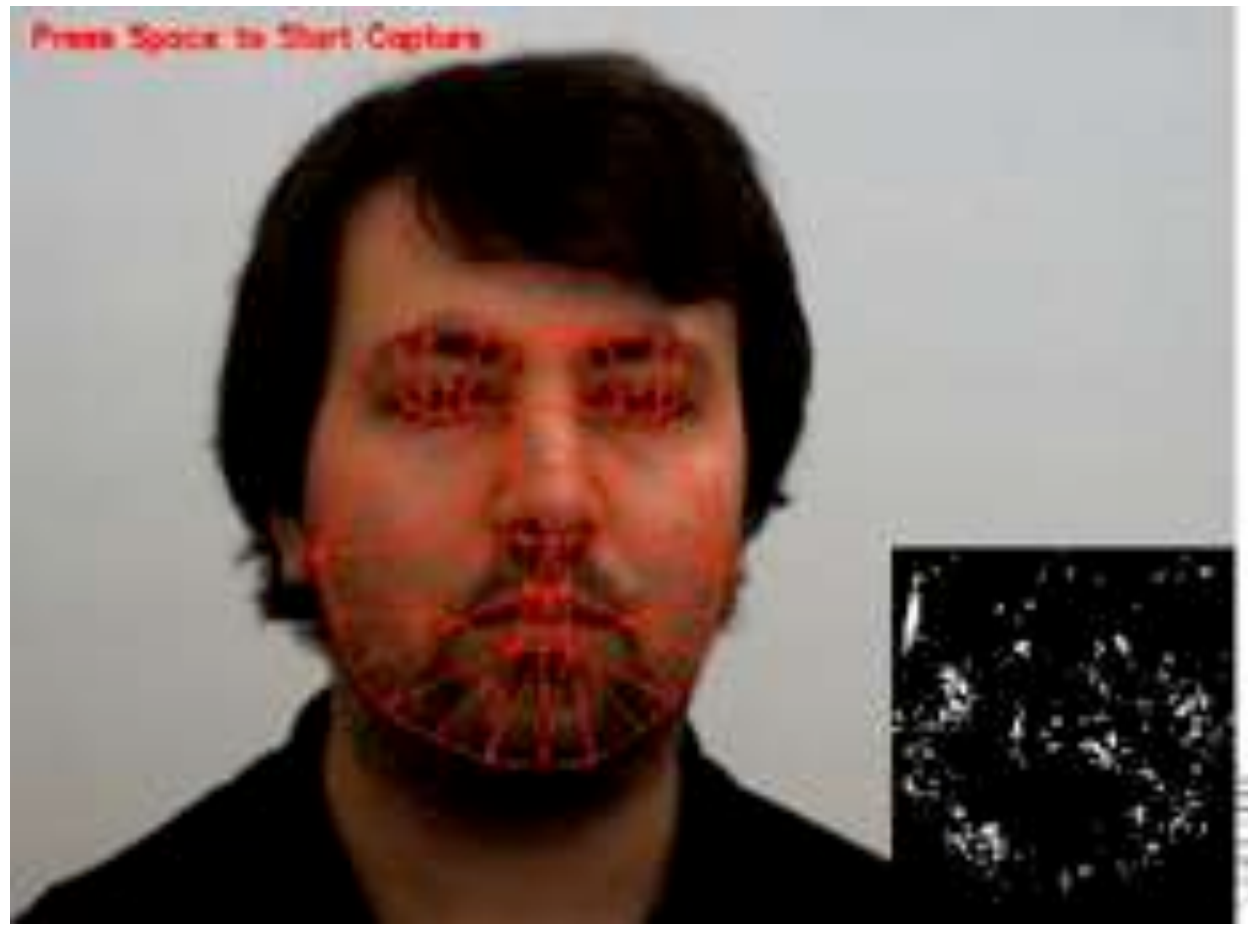
$-$

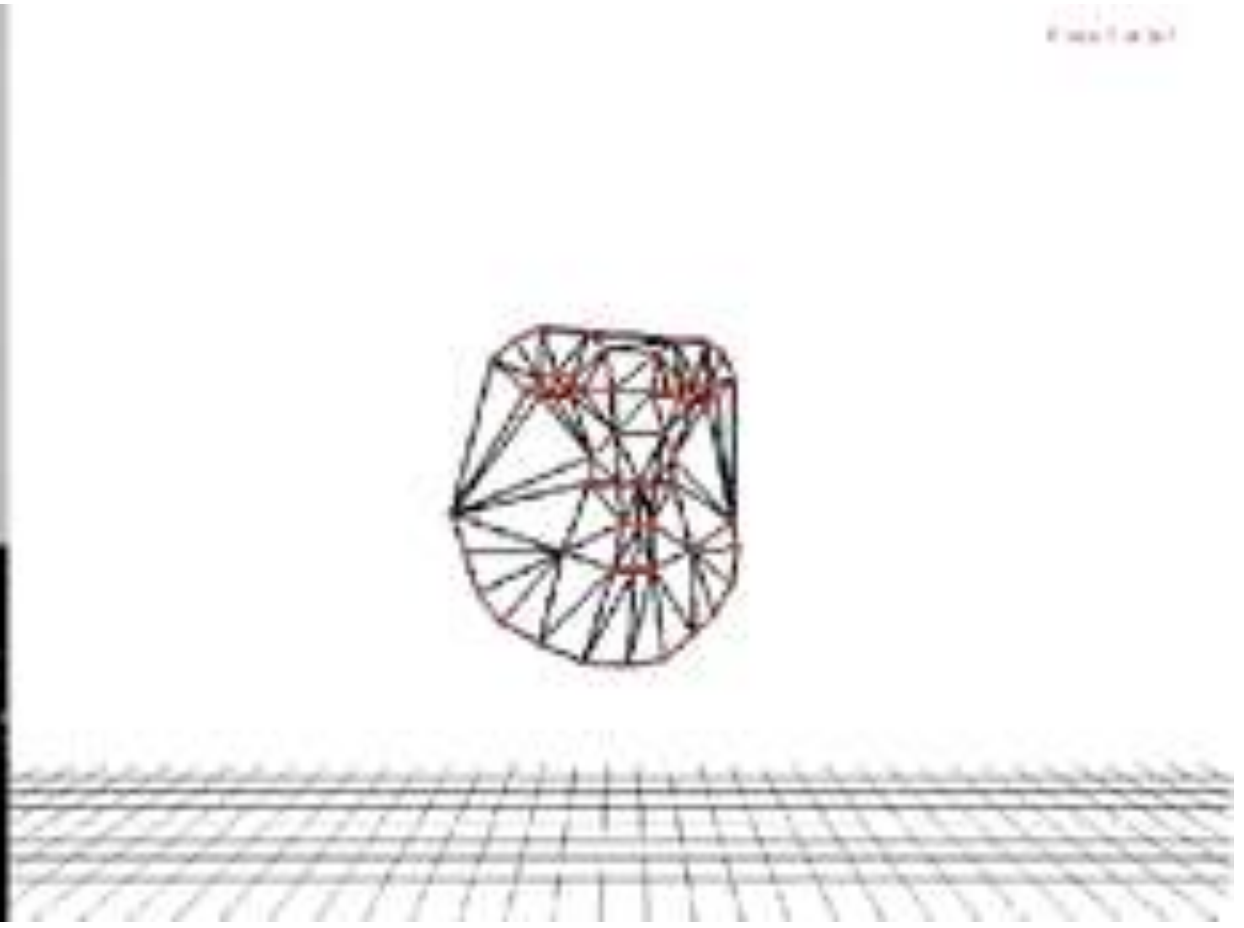
$\Big]^{\cancel{2}}$

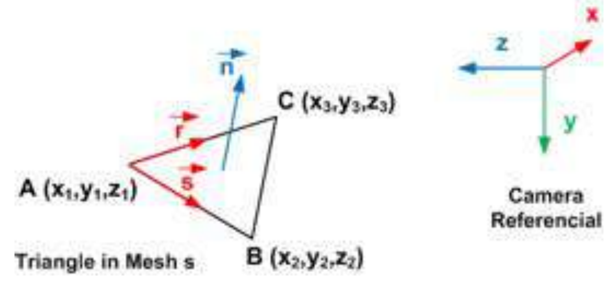
$$\arg\min_{\mathbf{p},\boldsymbol{\lambda}} \sum_{\mathbf{x}\in\mathbf{s}_0} \rho\left(\mathbf{E}(\mathbf{x}), \boldsymbol{\sigma}\right)$$





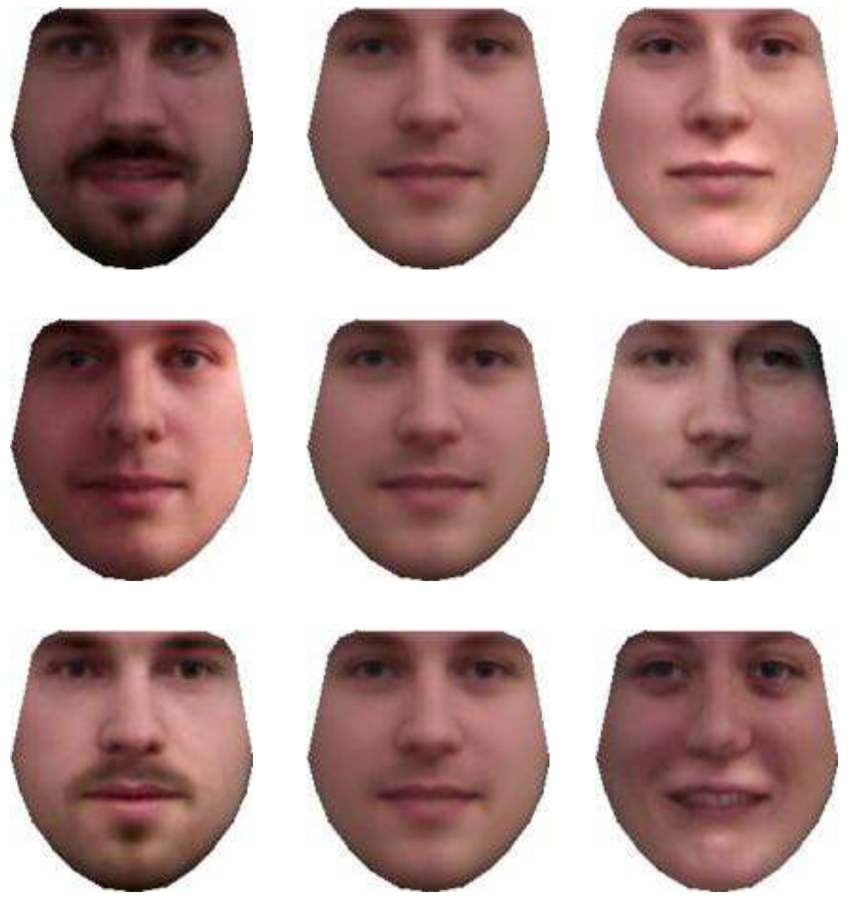


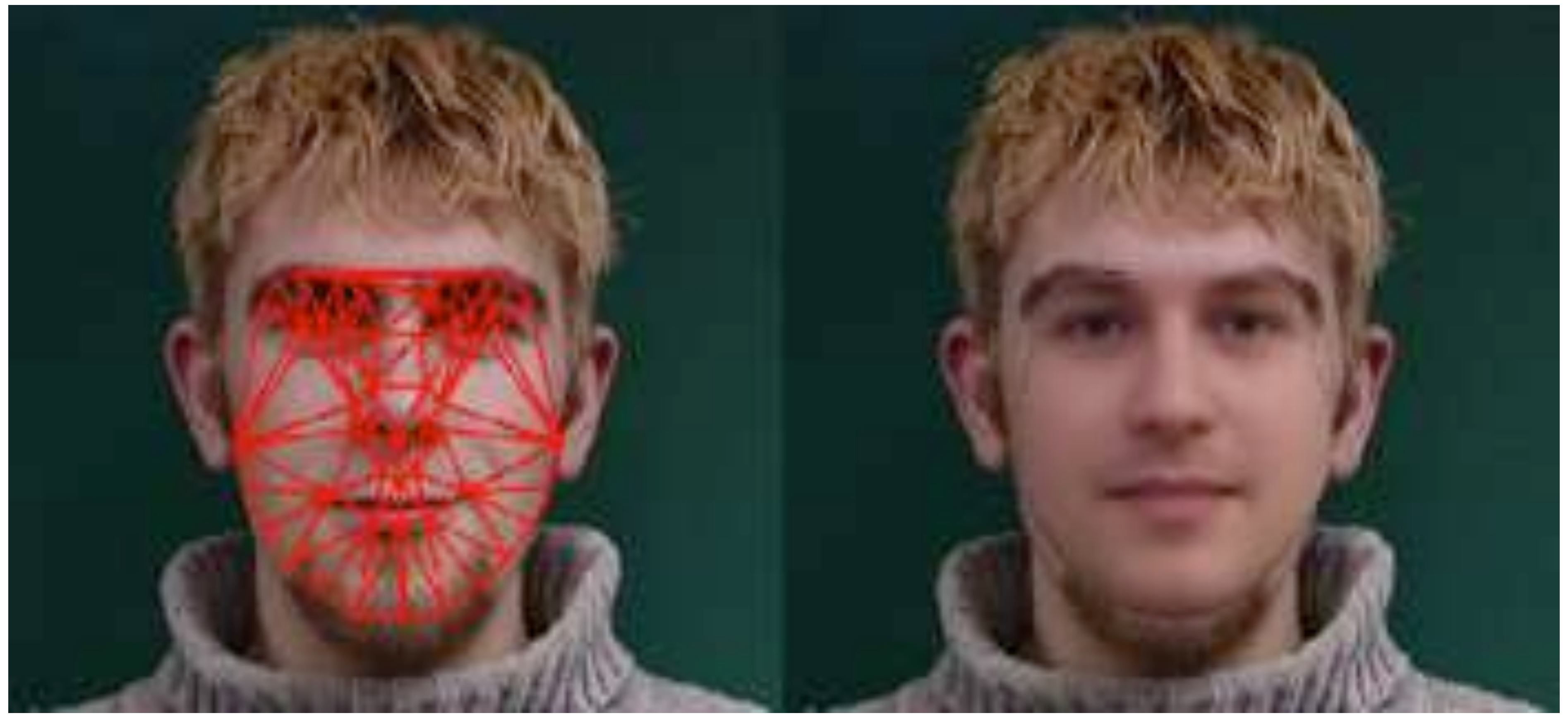




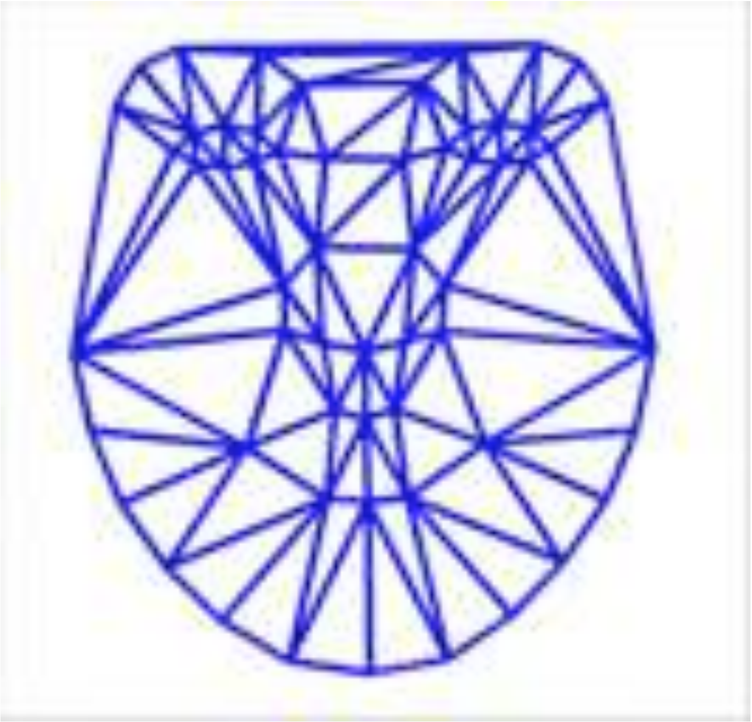
Generative vs Discriminative Face Alignment

- **Generative / Holistic Appearance Model**



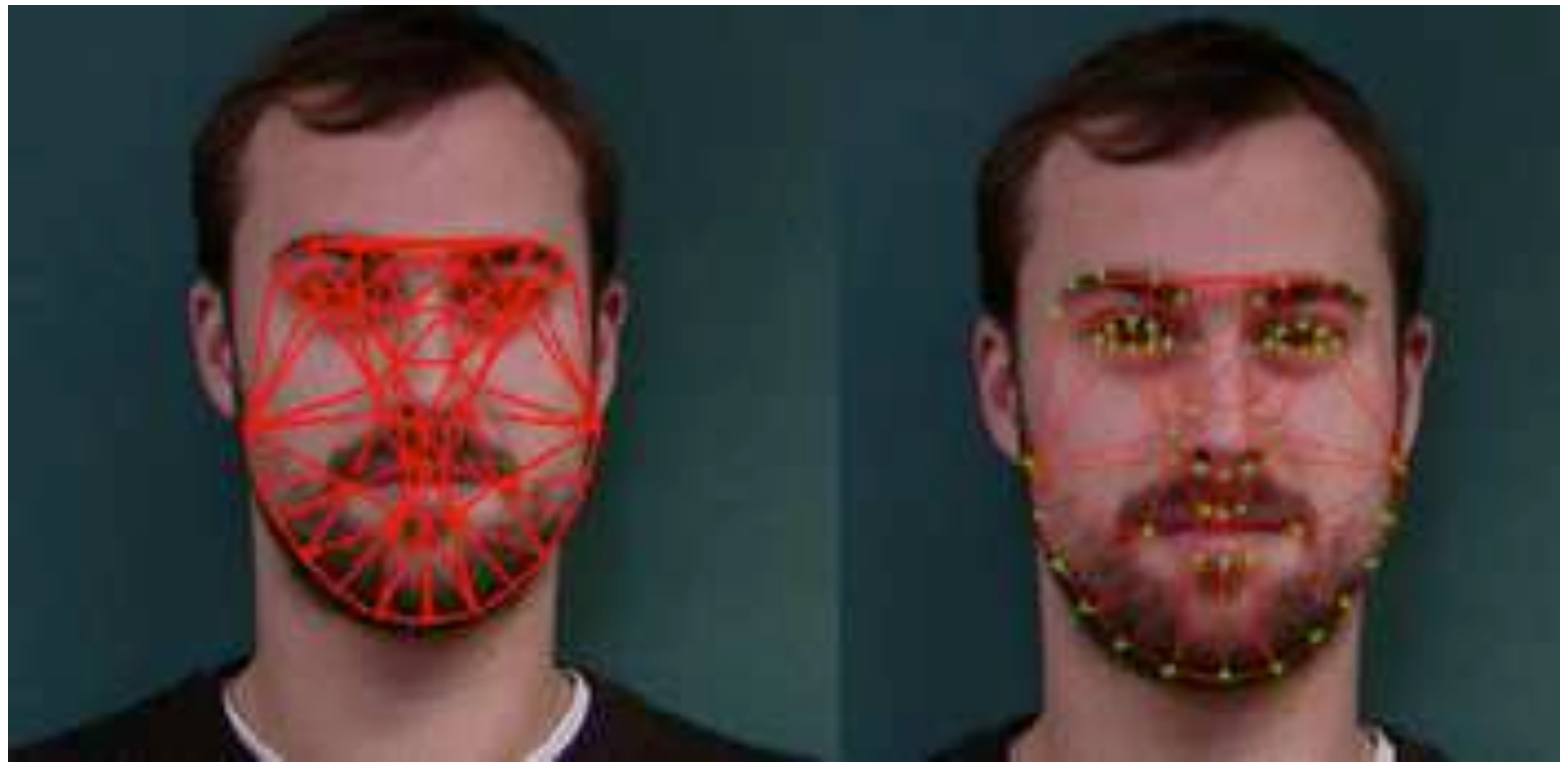


Shape Model



Point Distribution Model
(PDM)

