Bayesian Constrained Local Models with Depth Data

Pedro Martins, João Faro, Patrick Brandão, Jorge Batista
http://www.isr.uc.pt/~pedromartins
pedromartins@isr.uc.pt

Institute of Systems and Robotics (ISR)
University of Coimbra
Portugal

IEEE International Conference on Image Processing (ICIP) 2016, Phoenix, AZ, USA

Introduction

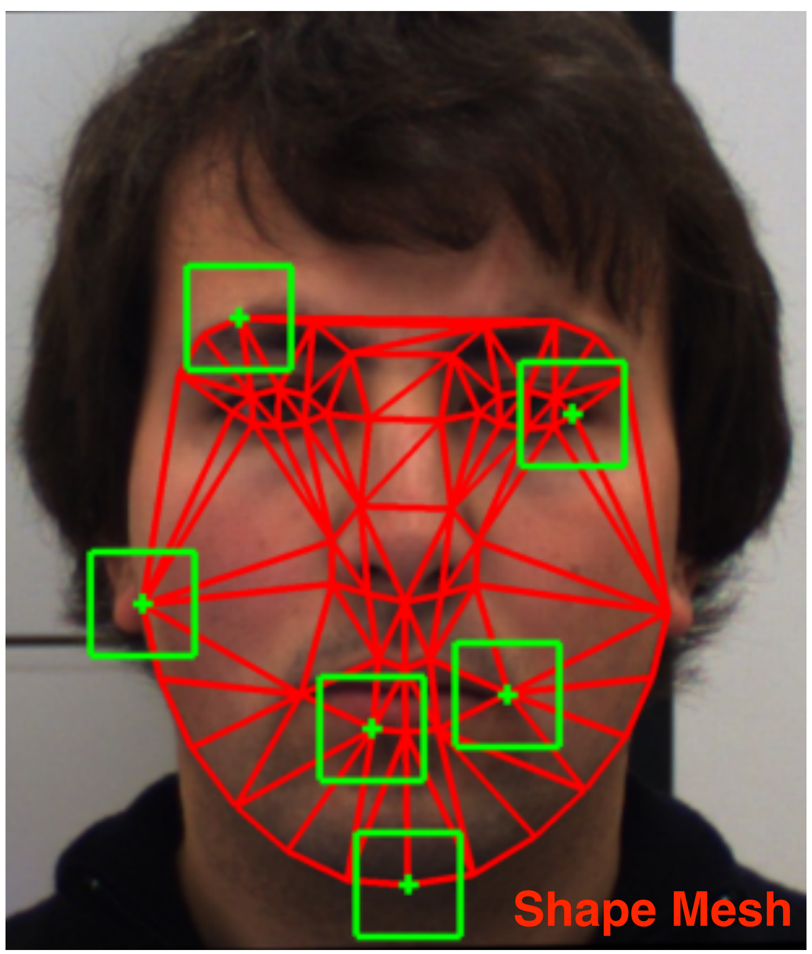
- Constrained Local Model (CLM) Extension
- Fitting with Intensity and Depth Data
- Bayesian CLM Framework
- Likelihood Fusion Strategies
- Evaluation Results (EURECOM and ISR-Z datasets)

CLM Fitting with RGBD Data (video)

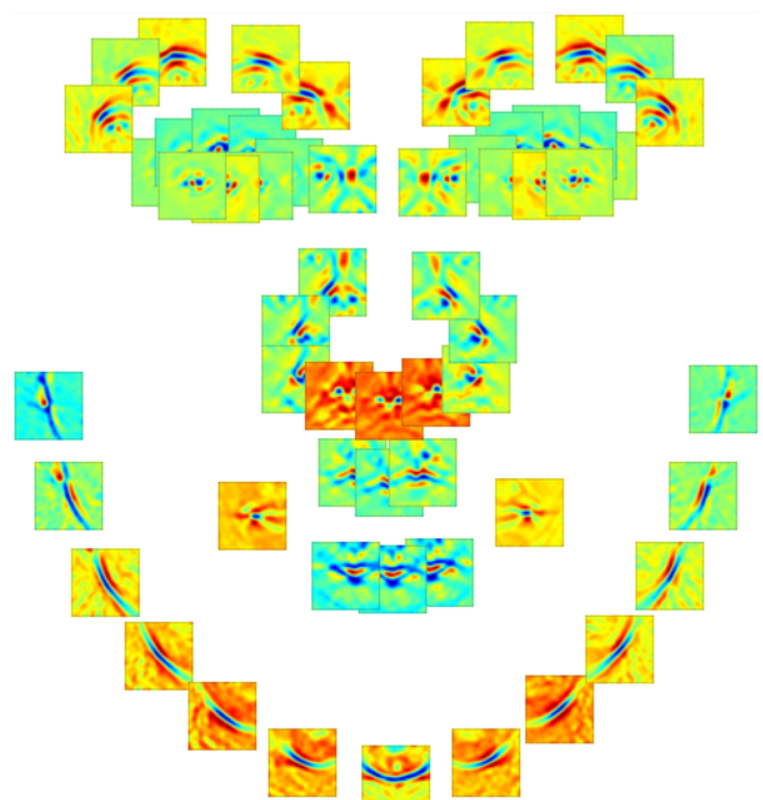
Constrained Local Model (CLM)

Data Term **Regularization Term**

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

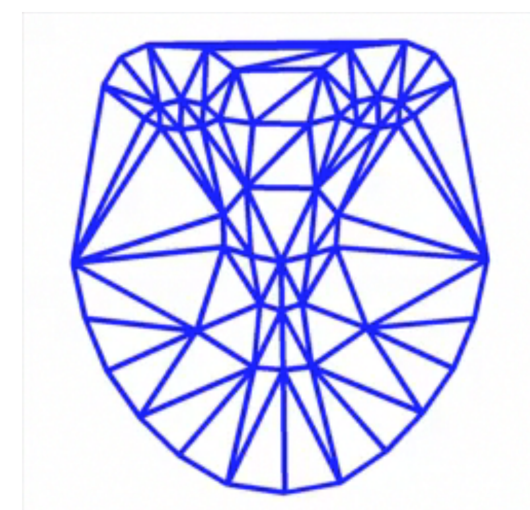


Local Search Regions



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

Local Detectors



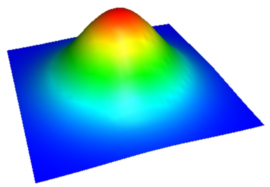
$$\mathbf{s} = \mathbf{s}_0 + \sum_{j=1}^{n} b_j \phi_j$$

Shape Model

Bayesian Inference CLM

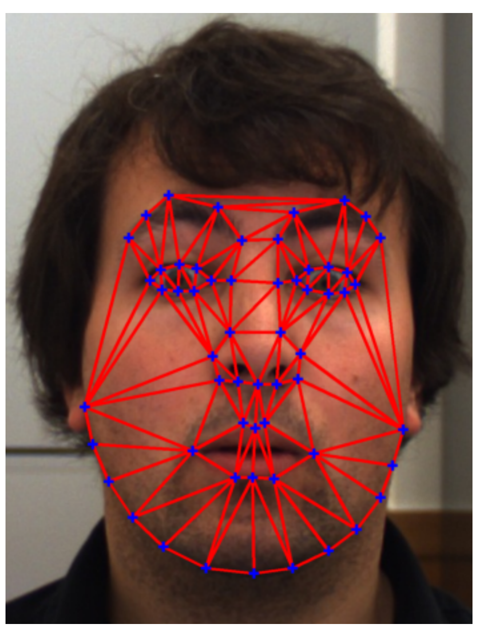
$$\hat{\mathbf{b}} = \arg\max_{\mathbf{b}} p(\mathbf{b}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{b})p(\mathbf{b})$$