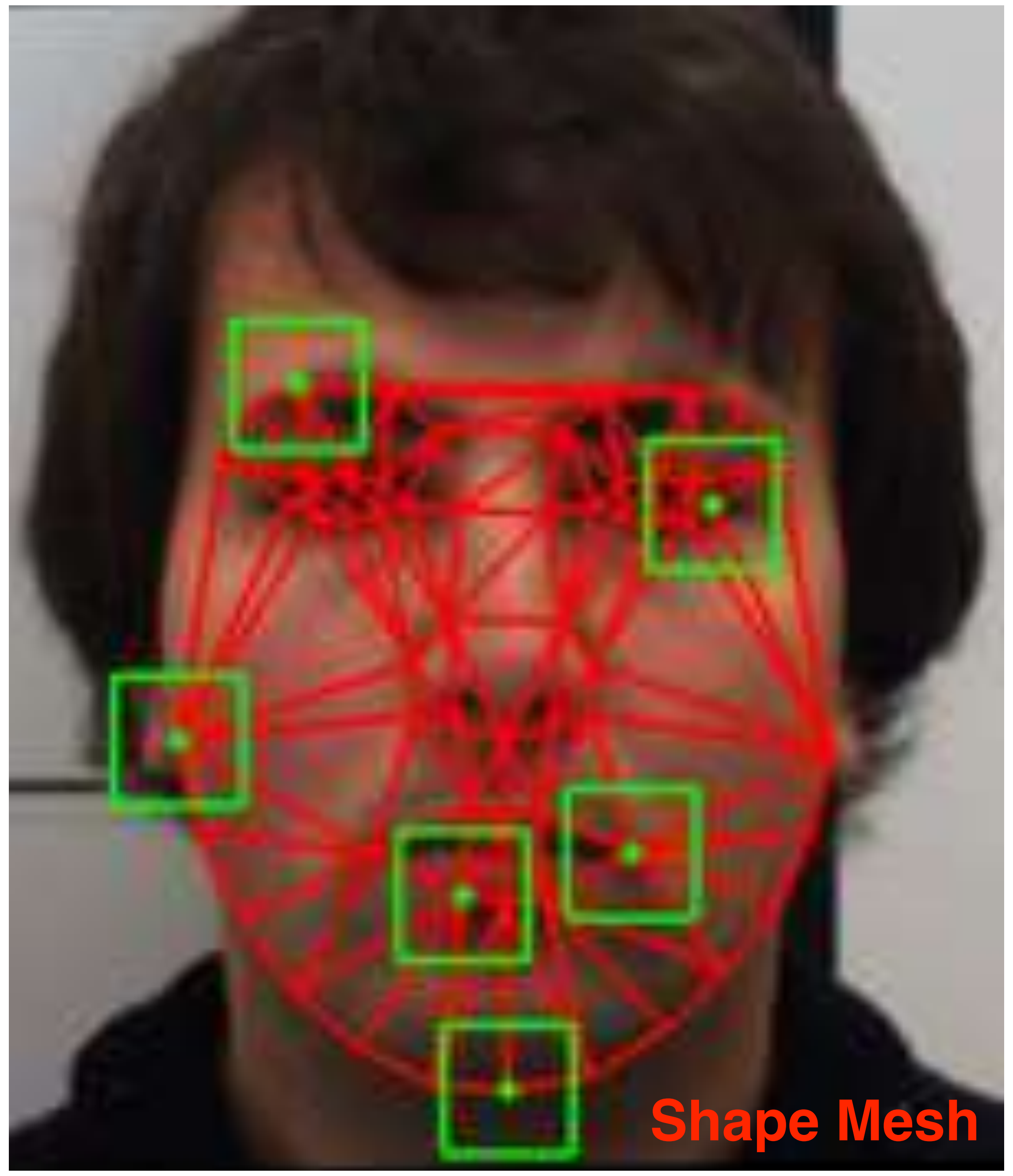
- **Discriminative / Patch Based Appearance Model**



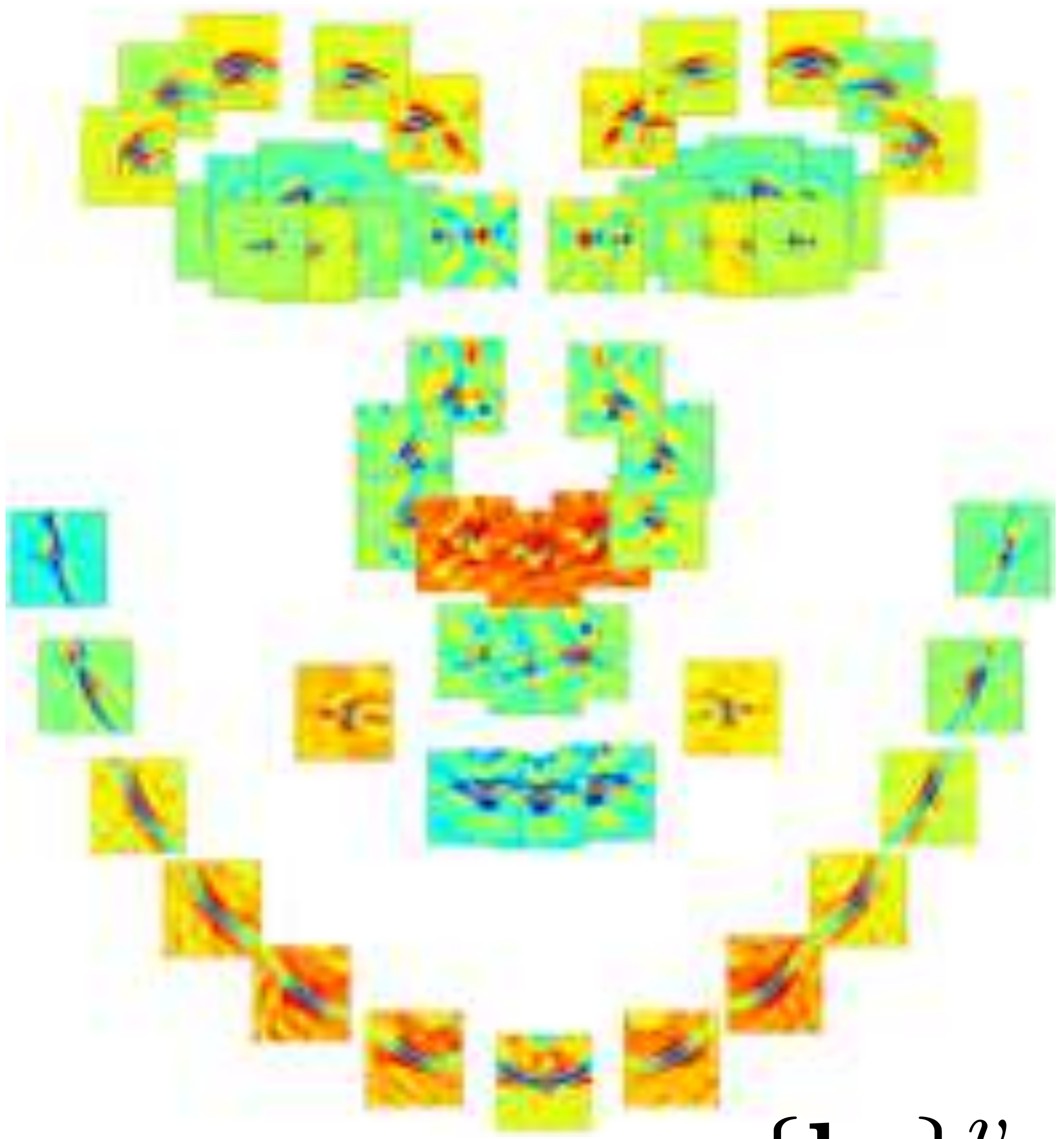
Constrained Local Model (CLM)

Data Term Regularization Term

$$\arg\max_{\mathbf{p}} \sum_{i=1}^{v} \mathbf{I}(\mathbf{s}_i) * \mathbf{h}_i - \lambda_0 \; \mathbf{p}^T \Sigma_{\mathbf{p}}^{-1} \mathbf{p}$$

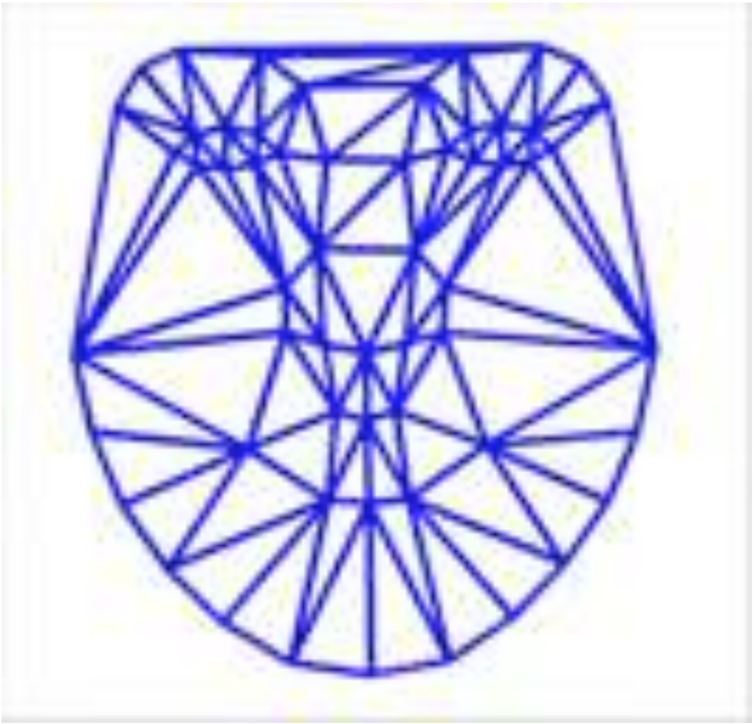


Local Search Regions



$\{\mathbf{h}_i\}_1^v$

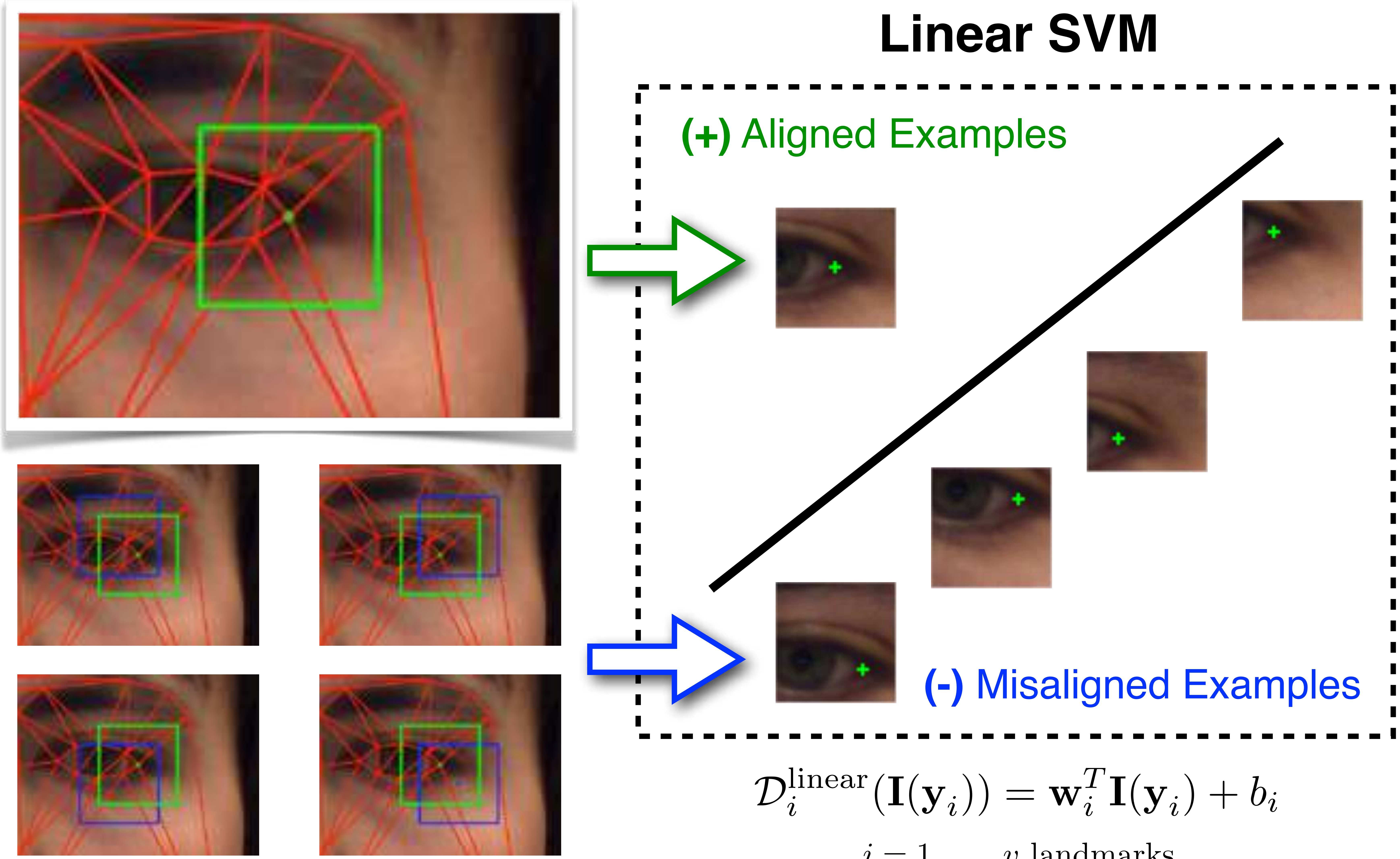
Local Detectors



$$\mathbf{s} = \mathbf{s}_0 + \sum_{i=1}^{n} p_i \phi_i$$

Shape Model

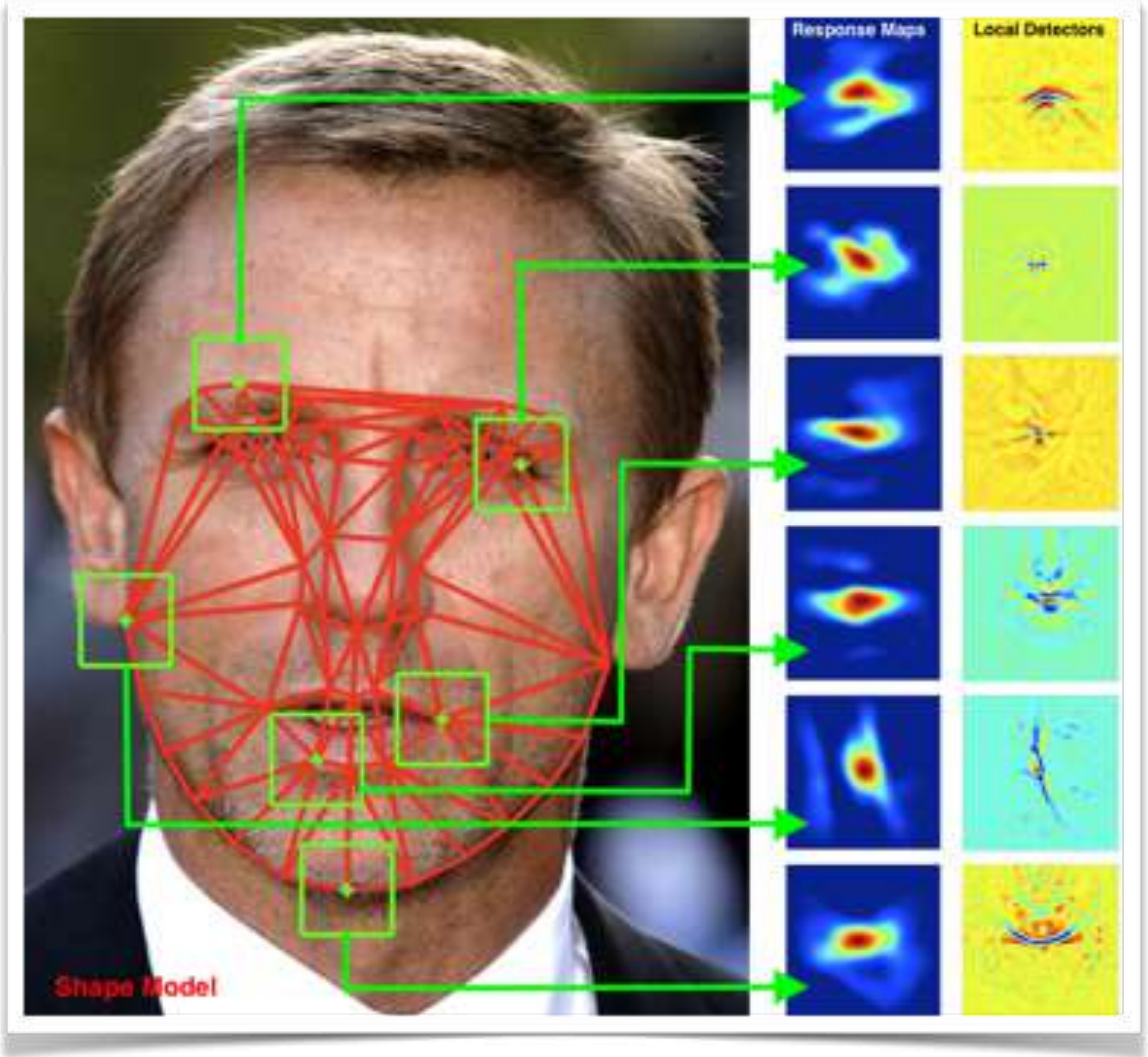
Local Landmark Detectors - SVM



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

$i = 1, \ldots, v$ landmarks

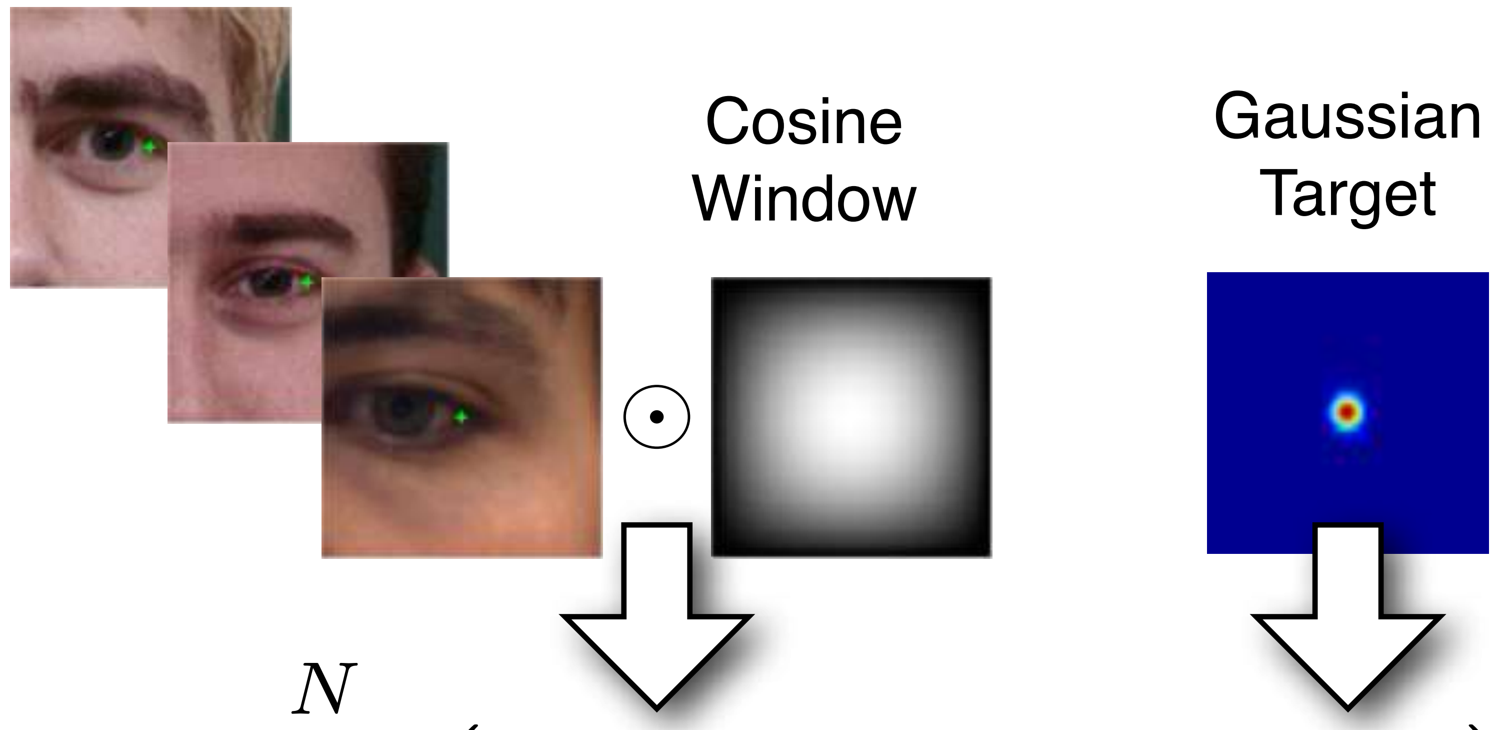
Local Landmark Detectors (MOSSE Filters)



$p(\mathbf{z}_i)$ $\mathbf{h}_i$

Regression Problem

$$\arg\min_{\mathbf{h}_i} \sum_{j=1}^{N} \left(\mathbf{h}_i * \mathbf{I}_j - \mathbf{g}_j\right)^2 + \lambda ||\mathbf{h}_i||^2$$



$$\min_{\mathbf{H}^\dagger} \sum_{j=1}^{N} \left(\mathcal{F}\{\mathbf{I}_j\} \odot \mathbf{H}_i^\dagger - \mathcal{F}\{\mathbf{g}_j\}\right)^2 + \lambda ||\mathbf{H}_i||^2$$

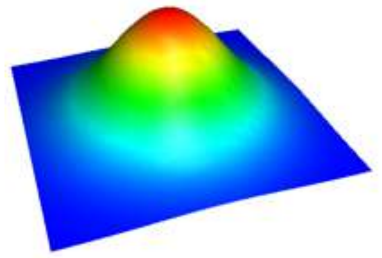
solution (spatial domain)

$$\mathbf{h}_i = \mathcal{F}^{-1}\left\{\frac{\sum_{j=1}^{N} \mathcal{F}\{\mathbf{g}_j\} \odot \mathcal{F}\{\mathbf{I}_j\}^\dagger}{\sum_{j=1}^{N} \mathcal{F}\{\mathbf{I}_j\} \odot \mathcal{F}\{\mathbf{I}_j\}^\dagger + \lambda}\right\}^\dagger$$

Bayesian Inference CLM

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

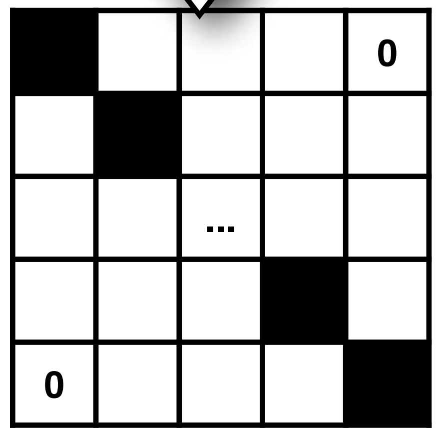
Likelihood Term



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

Shape
Observation

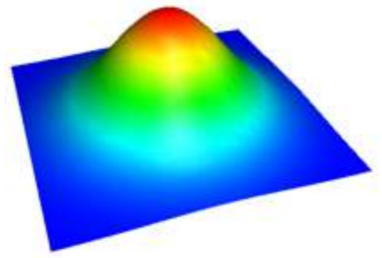
$(\mathbf{y}, \Sigma_{\mathbf{y}})$



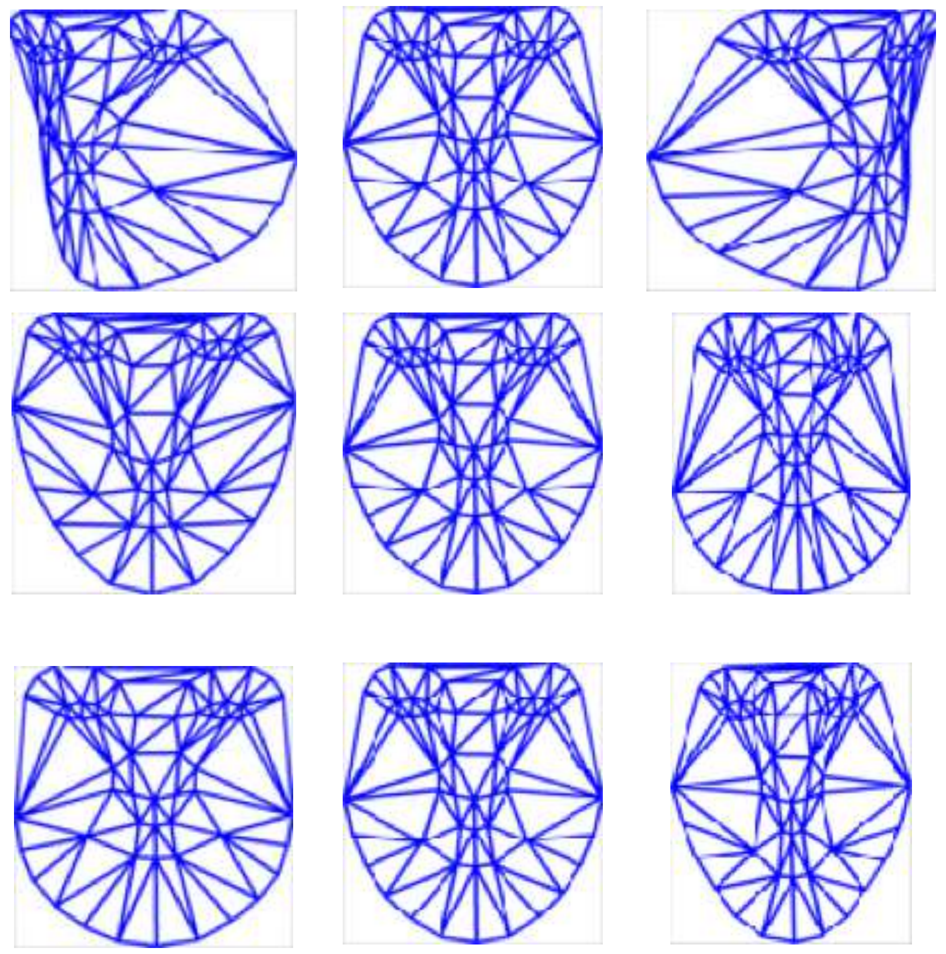
Uncertainty
Covariance

$2v \times 2v$
Block diagonal

Prior Term



$$p(\mathbf{b}) \propto \mathcal{N}(\mathbf{b}|\mathbf{0}, \Lambda)$$

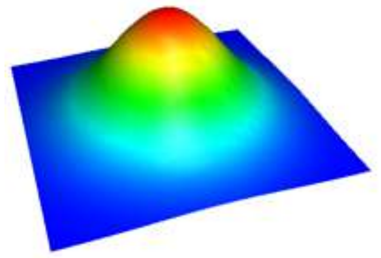


Linear Dynamic System (LDS)

$$\begin{aligned} \mathbf{b}_l &= \mathbf{I}_n \mathbf{b}_{l-1} + q, \quad & q \sim \mathcal{N}(\mathbf{0}, \Lambda) \\ \mathbf{y} - \mathbf{s}_0 &= \Phi\mathbf{b}_l + r, \quad & r \sim \mathcal{N}(\mathbf{0}, \Sigma_{\mathbf{y}}) \end{aligned}$$

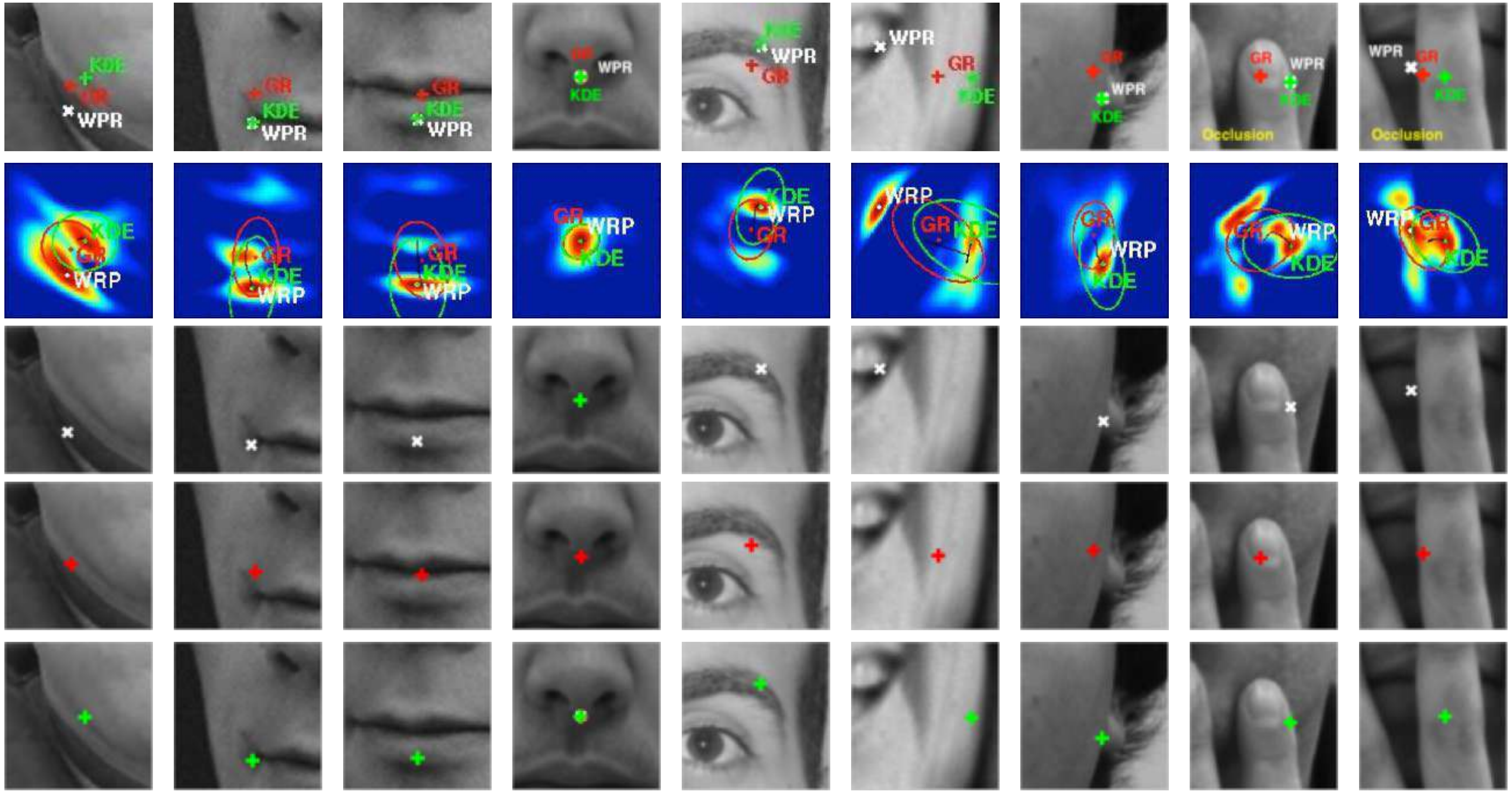
Posterior Term

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



Local Optimization Strategies

Patches under occlusion

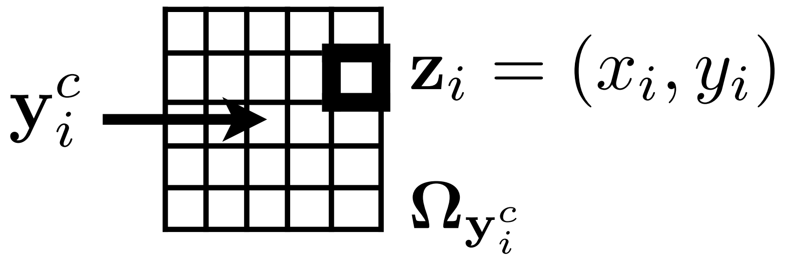


$p_i(\mathbf{z}_i)$

WPR — Weighted Peak Response

GR — Gaussian Response

KDE — Kernel Density Estimator



Weighted Peak Response

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

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

Gaussian Response

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

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

Kernel Density Estimator

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

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

Non-Parametric Bayesian Inference CLM

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

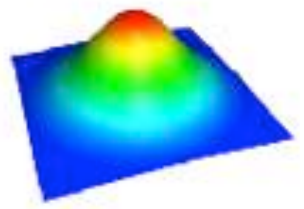
Posterior Expectation

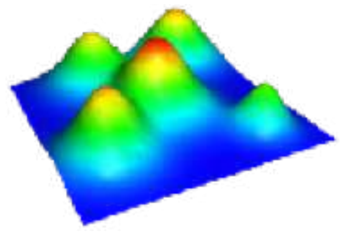
$$\hat{\mathbf{b}}_k = \frac{1}{N}\sum_{i=1}^{N}\tilde{\mathbf{b}}_k^{(i)}$$

Posterior

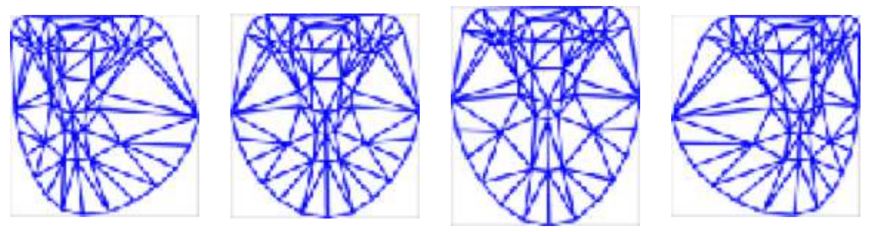
Likelihood

Prior



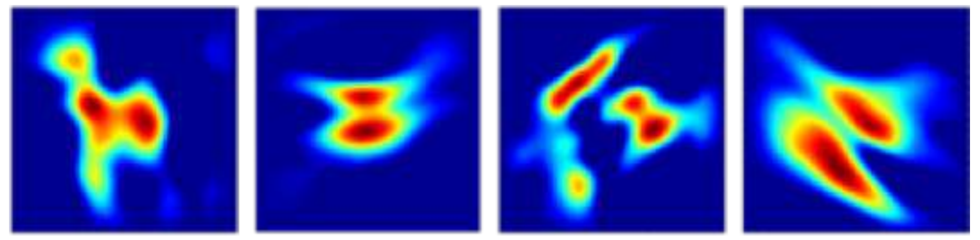


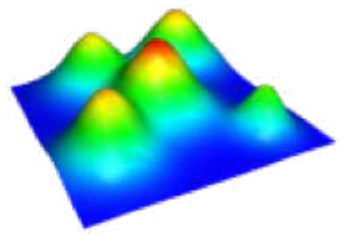
Prior Term



$$p(\mathbf{b}) \propto \mathcal{N}(\mathbf{b}|\mathbf{0}, \Lambda)$$

Multimodal Likelihood





Posterior Term

Kernel Density Estimator (KDE)

$$p(\mathbf{b}_k|\mathbf{y}_k,\ldots,\mathbf{y}_0) \approx \sum_{i=1}^{N} w_k^{(i)} K_h(\mathbf{b}_k - \mathbf{b}_k^{(i)})$$

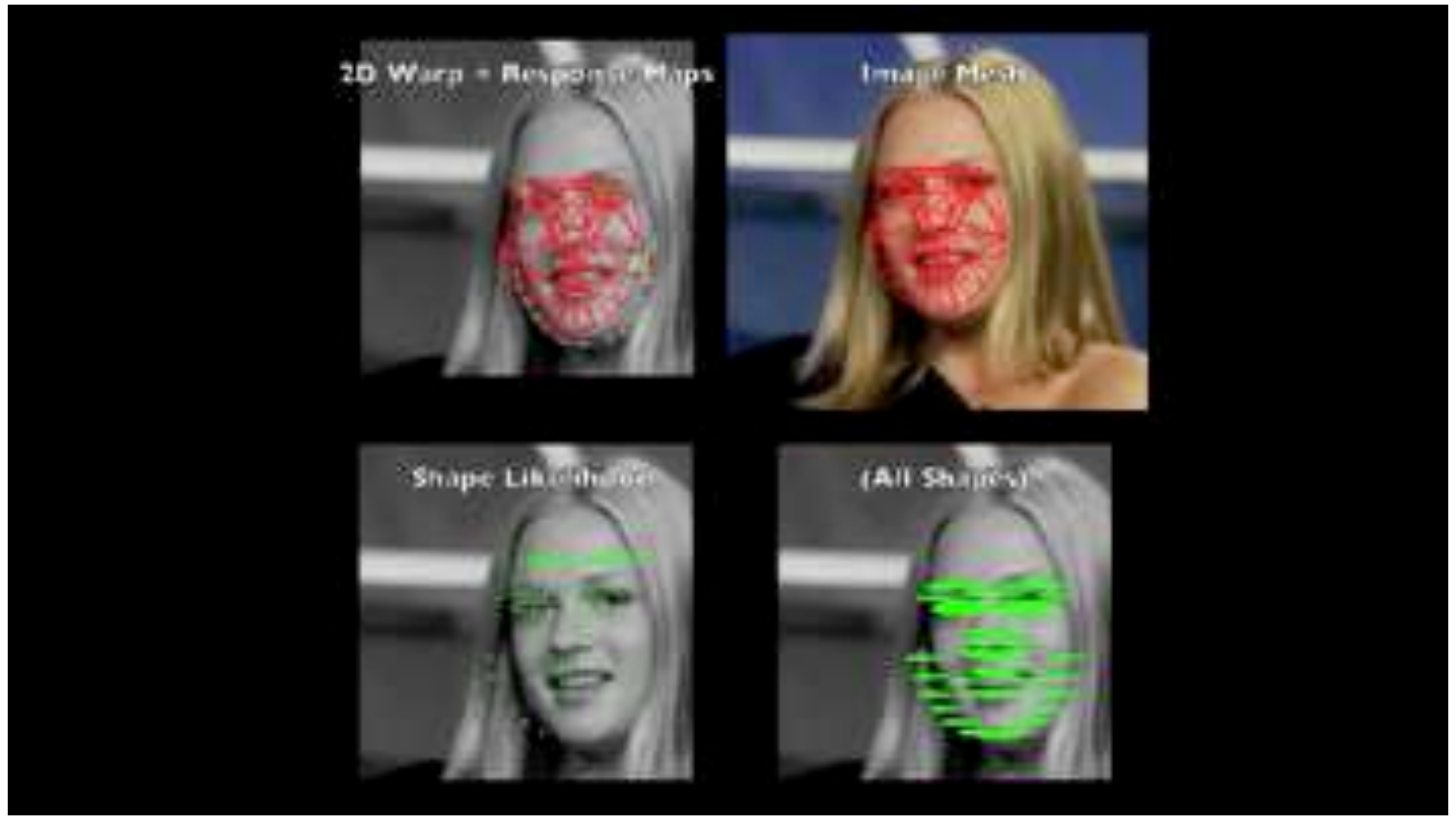
$$\{w_k^{(i)}, \mathbf{b}_k^{(i)}\}_{i=1}^{N}$$

(i) - Particle (possible shape)
(k) - Iteration

Inference by a Regularized Particle Filter (RPF)

$$w_k^{(i)} \propto p(\mathbf{y}_k|\mathbf{b}_k^{(i)}) = \rho\left(\prod_{j=1}^{v} p(a_j = 1|\mathcal{D}_j, \mathbf{I}(\mathbf{y}_j)); \sigma\right)$$

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



Multiple Detectors per Landmark

Training examples for a given landmark:

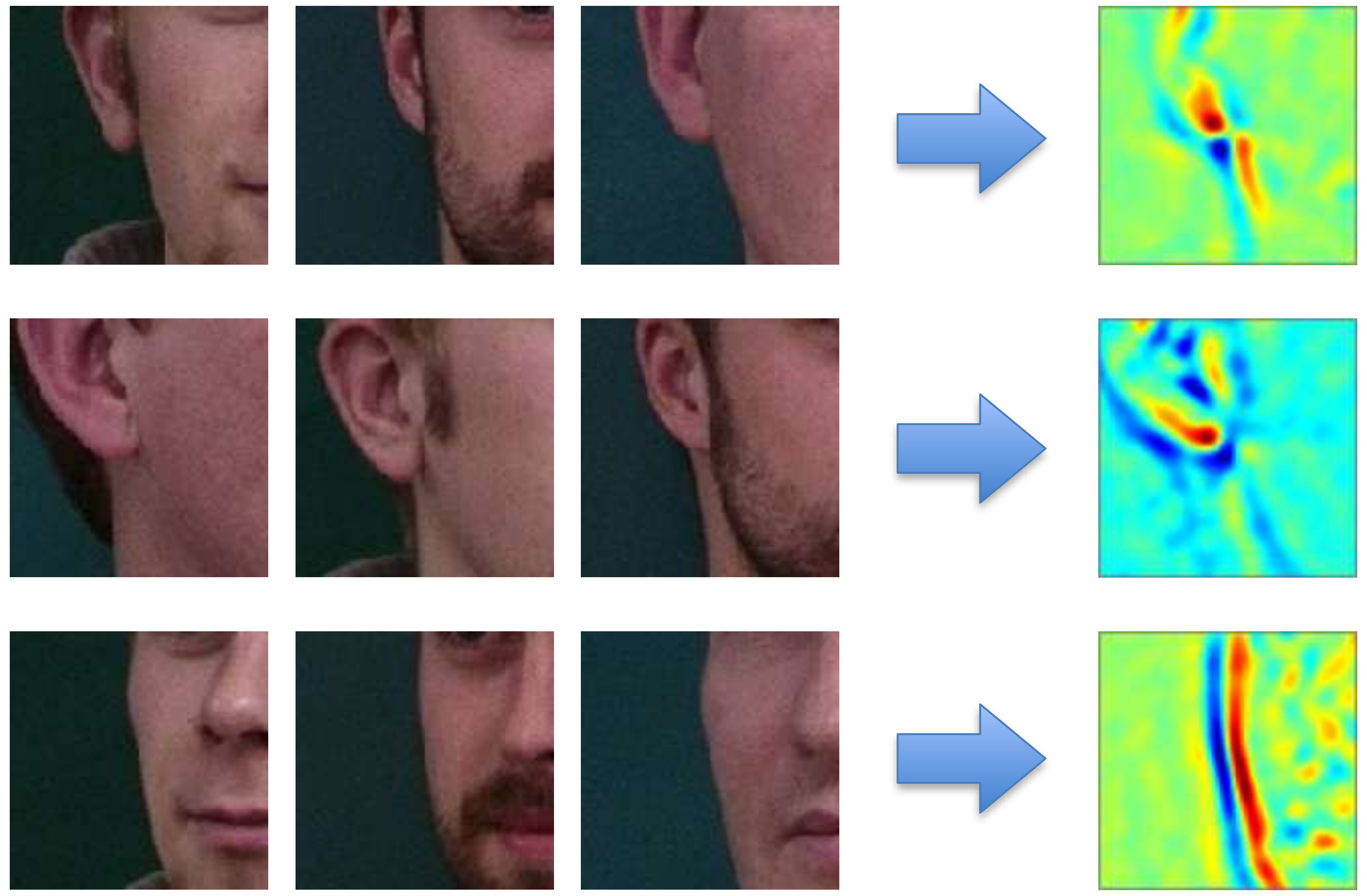
$\mathbf{I}(\mathbf{x}_i)$ MOSSE Filter

$\mathbf{h}_i$

Sampled Region

Clustering of training examples

MOSSE Filter



$\mathbf{I}(\mathbf{x}_j)$ $\mathbf{h}_i^{(m)}$

Unsupervised Clustering

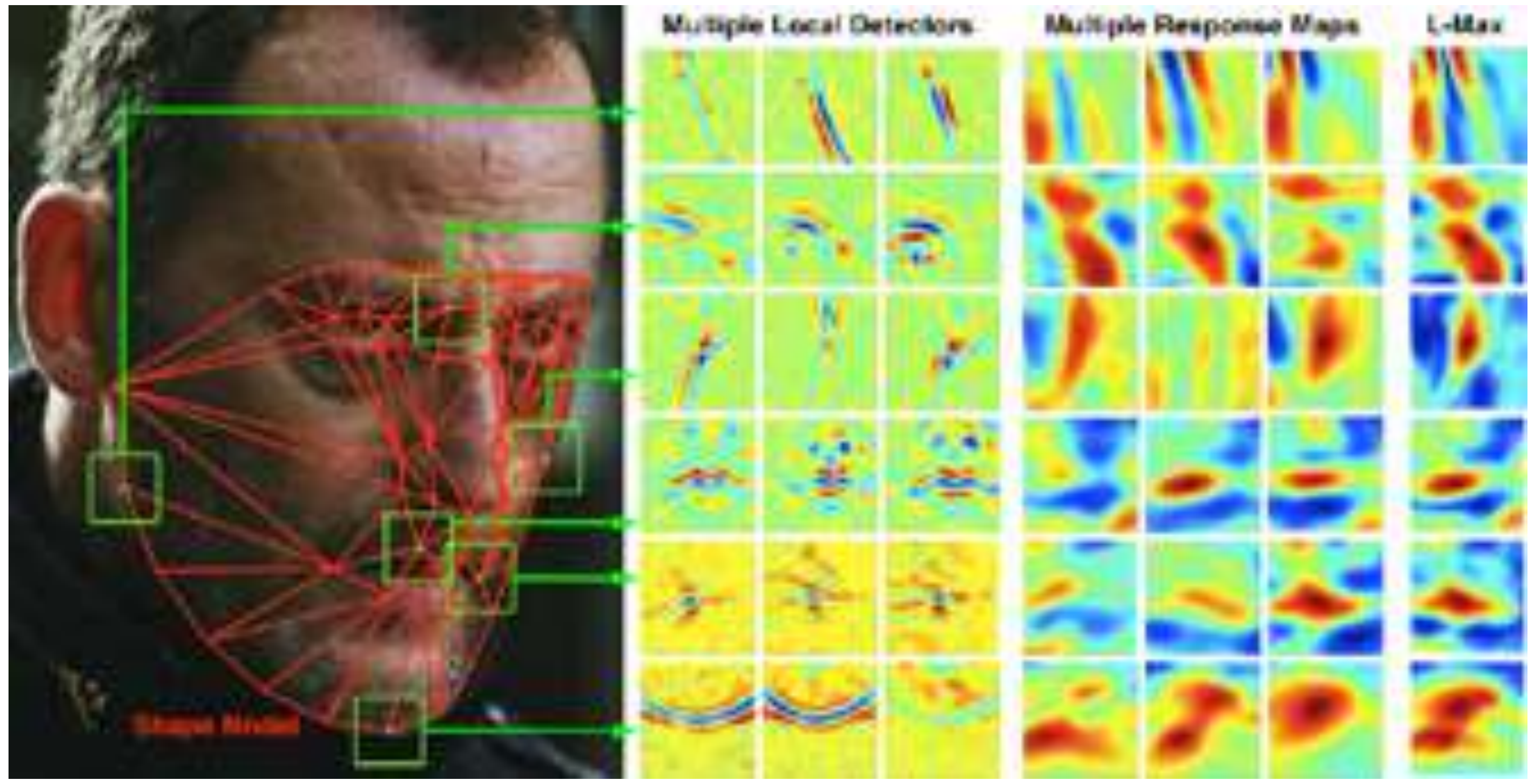
$$\arg\max_{\mathbf{h}_i^{(m)}} \sum_{j=1}^{N} \sum_{m=1}^{M} \mathbf{I}(\mathbf{x}_j) * \mathbf{h}_i^{(m)}$$

M - clusters
N - examples

Solve (for each landmark i) using a two step approach:
- Initial clustering by k-means

① Build basic detectors using the current clustering estimate
② Move samples to the cluster with highest correlation

- Repeat until no more samples change



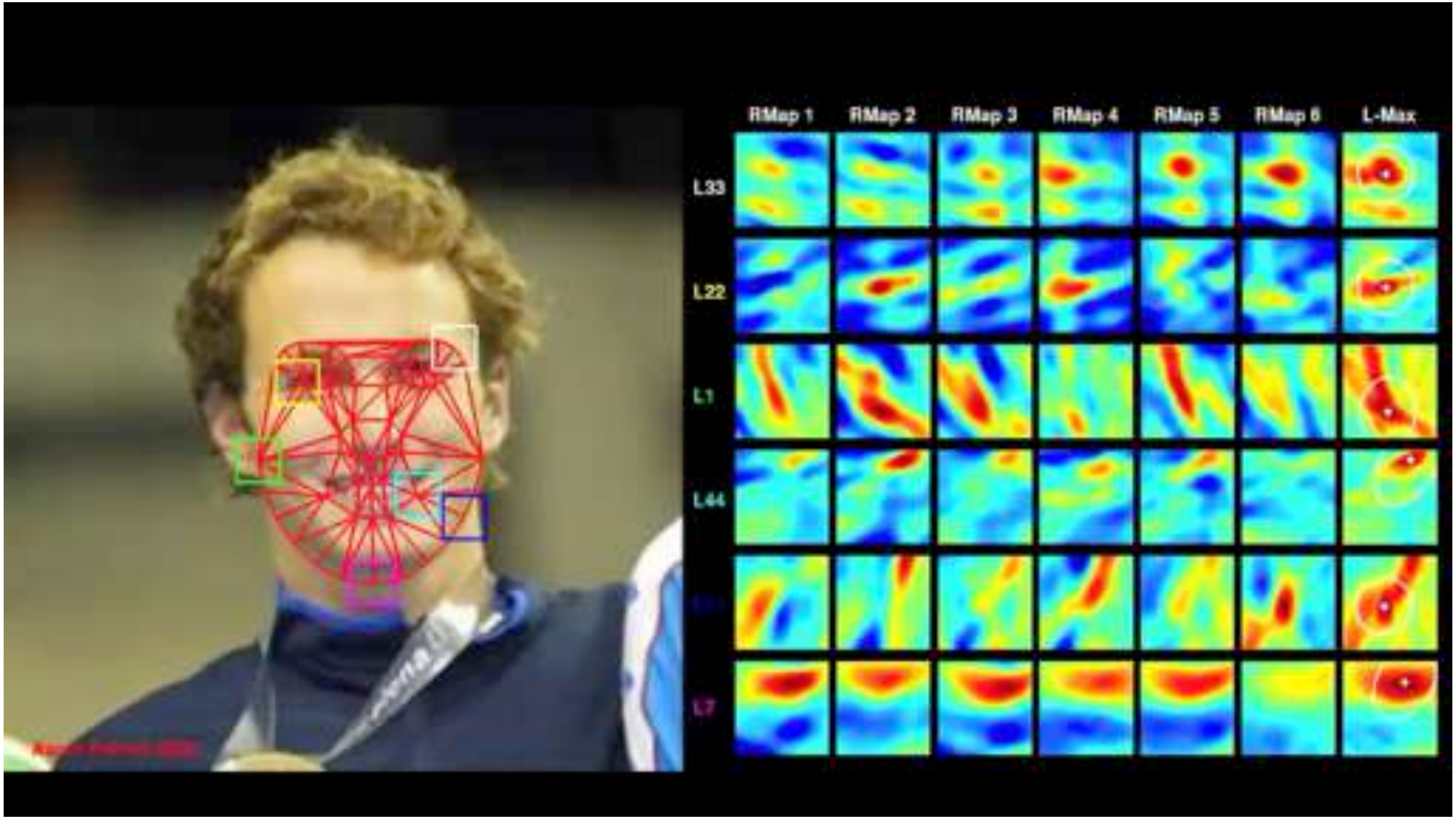
Multiple Response Maps

$$p_i(\mathbf{z}_i)^{(m)} = \frac{1}{1 + e^{-a_i \beta_1 \mathcal{D}_i(\mathbf{I}(\mathbf{z}_i)) + \beta_0}}$$

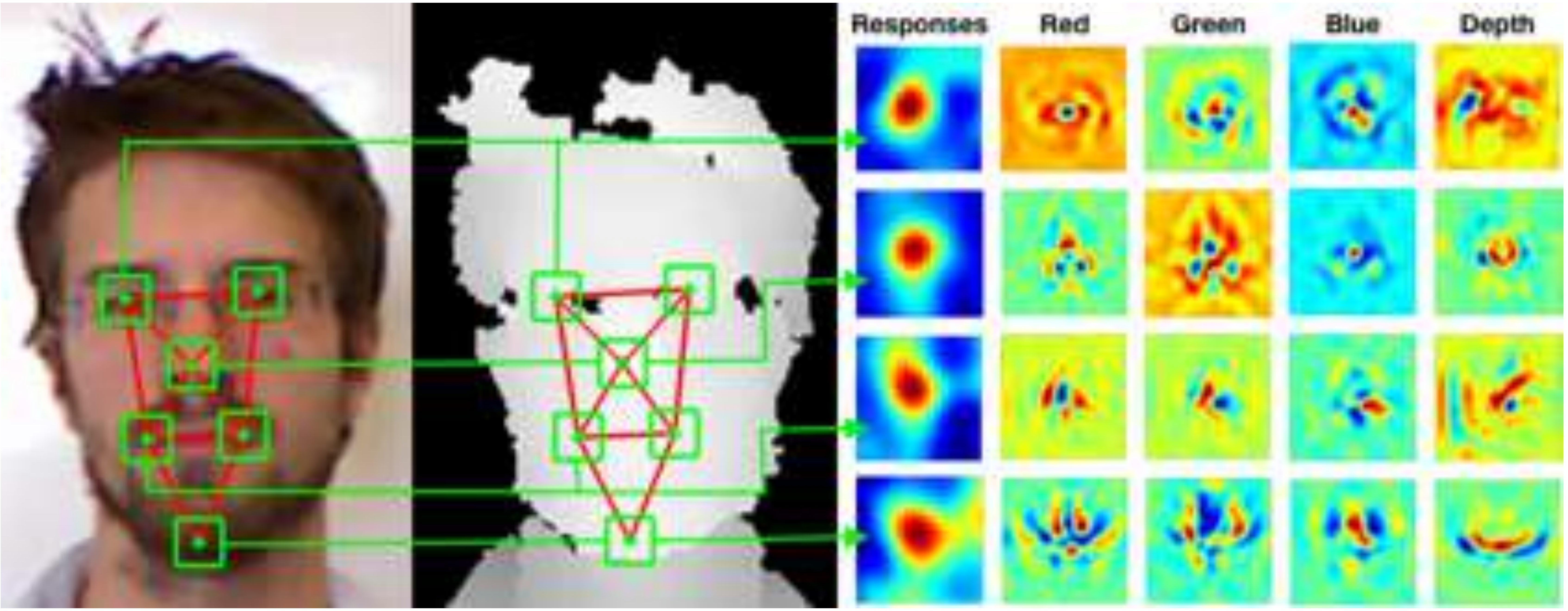
Combining Multiple Detections

$$p_i(\mathbf{z}_i)_\infty = \max_{\mathbf{z}_i}\{p_i(\mathbf{z}_i)^{(1)}, \ldots, p_i(\mathbf{z}_i)^{(M)}\}$$

Multiple Detectors por Landmark (video)

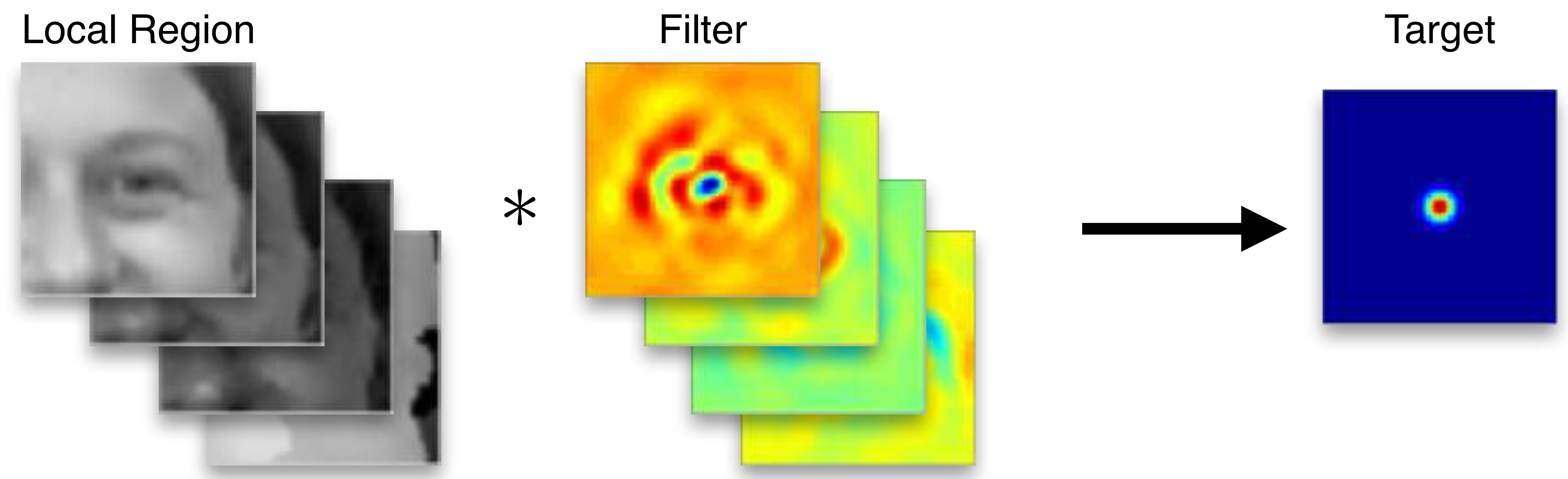


CLM with Depth Data



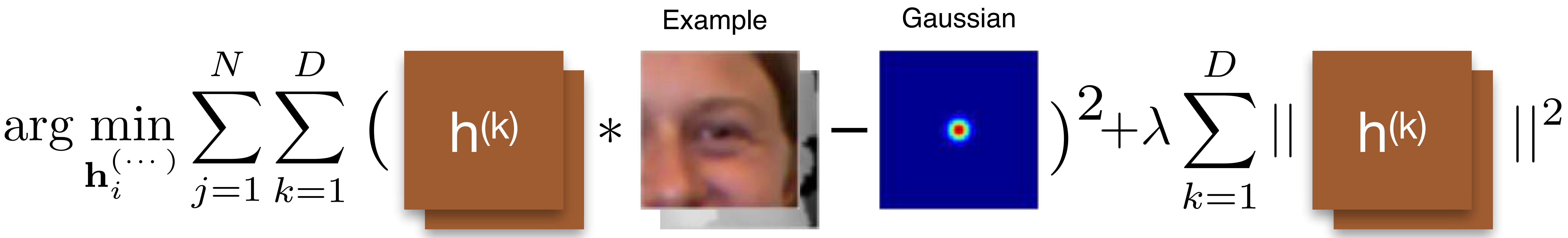
- Strategy Employed:
  - Multiple Channel Local Detectors (RGBD - w/ single response map)
  - Fast CLM Inference (Gaussian)

Multiple Channel Correlation Filters



Spatial Domain

$$\arg \min_{\mathbf{h}_i^{(1)},\ldots,\mathbf{h}_i^{(D)}} \sum_{j=1}^{N} \sum_{k=1}^{D} \left(\mathbf{h}_i^{(k)} * \mathbf{I}_j^{(k)} - \mathbf{g}_j \right)^2 + \lambda \sum_{k=1}^{D} ||\mathbf{h}_i^{(k)}||^2$$



Minimization across all channels

BCLM w/ Depth Data

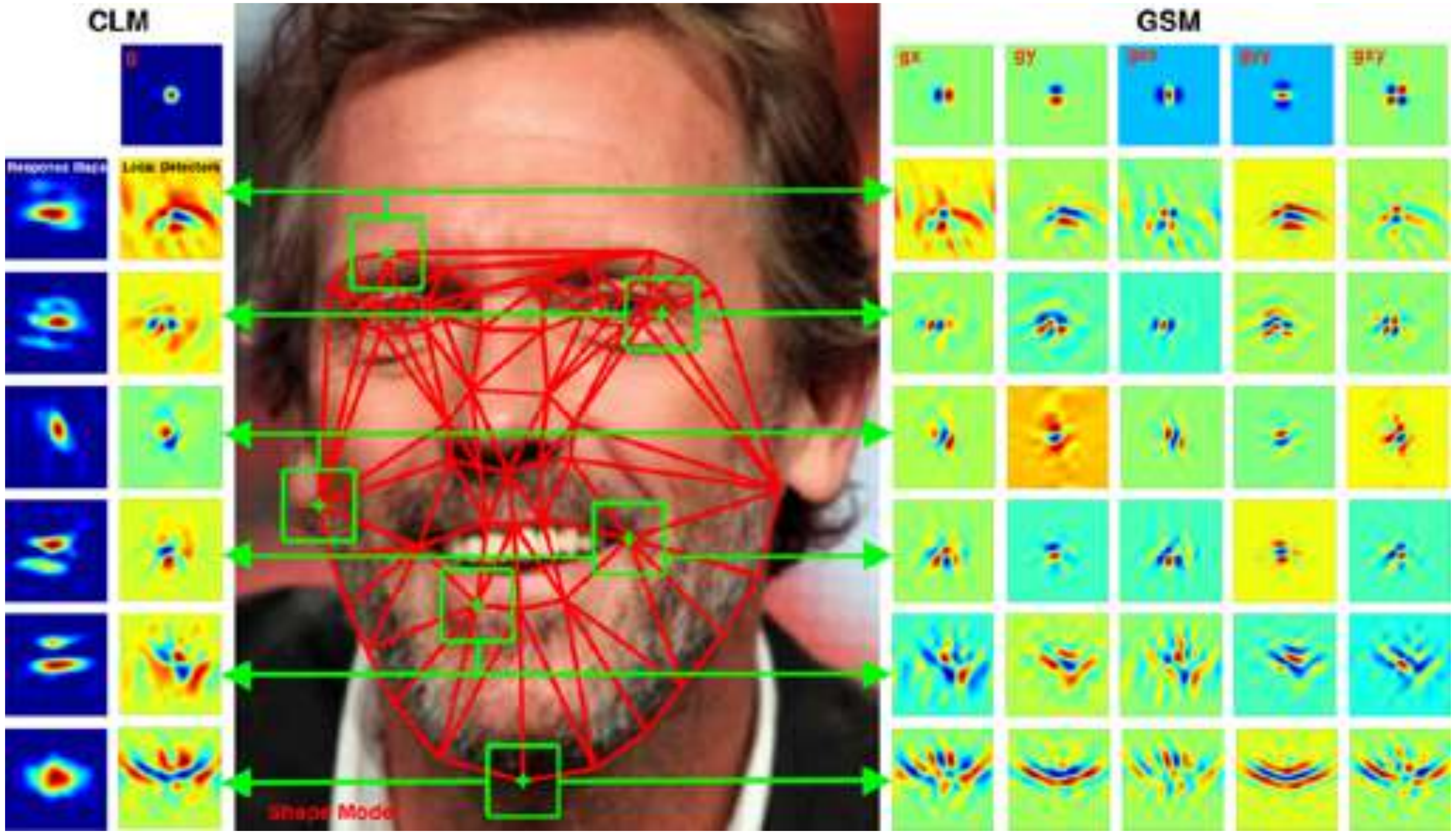
Gradient Shape Model (GSM)

s - shape vector at image frame

Data Term — Regularization Term

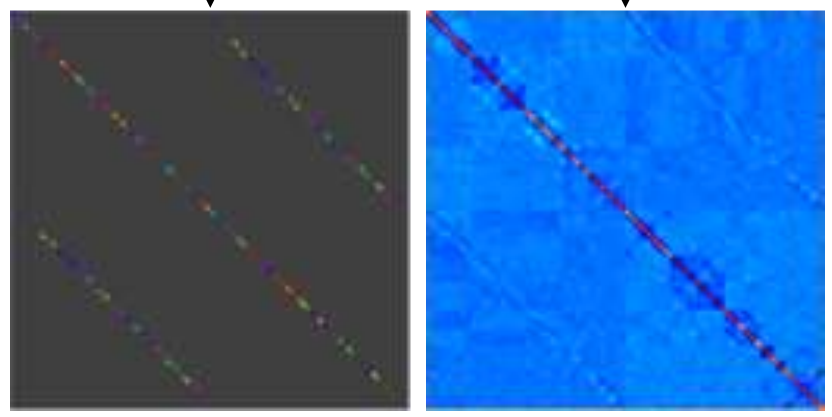
$$\arg\min_{\mathbf{s},\boldsymbol{\theta}} - \sum_{i=1}^{v} D_i(\mathbf{I}(\mathbf{s}_i), \boldsymbol{\theta}) + \lambda_1 \left(\mathcal{S}(\mathbf{s}, \boldsymbol{\theta}) - \mathbf{s}_0\right)^T \Sigma_{\mathbf{s}}^{-1} \left(\mathcal{S}(\mathbf{s}, \boldsymbol{\theta}) - \mathbf{s}_0\right)$$