Likelihood Term

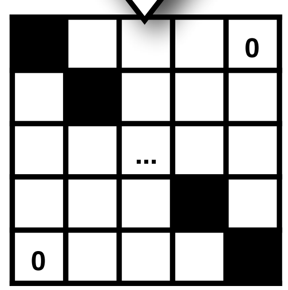


$$p(\mathbf{y}|\mathbf{b}) \propto \exp\left(-\frac{1}{2}(\mathbf{y} - (\mathbf{s}_0 + \Phi\mathbf{b}))^T \Sigma_{\mathbf{y}}^{-1} (\mathbf{y} - (\mathbf{s}_0 + \Phi\mathbf{b}))\right)$$

Shape Observation



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

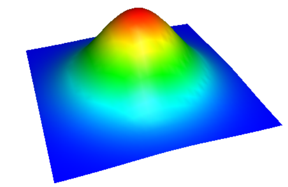


$2v \times 2v$

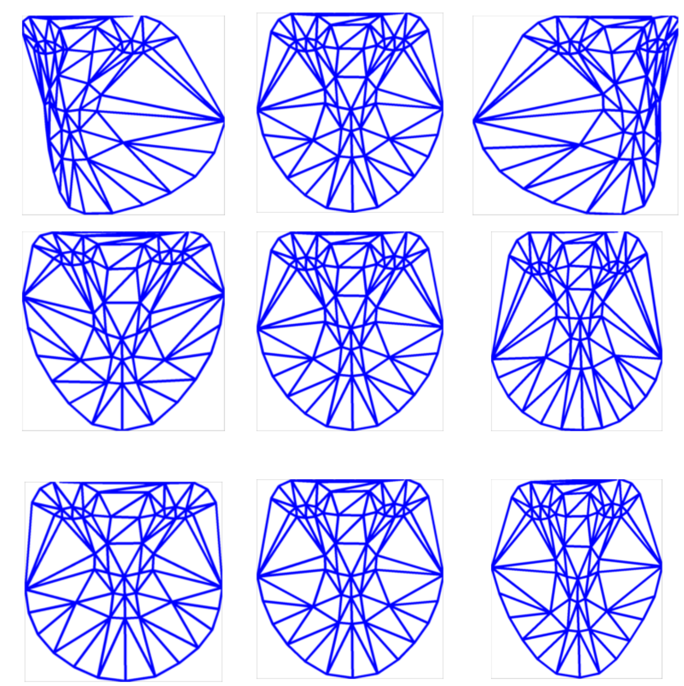
Block diagonal

Uncertainty Covariance

Prior Term



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

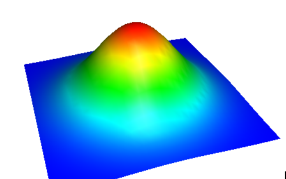


Linear Dynamic System (LDS)