Similarity Transform — Similarity Transform



$$\nabla_f(\mathbf{s}, \boldsymbol{\theta}) = \nabla_{\mathrm{D}}(\mathbf{s}) + \lambda_1 \nabla_{\mathrm{R}}(\mathbf{s}, \boldsymbol{\theta})$$

$$\mathbf{H}_f(\mathbf{s}, \boldsymbol{\theta}) = \mathbf{H}_{\mathrm{D}}(\mathbf{s}) + \lambda_1 \mathbf{H}_{\mathrm{R}}(\mathbf{s}, \boldsymbol{\theta})$$

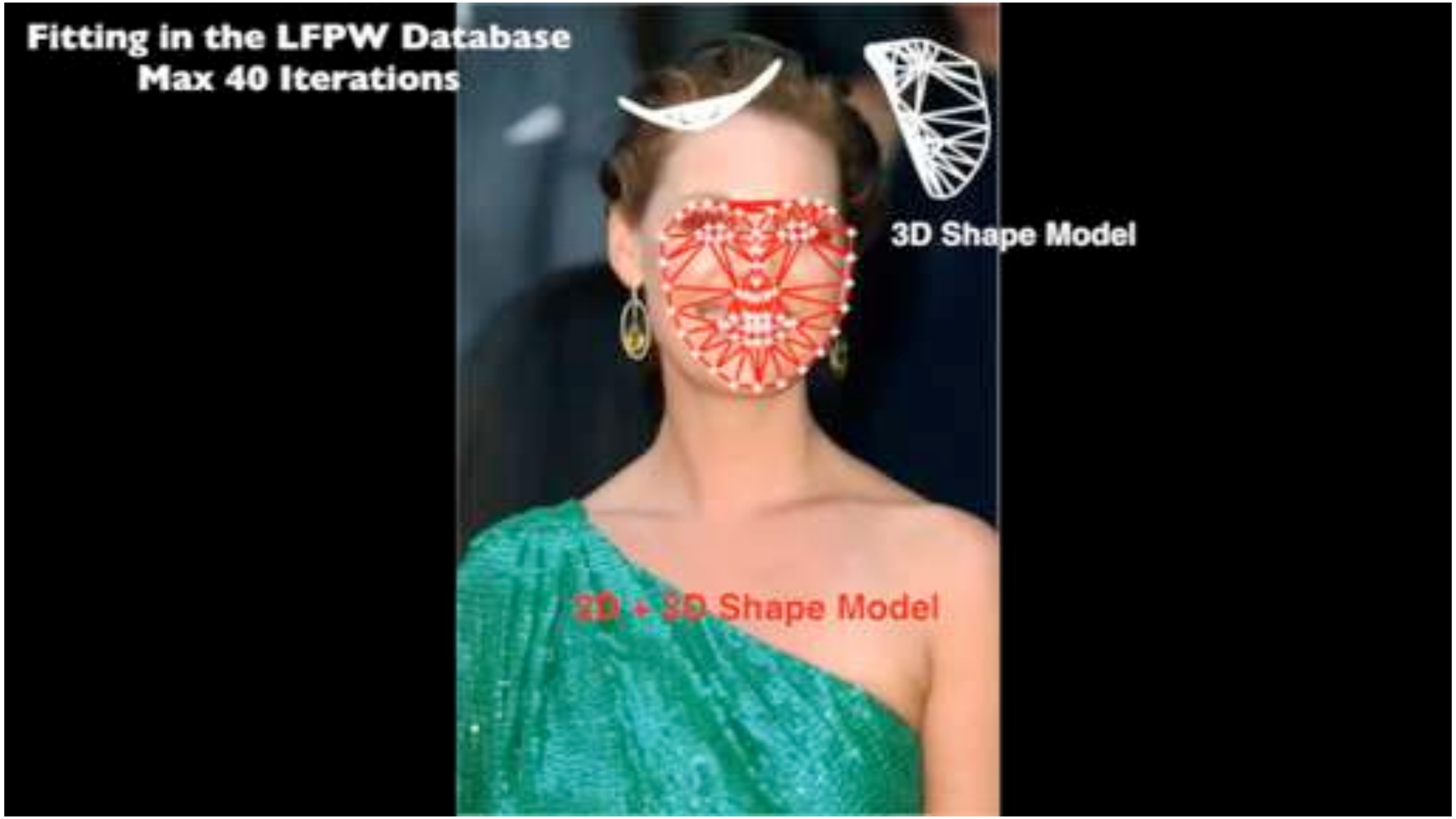


$$\nabla_{\mathrm{D}}(\mathbf{s}) = \left[\ \mathcal{I}_1^T \frac{\partial \mathbf{h}_1}{\partial x_1}\ \ldots\ \mathcal{I}_v^T \frac{\partial \mathbf{h}_v}{\partial x_v}\quad \mathcal{I}_1^T \frac{\partial \mathbf{h}_1}{\partial y_1}\ \ldots\ \mathcal{I}_v^T \frac{\partial \mathbf{h}_v}{\partial y_v}\quad \mathbf{0}_4\ \right]^T$$

2D + 3D Gradient Shape Model (GSM)

Data Term — 2D Regularization Term — 3D Regularization Term — Projection Residual

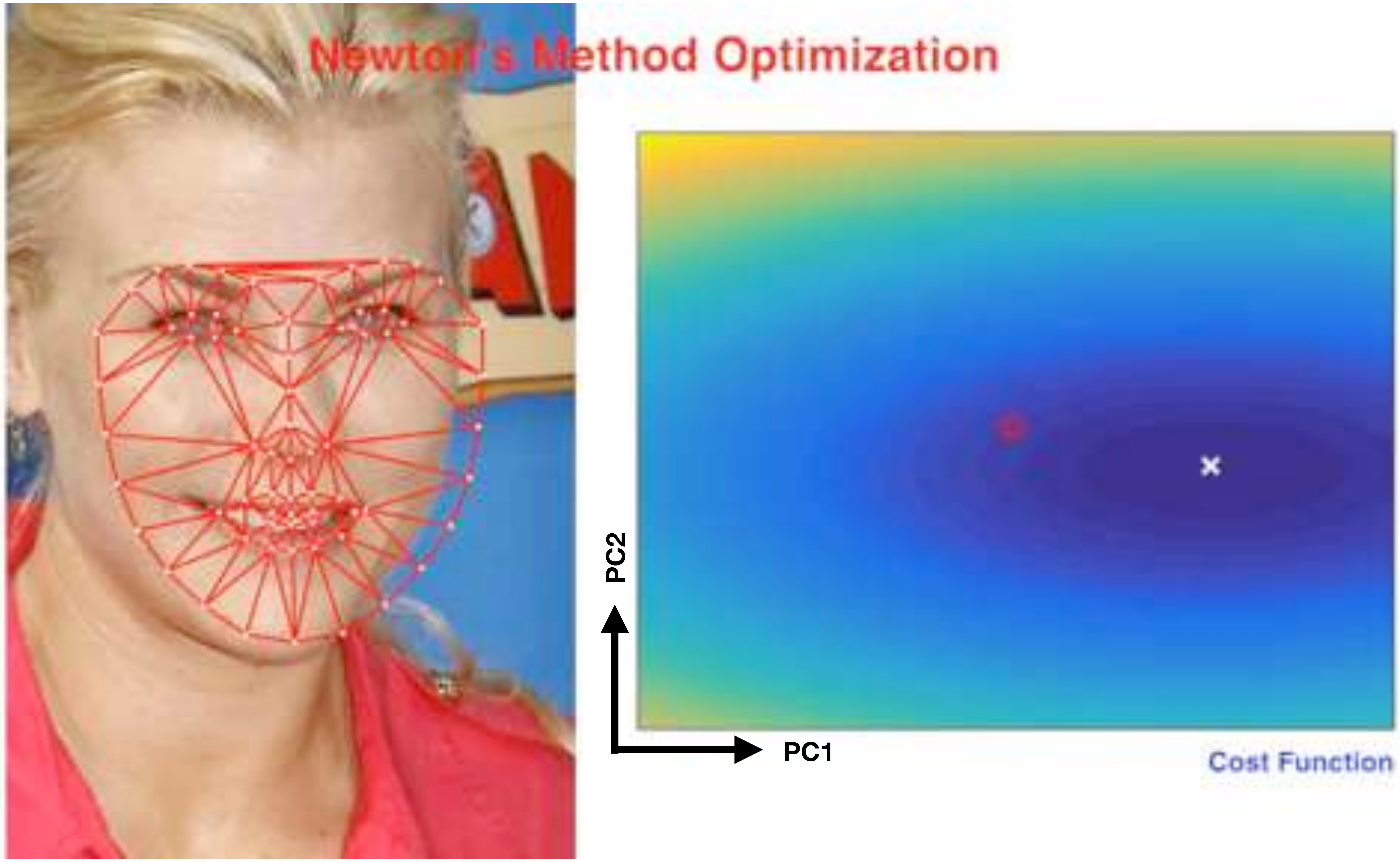
$$\arg\min_{\mathbf{s},\boldsymbol{\theta},\bar{\mathbf{s}},\mathbf{P}} -\sum_{i=1}^{v} D_i(\mathbf{I}(\mathbf{s}_i),\boldsymbol{\theta}) + \lambda_1 \left(\underbrace{\mathcal{S}(\mathbf{s},\boldsymbol{\theta})}_{\text{Similarity Transform}} - \mathbf{s}_0\right)^T \Sigma_{\mathbf{s}}^{-1} \left(\underbrace{\mathcal{S}(\mathbf{s},\boldsymbol{\theta})}_{\text{Similarity Transform}} - \mathbf{s}_0\right) + \lambda_2(\bar{\mathbf{s}} - \bar{\mathbf{s}}_0)^T \Sigma_{\bar{\mathbf{s}}}^{-1}(\bar{\mathbf{s}} - \bar{\mathbf{s}}_0) + \lambda_3 \|\mathbf{r}\|^2$$



Scaled Orthographic Projection

$$\mathbf{r} = \mathbf{s} - \sigma \underbrace{\begin{pmatrix} i_x & i_y & i_z \\ j_x & j_y & j_z \end{pmatrix}}_{\mathbf{R}} \otimes \mathbf{I}_v\, \bar{\mathbf{s}} - \begin{pmatrix} o_x \\ o_y \end{pmatrix} \otimes \mathbf{1}_v$$

Gradient Descent vs. Cascaded Regression



Gradient Descent

- Requires 'good' initialization.
- In general, requires to compute the Jacobian at each iteration.
- Require to compute the Hessian and its inverse (2nd order methods).
- Learning Fast.
- Testing Slow.

Cascaded Regression

- Captures the variance of the initialization.
- Precomputed Regression matrix.
- Learning Slow.
- Testing Fast.

Cascade Regression Framework

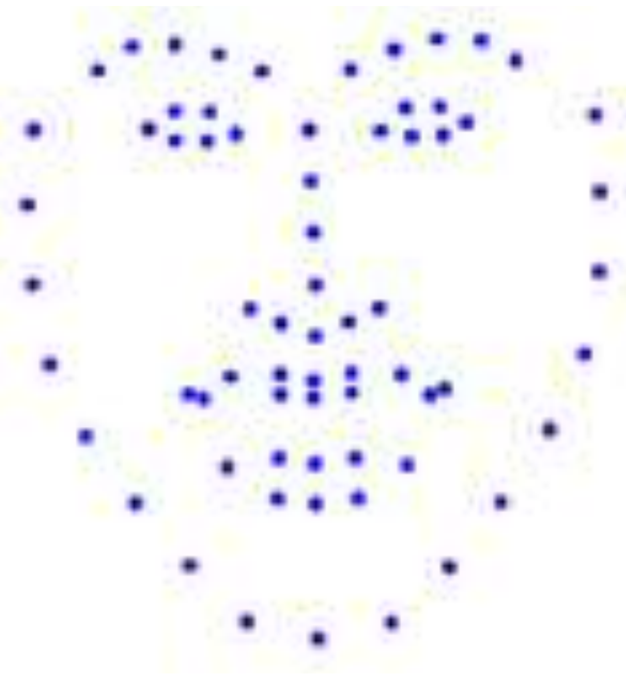
Level 1 → Level 2 → Level 3 → • • • Level K

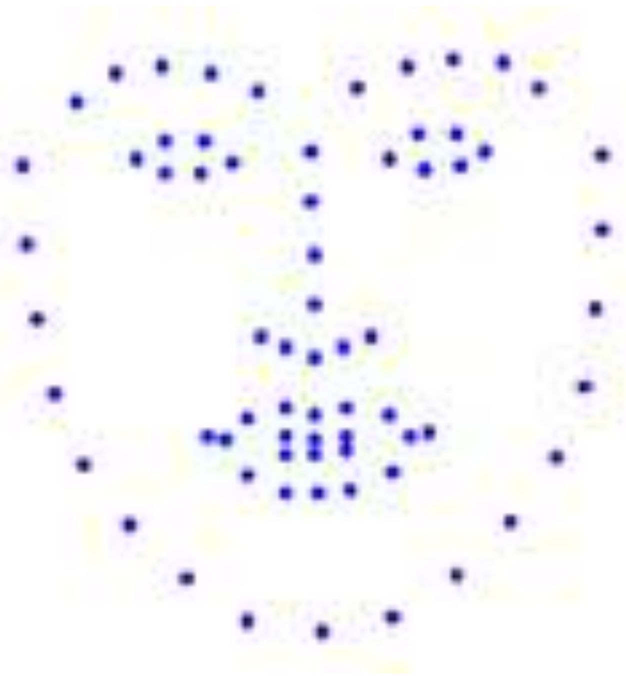
$$\mathbf{s}^k = \mathbf{s}^{k-1} + \mathbf{R}^{k-1}\mathcal{F}(\mathbf{I}, \mathbf{s}^{k-1})$$

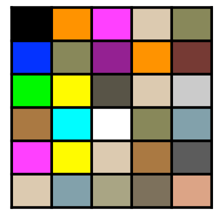
k **- cascade level**

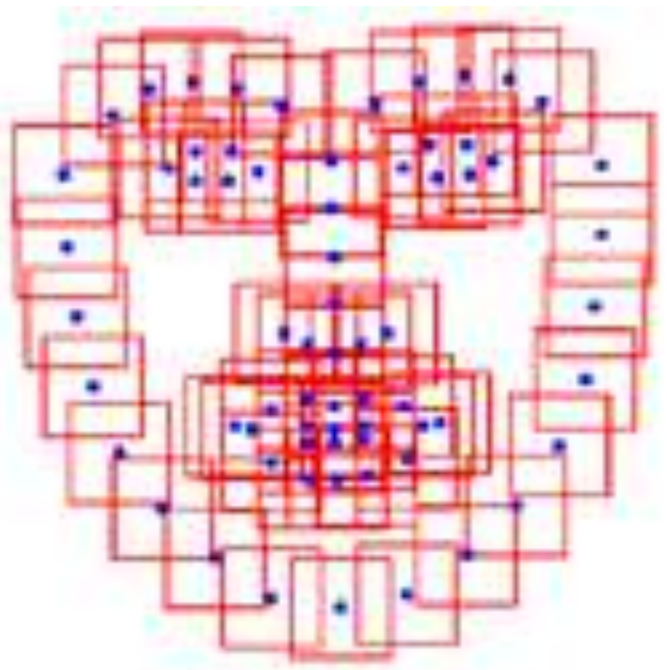
Updated shape vector | **Previous shape vector** | **Regression Matrix** | **Feature Extraction**

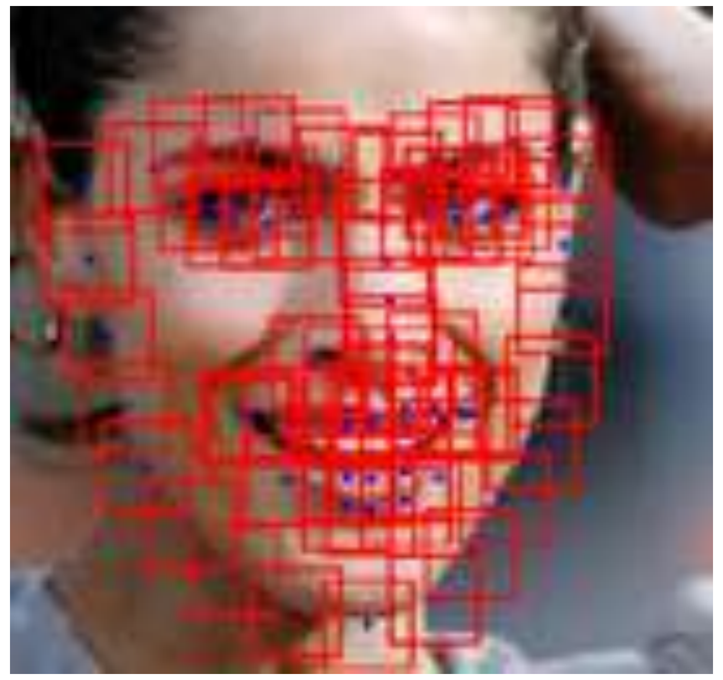
$$\mathbf{s} = \begin{pmatrix} x_0 \\ \vdots \\ x_v \\ y_0 \\ \vdots \\ y_v \end{pmatrix}$$











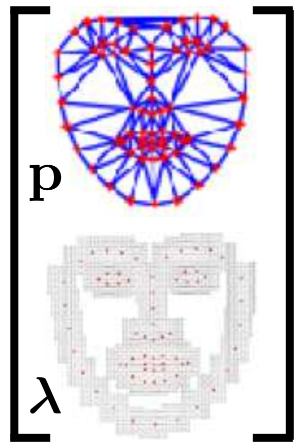
v - landmarks

$2v \times 1$ | $2v \times 1$ | $2v \times d$ | $d \times 1$

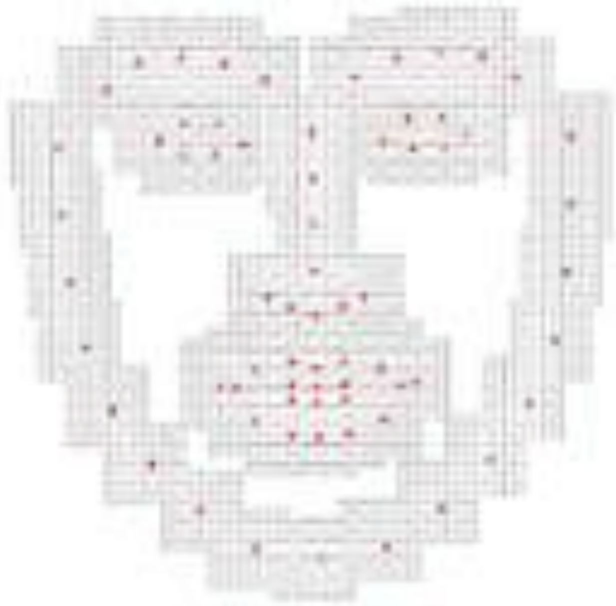
RGB | **HoG**

Simultaneous Cascaded Regression (SCR)

Regression with both shape and appearance structure



Level 1 → Level 2 → Level 3 → • • • Level K

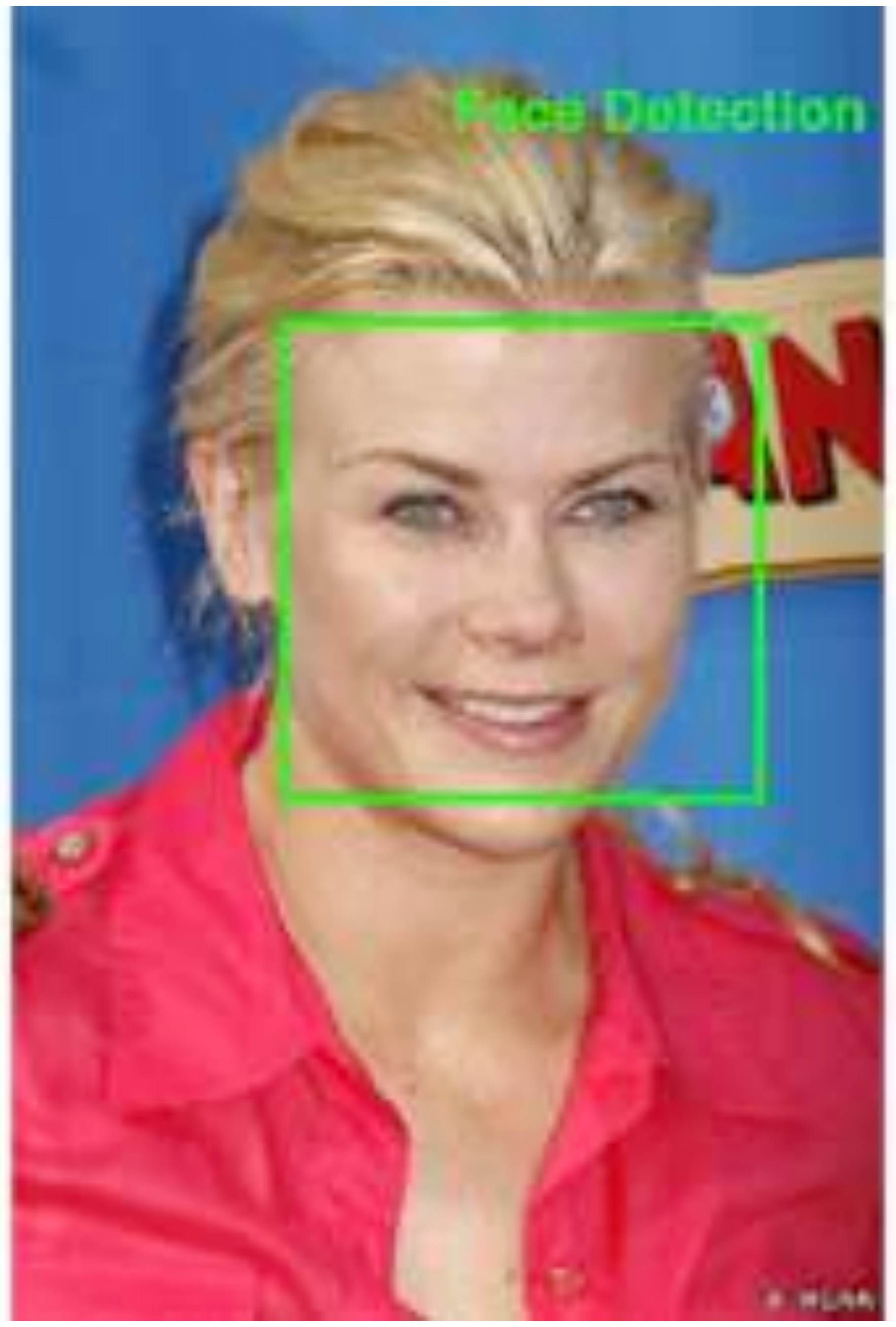


$$\begin{bmatrix}\mathbf{p}\\ \boldsymbol{\lambda}\end{bmatrix}^{k} = \begin{bmatrix}\mathbf{p}\\ \boldsymbol{\lambda}\end{bmatrix}^{k-1} + \mathbf{R}^{k-1}\left(\mathbf{I}(\mathcal{W}(\mathbf{p}^{k-1})) - \mathbf{A}_0 - \mathbf{A}\boldsymbol{\lambda}^{k-1}\right), \quad k = 1, \ldots, K$$

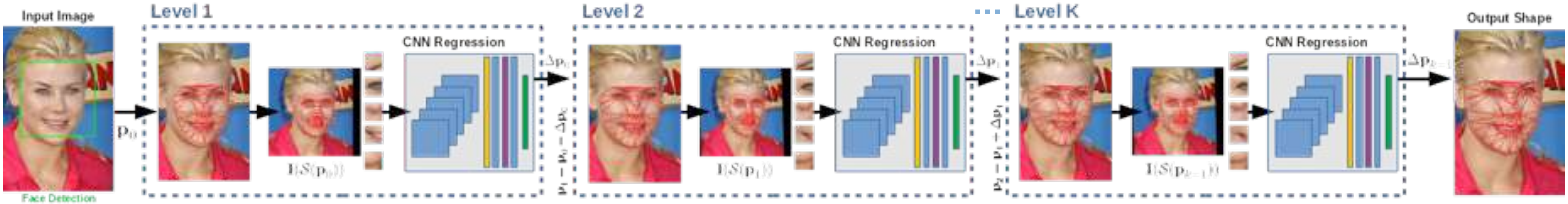
Shape + Appearance parameters

Features extracted at previous level

Features generated by the Model



Nonlinear Cascade Regression



Combined shape + pose parameters

$$\mathbf{p} = \begin{bmatrix} \mathbf{b} \\ \mathbf{q} \end{bmatrix} \in \mathbb{R}^{n+4}$$

$$\mathbf{p}^k = \mathbf{p}^{k-1} + \gamma \boldsymbol{\mathcal{R}}^{k-1}\{\mathcal{L}(\mathbf{I}(\mathcal{S}(\mathbf{p}^{k-1})))\}$$

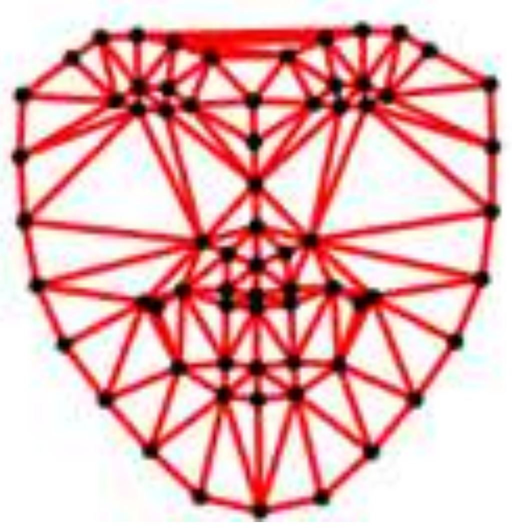
***k* - cascade level**

Updated shape instance

Previous shape instance

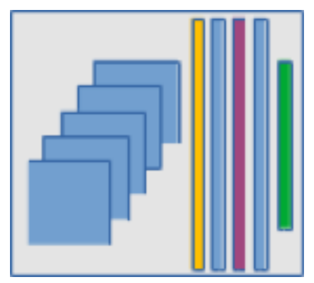
Nonlinear Mapping

Local Feature Extraction at Normalized Frame



$(n+4)\times 1$

$(n+4)\times 1$

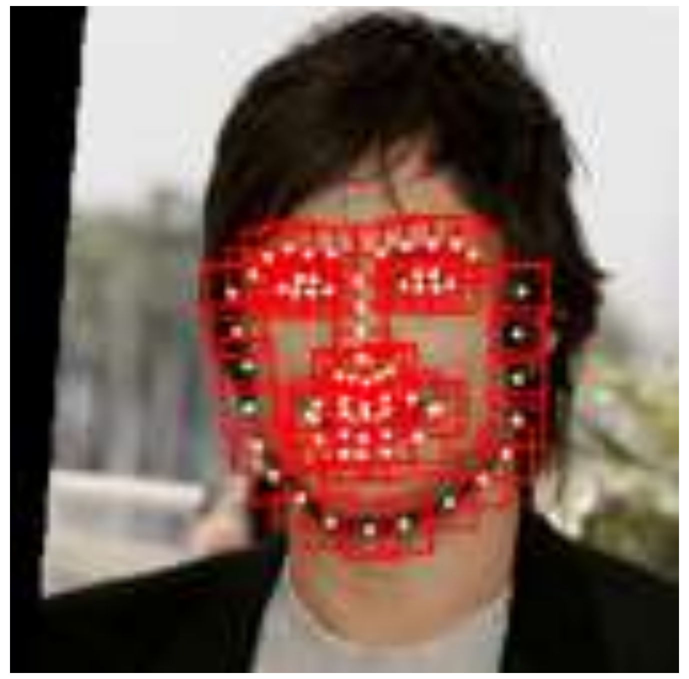


CNN[k]

$\mathbf{I}(.)$

Similarity Warp

$\mathbf{p}(n+1:n+4)$



$\mathbf{I}(\mathcal{S}(\mathbf{p}))$

Sampled 3D Array

$P \times P \times v$

$\mathcal{L}(\mathbf{I}(\mathcal{S}(\mathbf{p})))$

CNN Regression Architecture

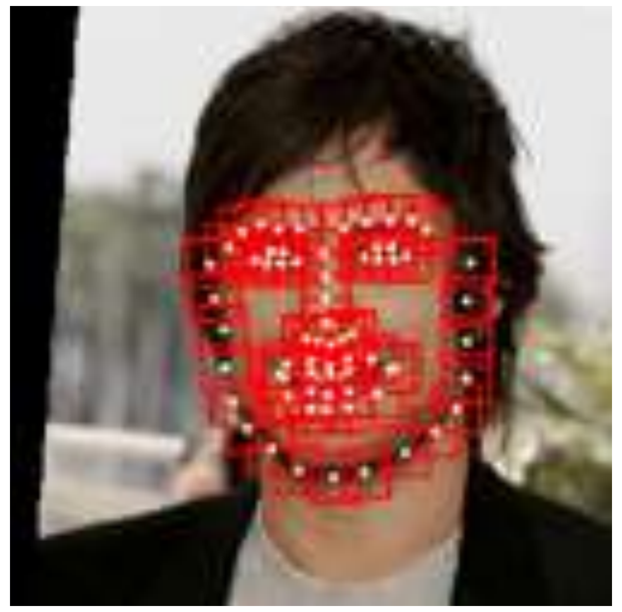
Nonlinear Regression

$$\arg\min_{\mathcal{R}^k} \sum_{i=1}^{N} \sum_{j=1}^{M} \|\Delta \mathbf{p}_{ij}^k - r_L(\ldots r_1(\mathcal{L}(\mathbf{I}_i(\mathcal{S}(\mathbf{p}_j^k)))))\|_{\Sigma^k}^2$$

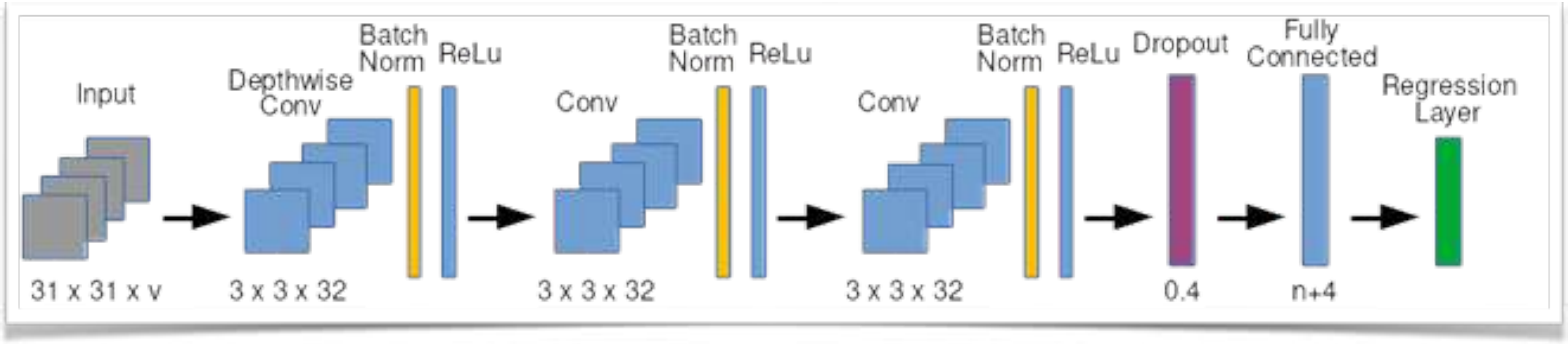
CNN Topology

Input

$\mathcal{L}(\mathbf{I}(\mathcal{S}(\mathbf{p})))$



Pose Normalized Image



Depthwise Convolution 32 filters (3x3) for each local patch

Convolution 32 filters (3x3)

Convolution 32 filters (3x3)

Dropout 0.4

Regression Layer (n+4)

Output

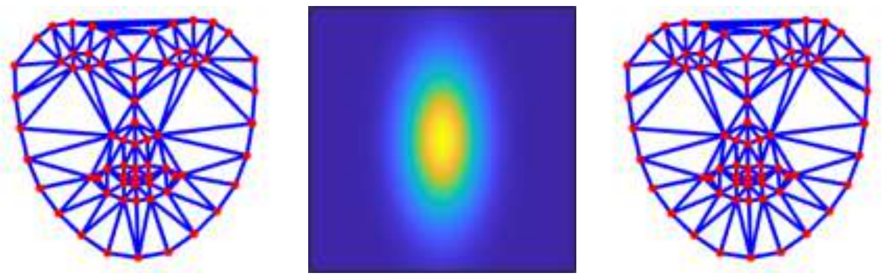
shape parameters update

$\Delta \mathbf{p}$

Loss function

$$L_r = \frac{1}{N} \sum_{j=1}^{N} \Delta \mathbf{p}_j^T \Sigma_{\mathbf{p}}^{-1} \Delta \mathbf{p}_j$$

Mahalanobis Distance



CNN Learning - Data Collection

$$\arg\min_{\mathcal{R}^k} \sum_{i=1}^{N} \sum_{j=1}^{M} \|\Delta \mathbf{p}_{ij}^k - \mathrm{CNN}^k(\mathcal{L}(\mathbf{I}_i(\mathcal{S}(\mathbf{p}_j^k))))\|_{\Sigma^k}^2$$

k - cascade level
i - training image
j - virtual sample

Estimate noise

$$\Sigma^k = \mathrm{cov}(\mathbf{p}_* - \mathbf{p}_{ij})$$

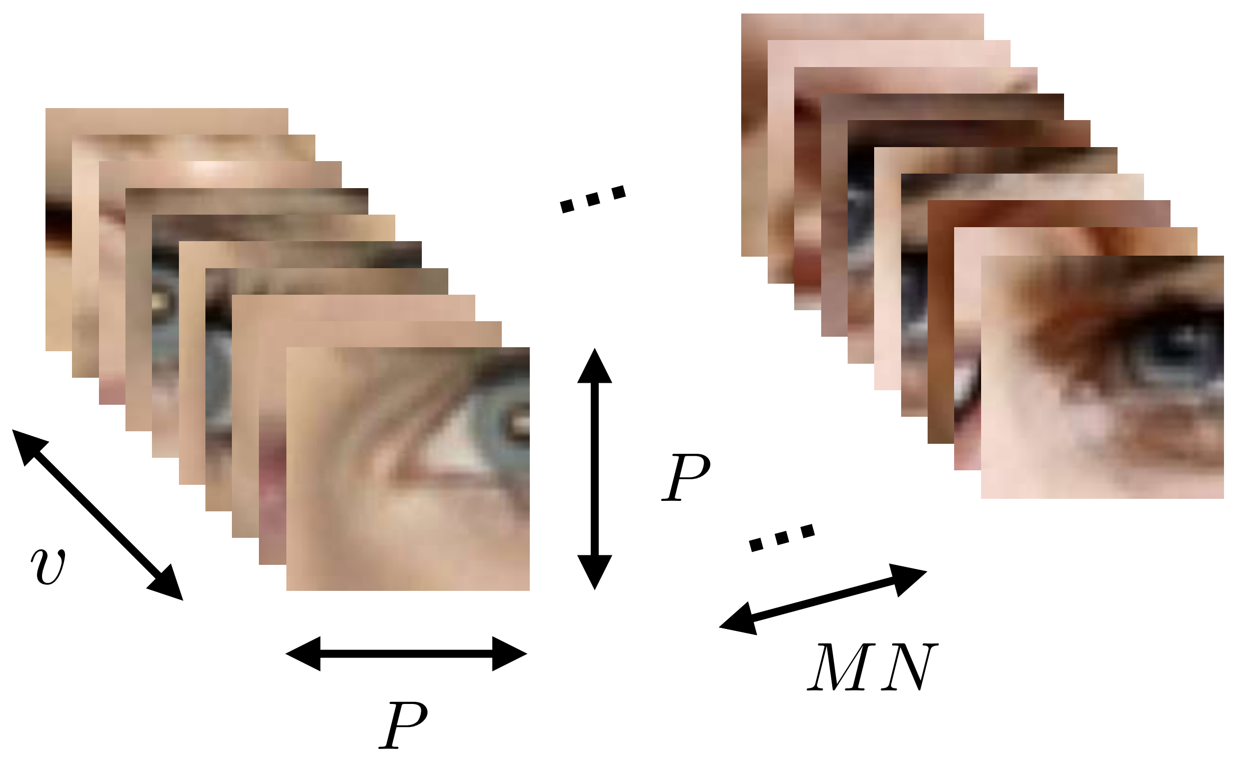
Deviation from Ground Truth

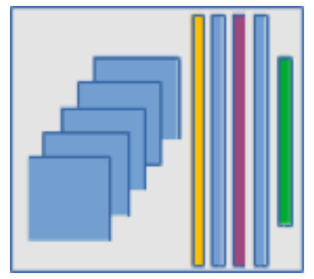
Data Matrix (local normalized patches)

Regression Labels

$$\Delta \mathbf{p}_{ij} = \mathbf{p}_* - \mathbf{p}_{ij}$$

$$\mathbf{D}_{ij} = \mathcal{L}(\mathbf{I}_i(\mathcal{S}(\mathbf{p}_j)))$$

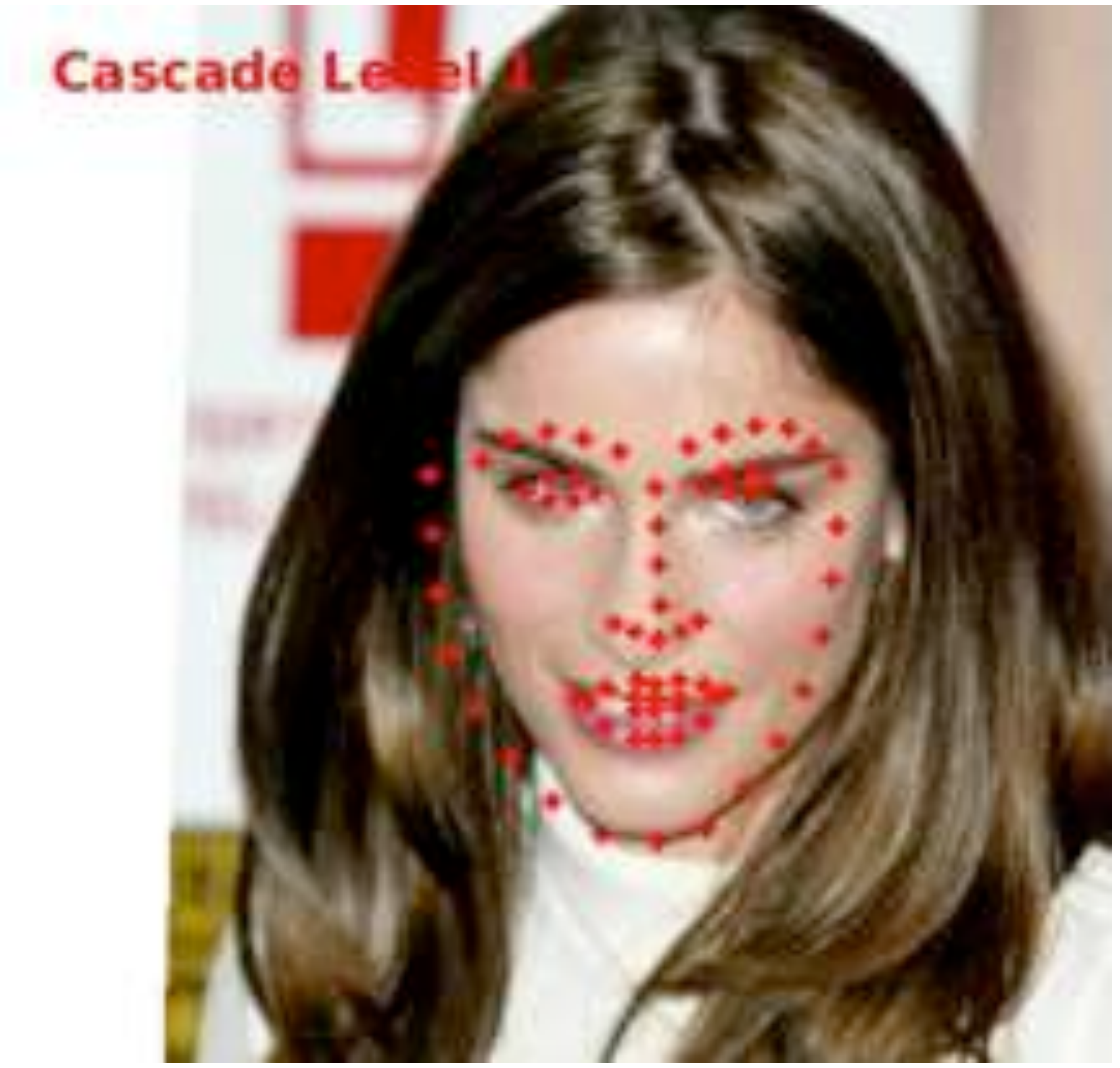




CNN^k

M - augmented examples
N - real examples

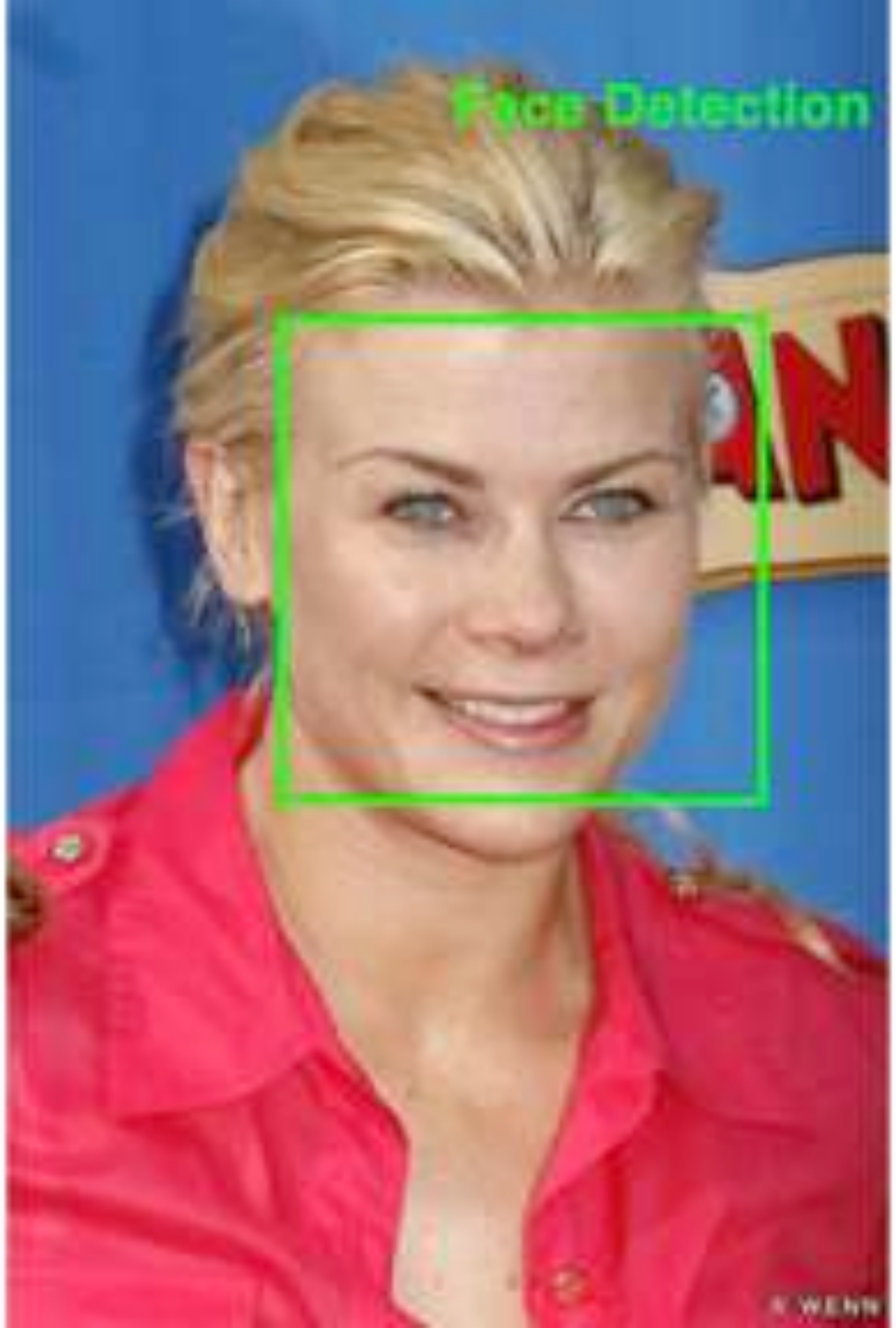
Data Matrix: 4D Array
(P x P x v x N.M)



Data Collection (D matrix)

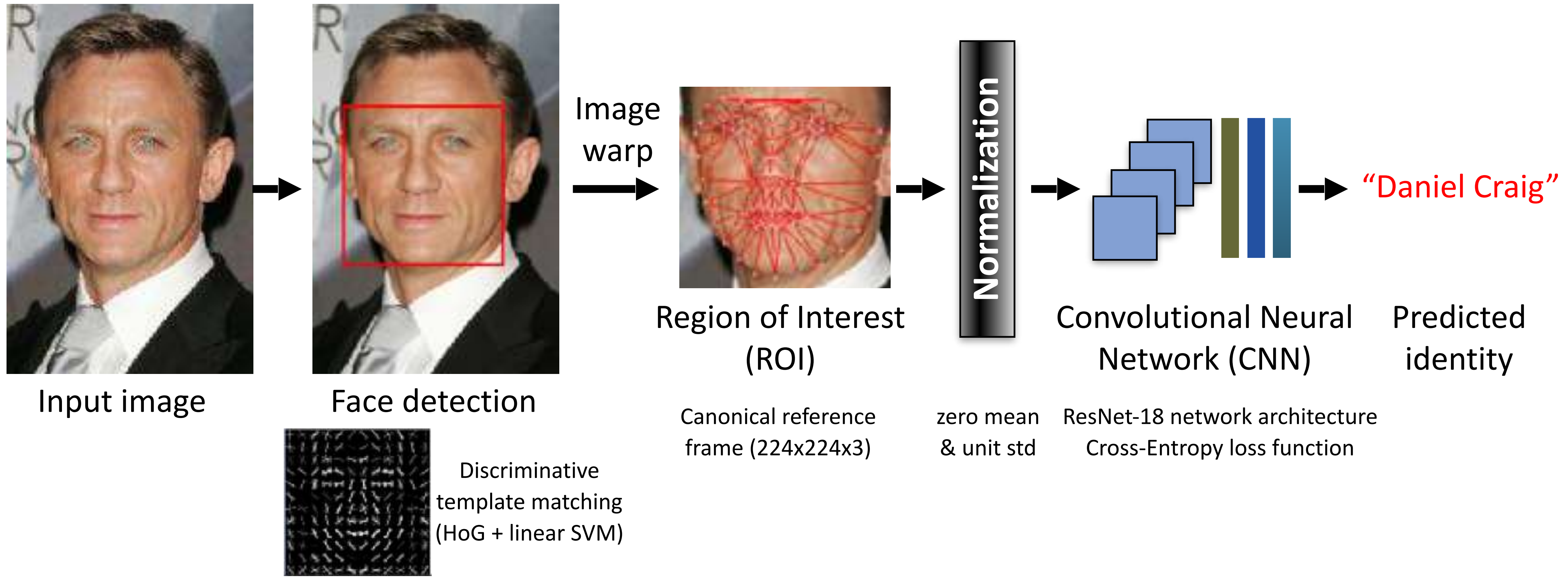
$$\mathbf{p}_{ij} \sim \mathcal{N}(\mathbf{p}_i, \Sigma^k)$$

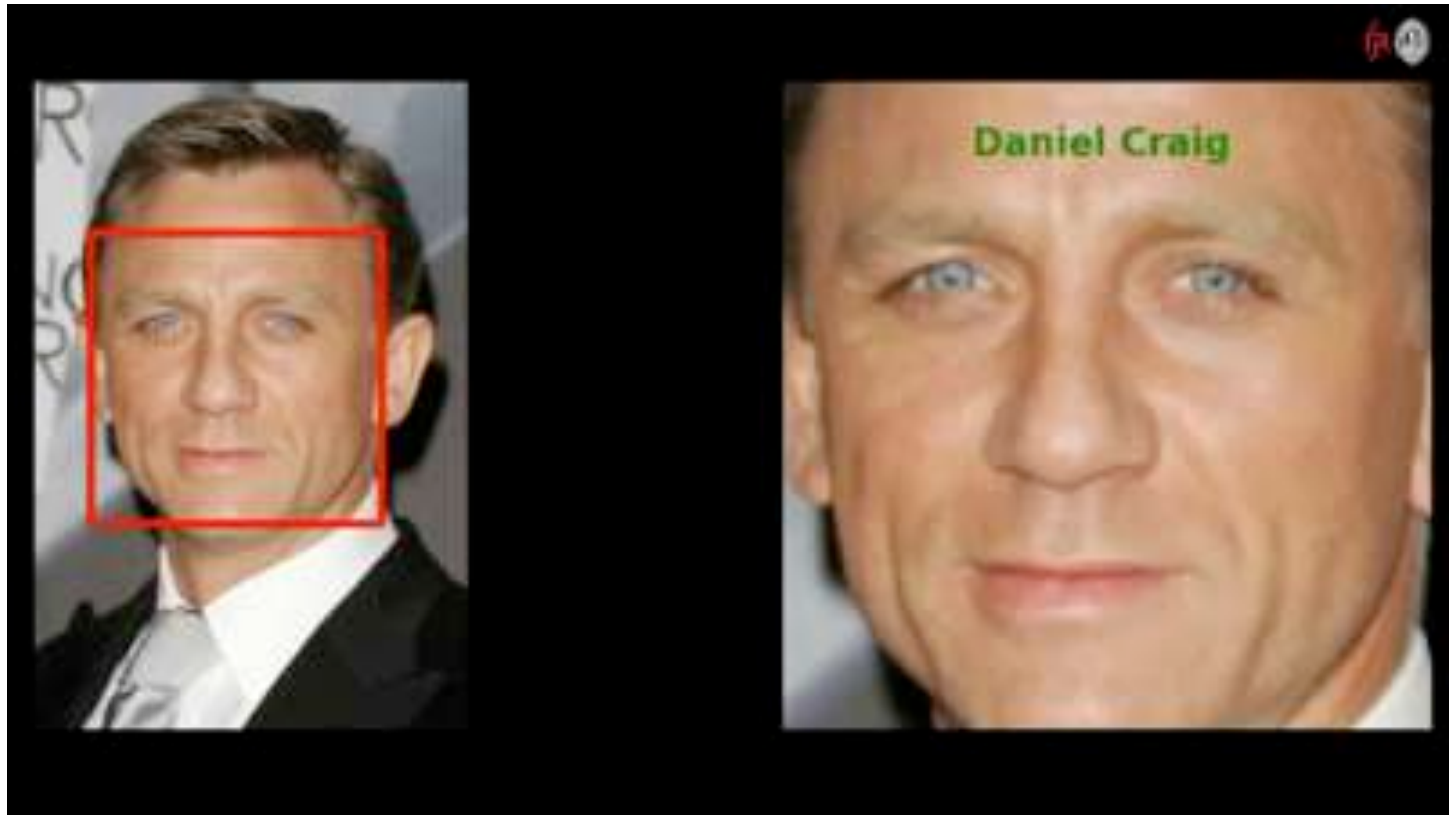
virtual shape sample


Face Detection
© WENN

Demo Applications

Face Recognition



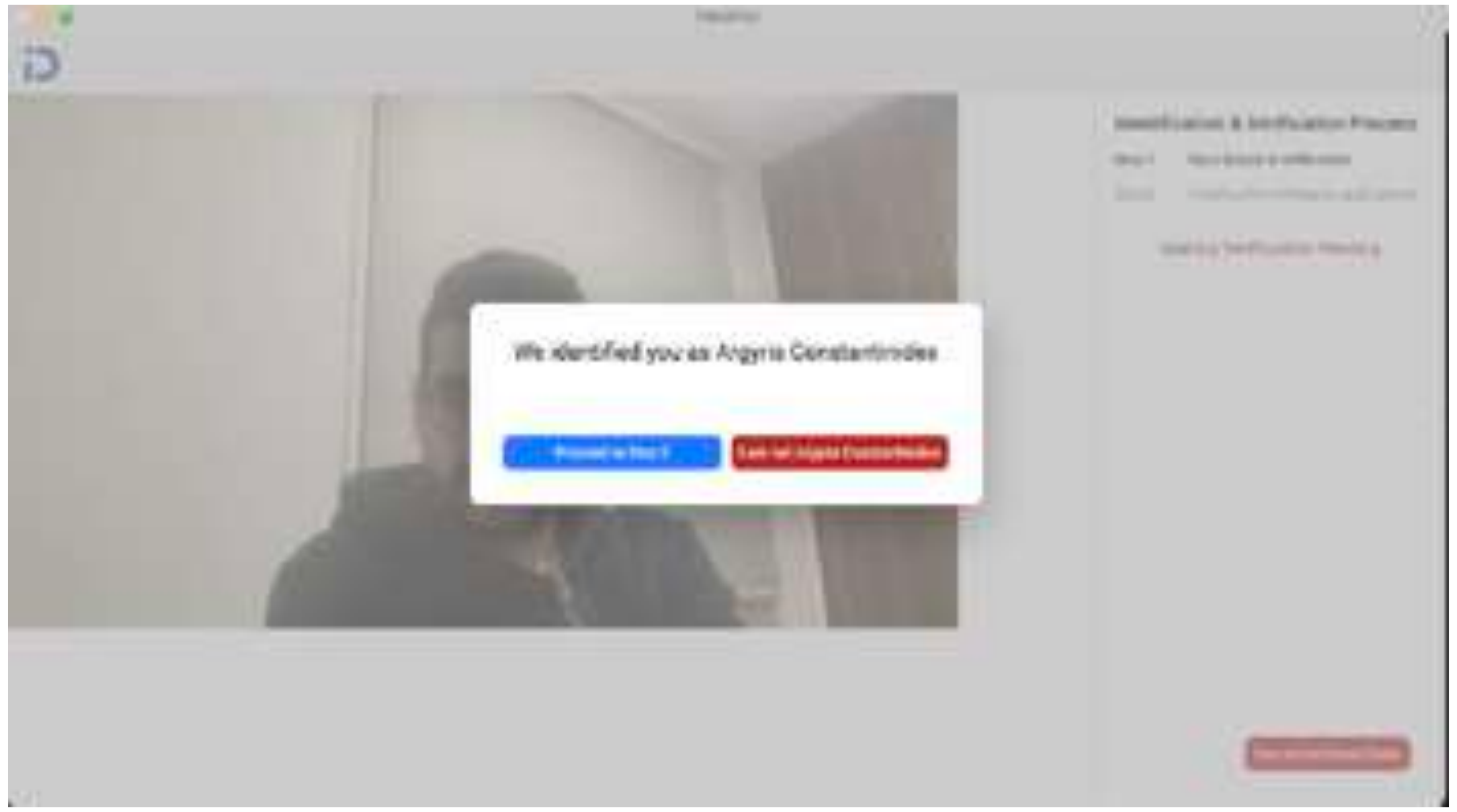


TrustID Project

TRUSTiD
Intelligent Student Identity Management
for Higher Eduation Institutions

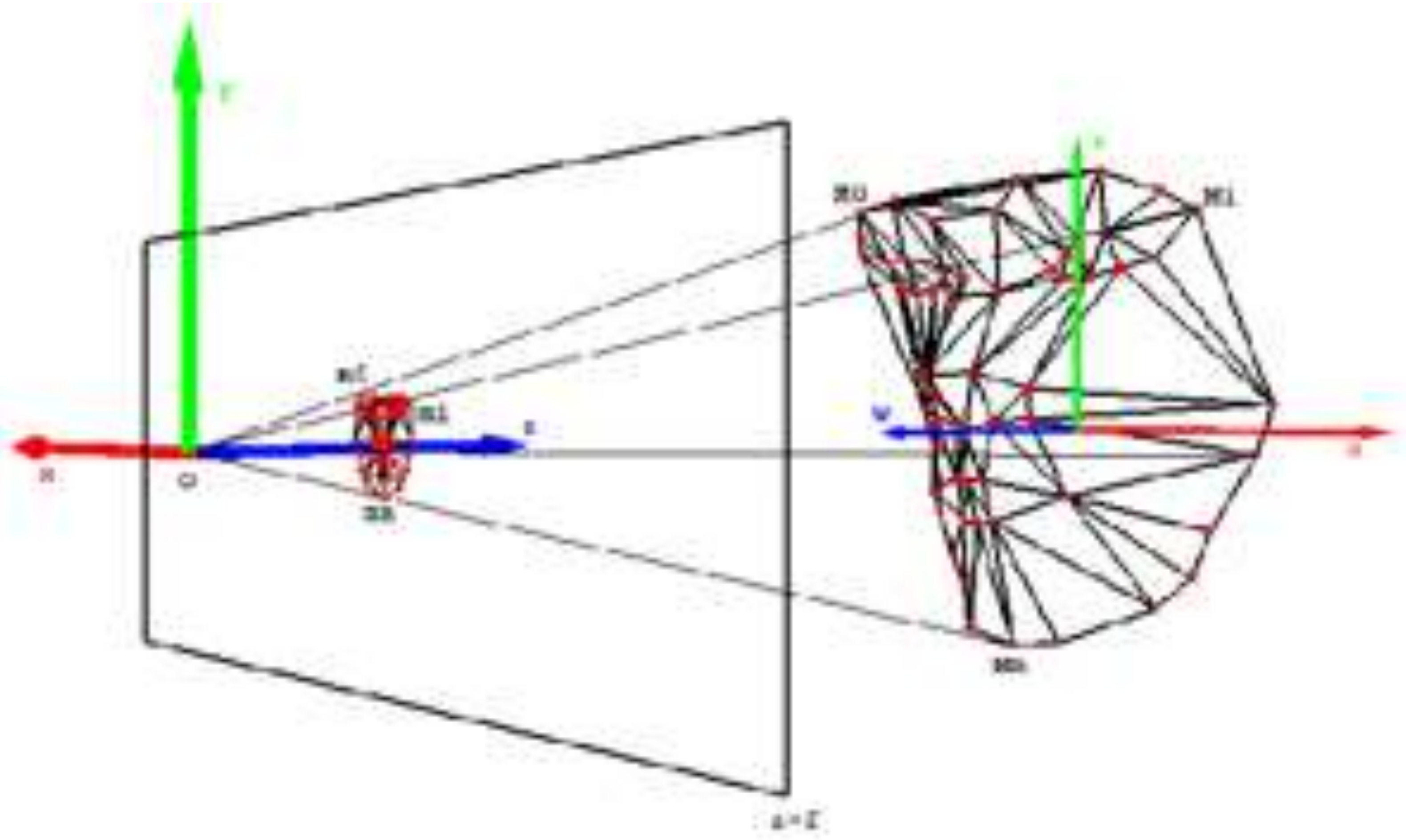
- Intelligent and Continuous Online Student Identity Management for Improving Security and Trust in European Higher Education Institutions.
- https://trustid-project.eu/
- Project approved by the European Commission under the Erasmus+ 2020 program, with a total funding of €291K and two years duration (June/2021 - May/2023).
- Partners:
    - University of Patras (Greece) [coordination]
    - University of Cyprus (Cyprus)
    - **Institute of Systems and Robotics, University of Coimbra (Portugal)**
    - Cognitive UX GmbH (Germany).

Enrol users web page.



User interface.

3D Head Pose Estimation

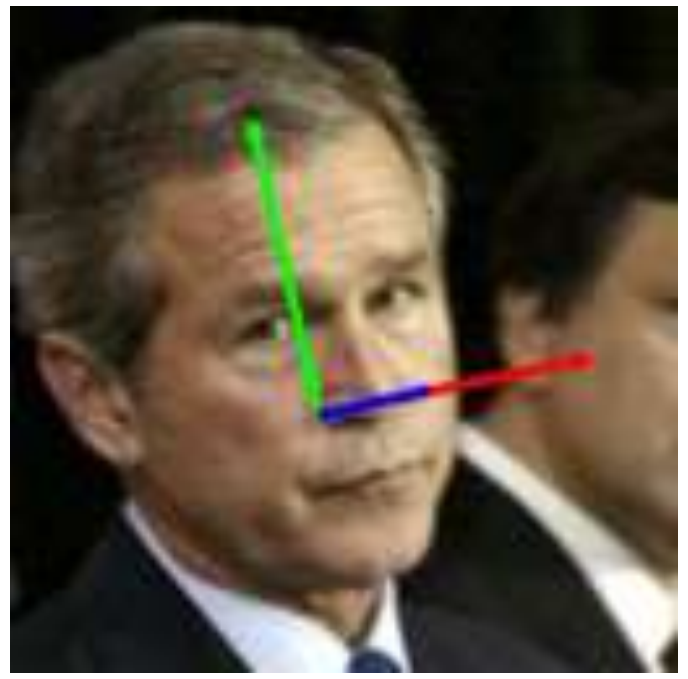


(old) 3D Model

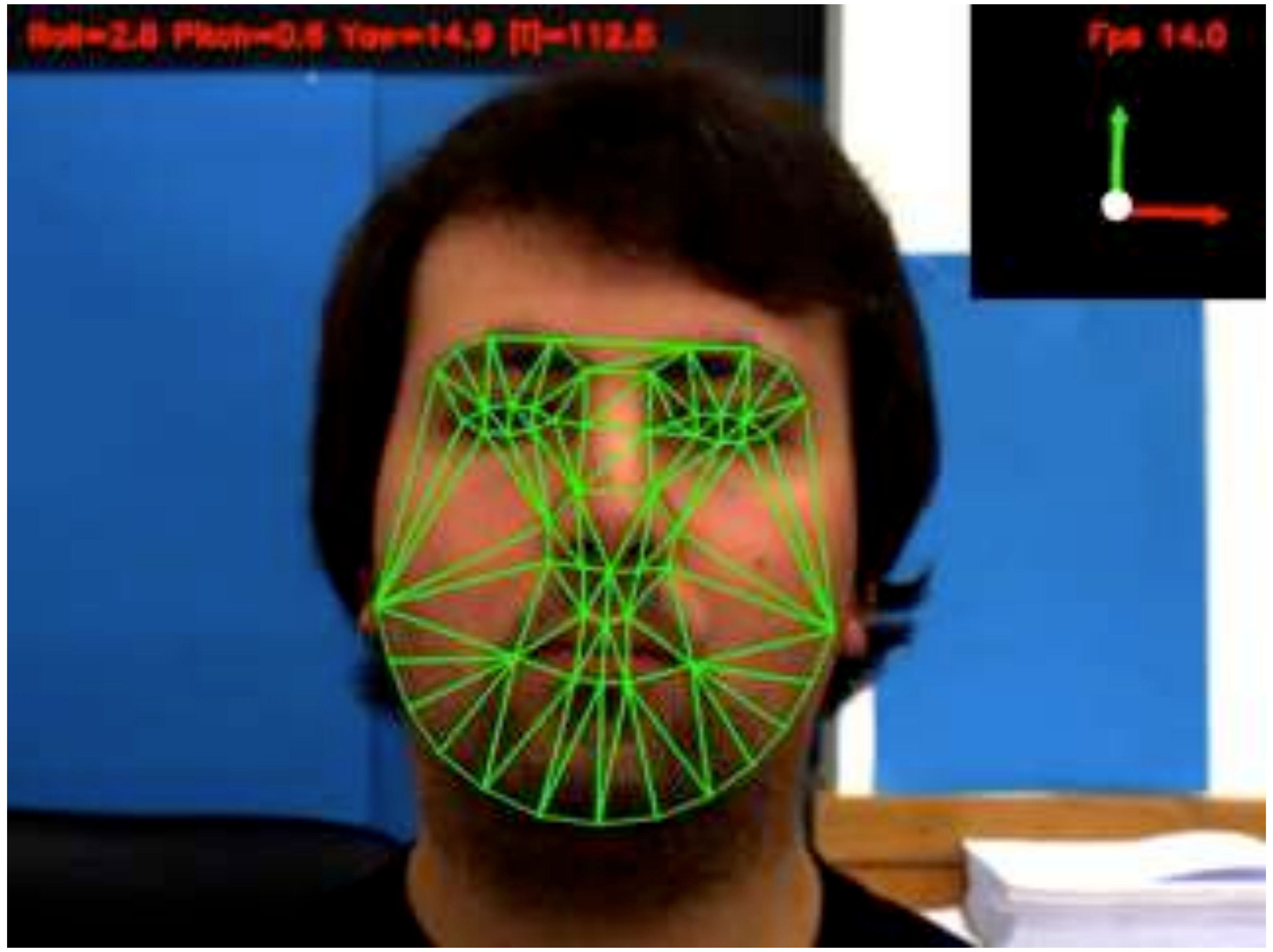
Improved 3D Model

2D landmarks

3D model projection

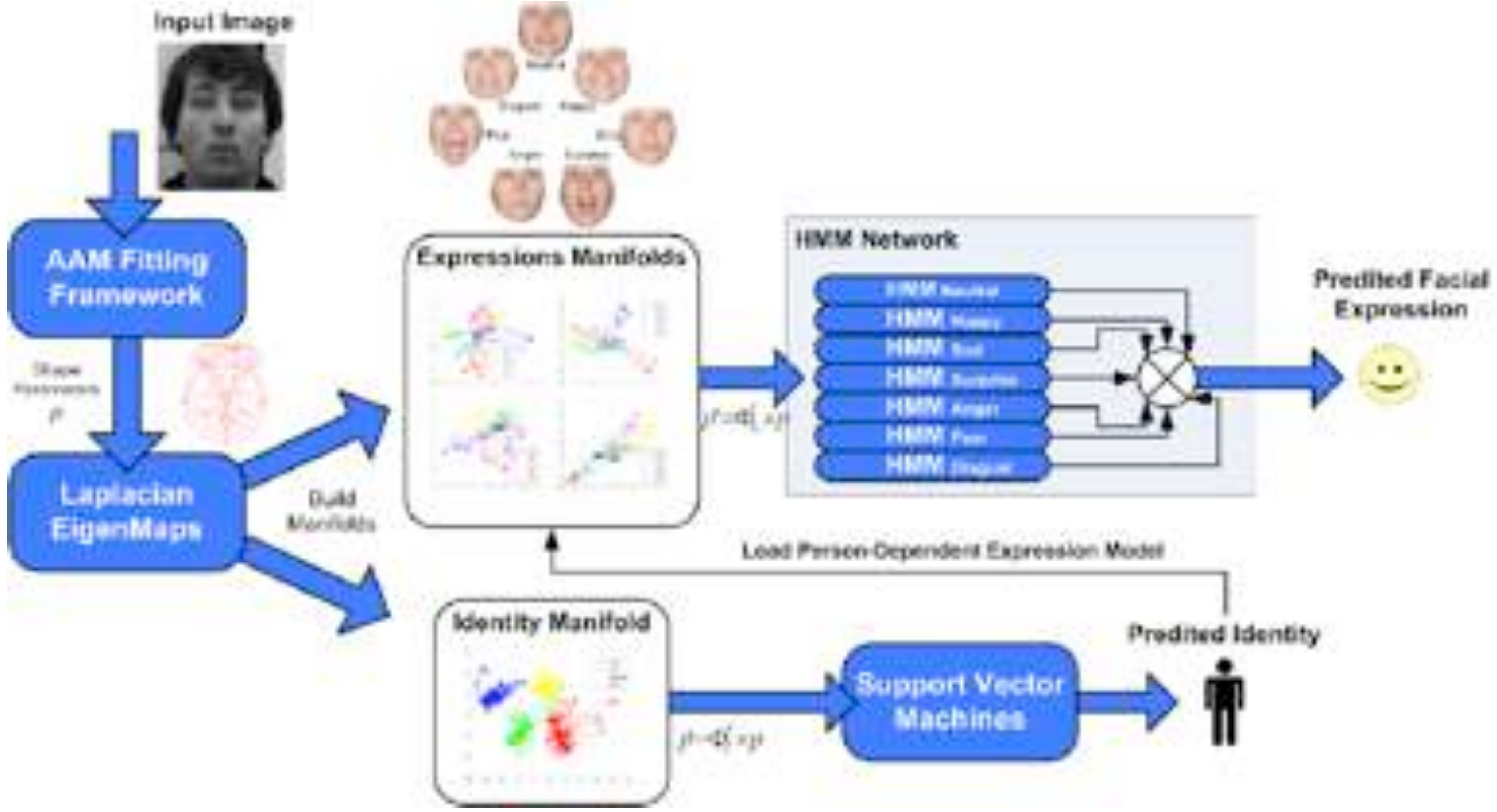


Pose representation

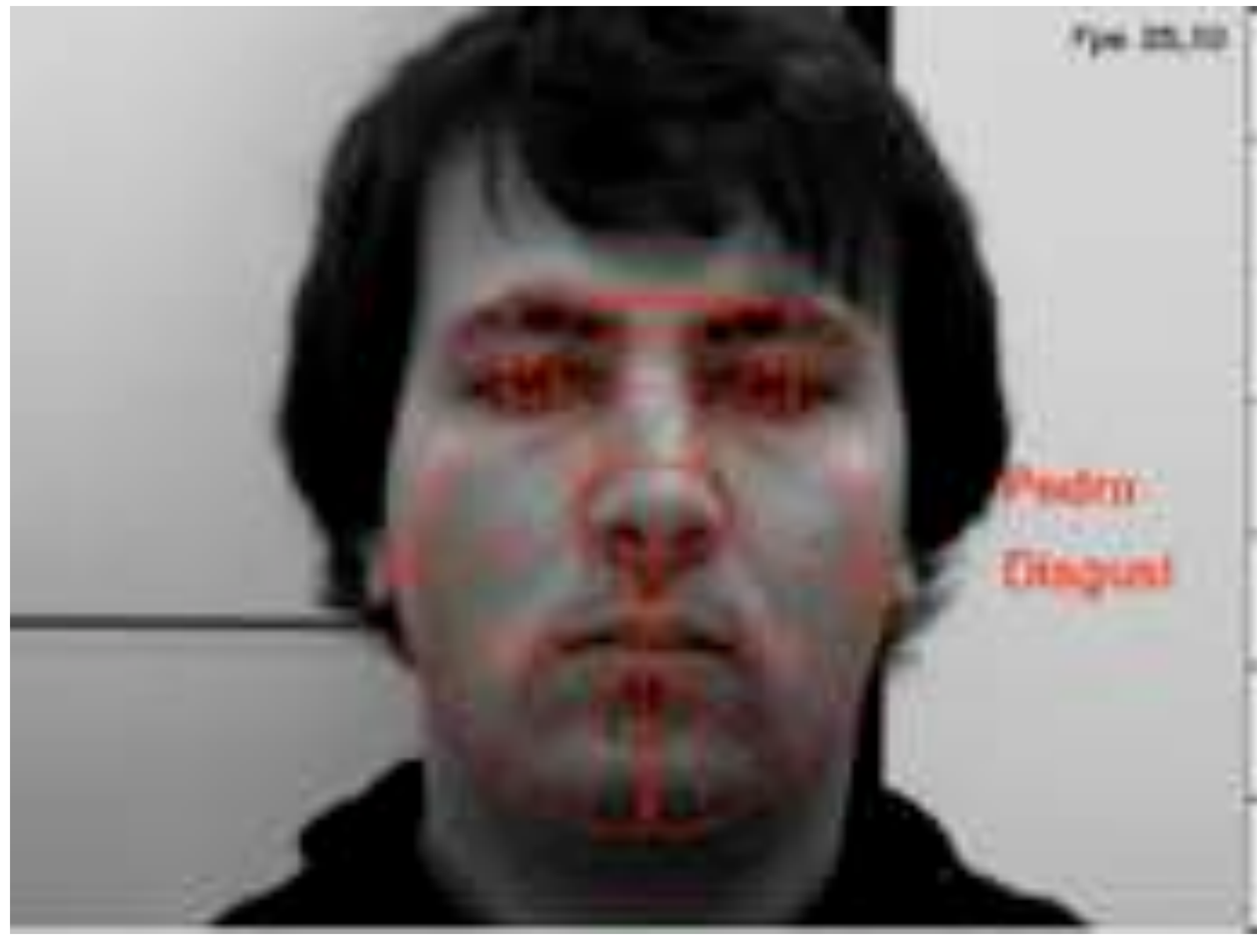


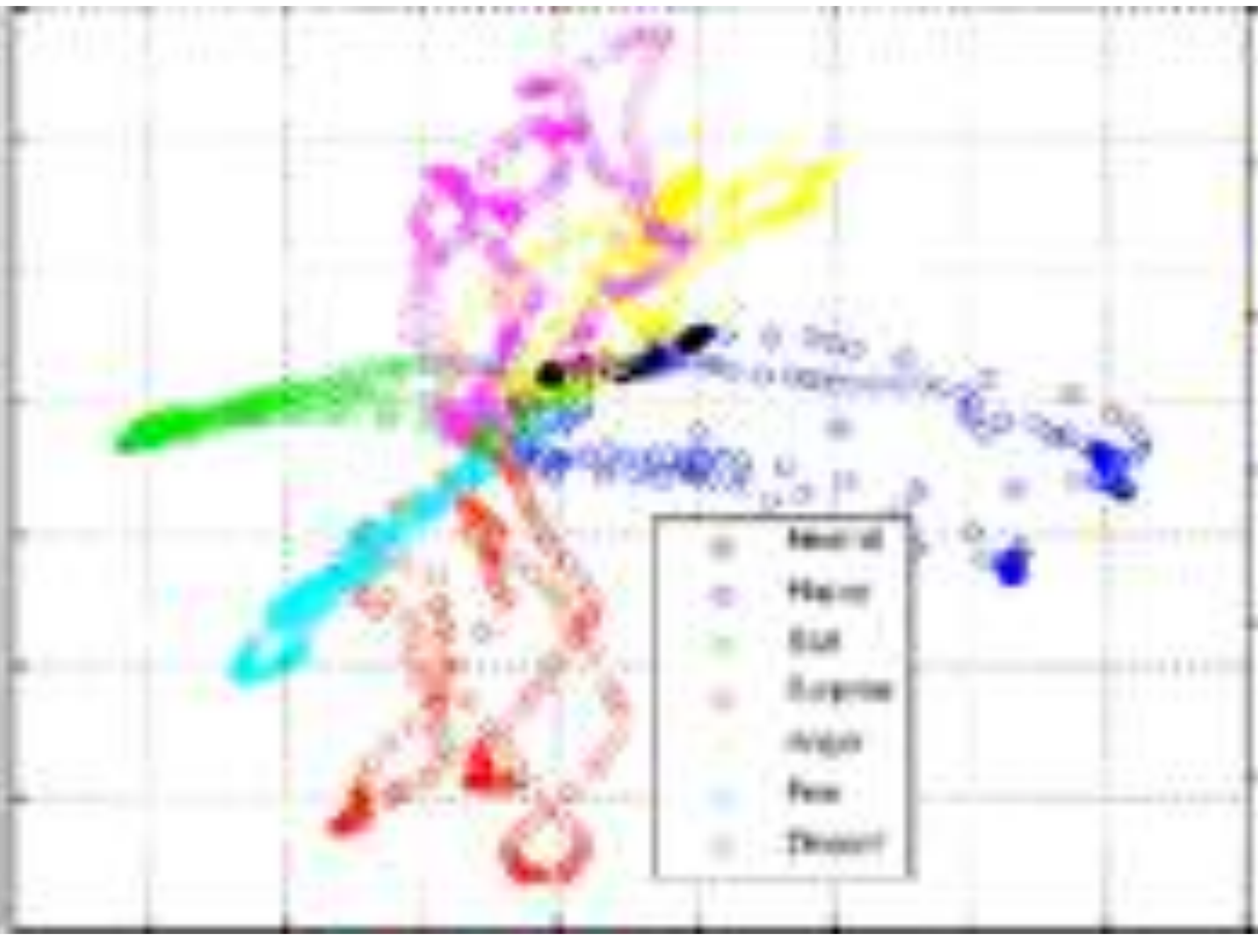

Rx=8.3 Ry=-0.1 Rz=-1.7 d=50.9

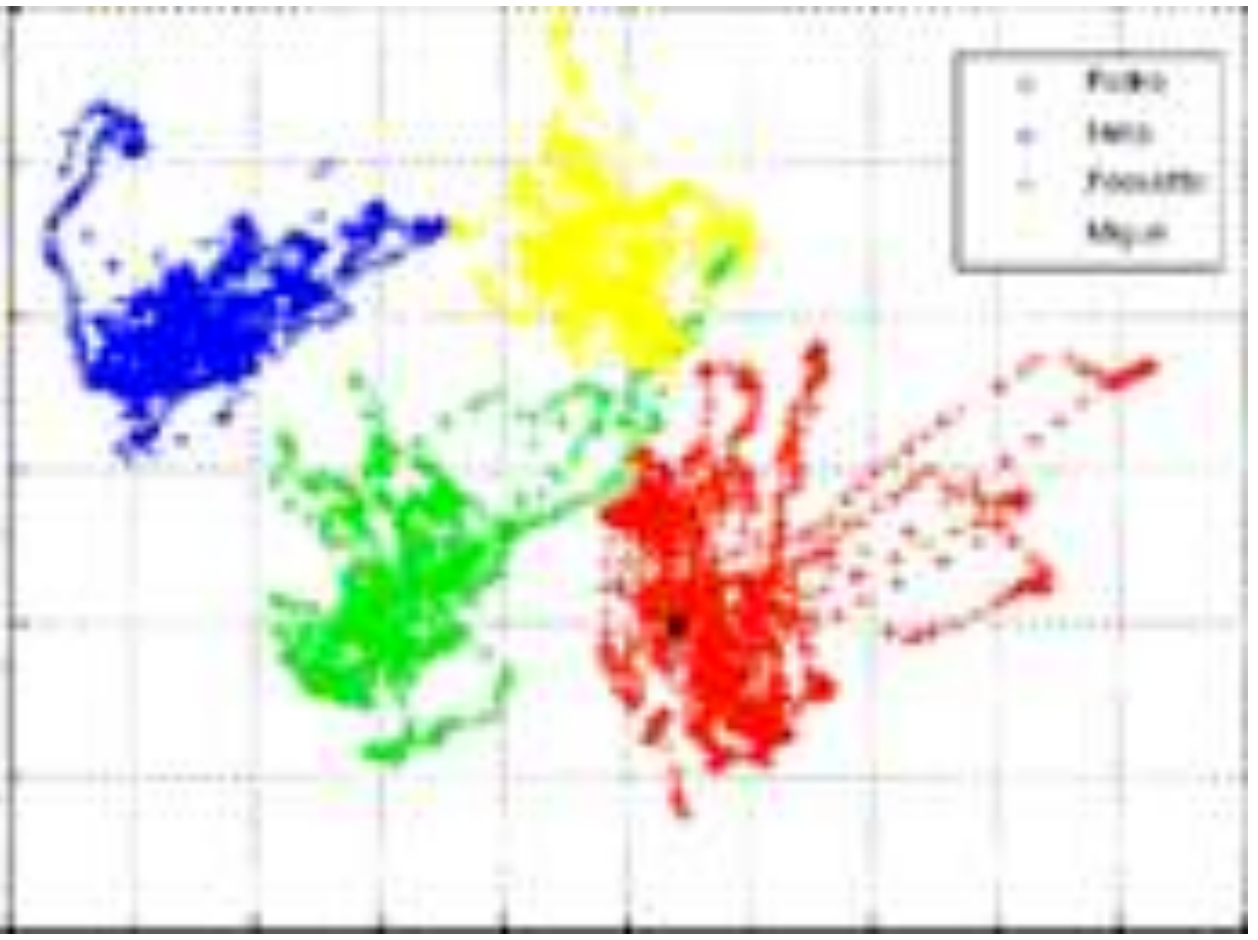
Facial Expression Recognition



Input (SIC Fitting) **Expression (HMM)** **Identity (SVM)**







Face Swapping w/ Blending

Age Estimation

- UTK Faces Database.
- 20K+ images of faces in the wild
- 200x200 RGB images
- CNN Regression (ResNet18)

Demo 3D Head Pose Estimation - Super Mario World

Thank you

https://www.isr.uc.pt/~pedromartins
pedromartins@isr.uc.pt