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

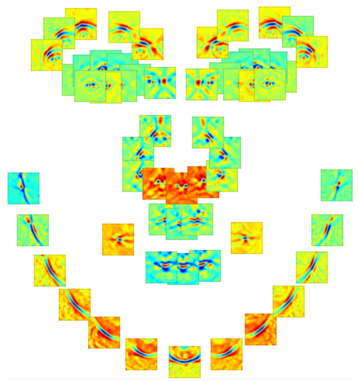
Posterior Term

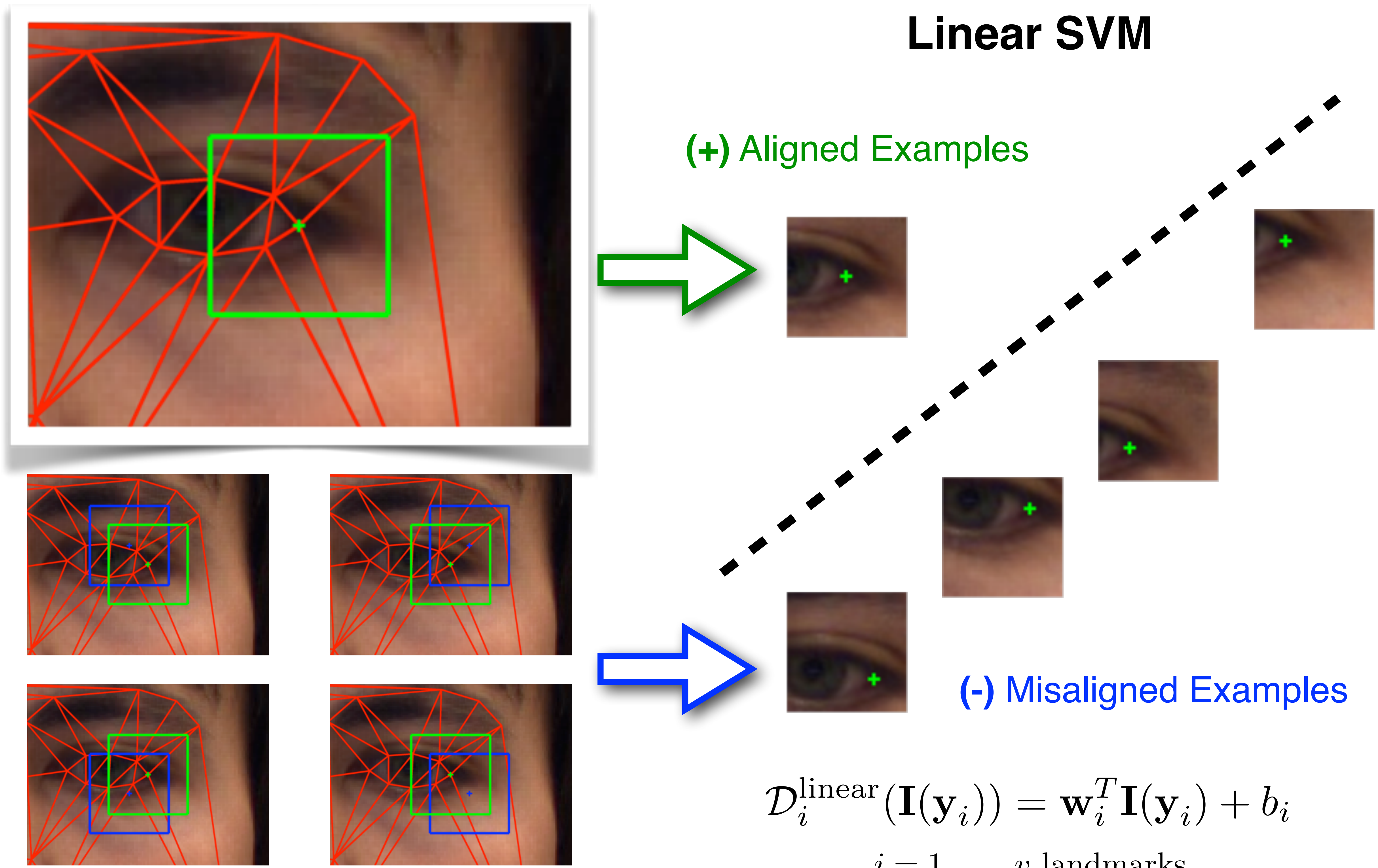
$$p(\mathbf{b}_l|\mathbf{y}_l, \ldots, \mathbf{y}_0) \propto \mathcal{N}(\mathbf{b}_l|\boldsymbol{\mu}_l^{\mathbf{F}}, \boldsymbol{\Sigma}_l^{\mathbf{F}})$$



[P. Martins et.al., ECCV 2012]
[P. Martins et. al., TPAMI 2016]

Local Landmark Detectors (SVM)

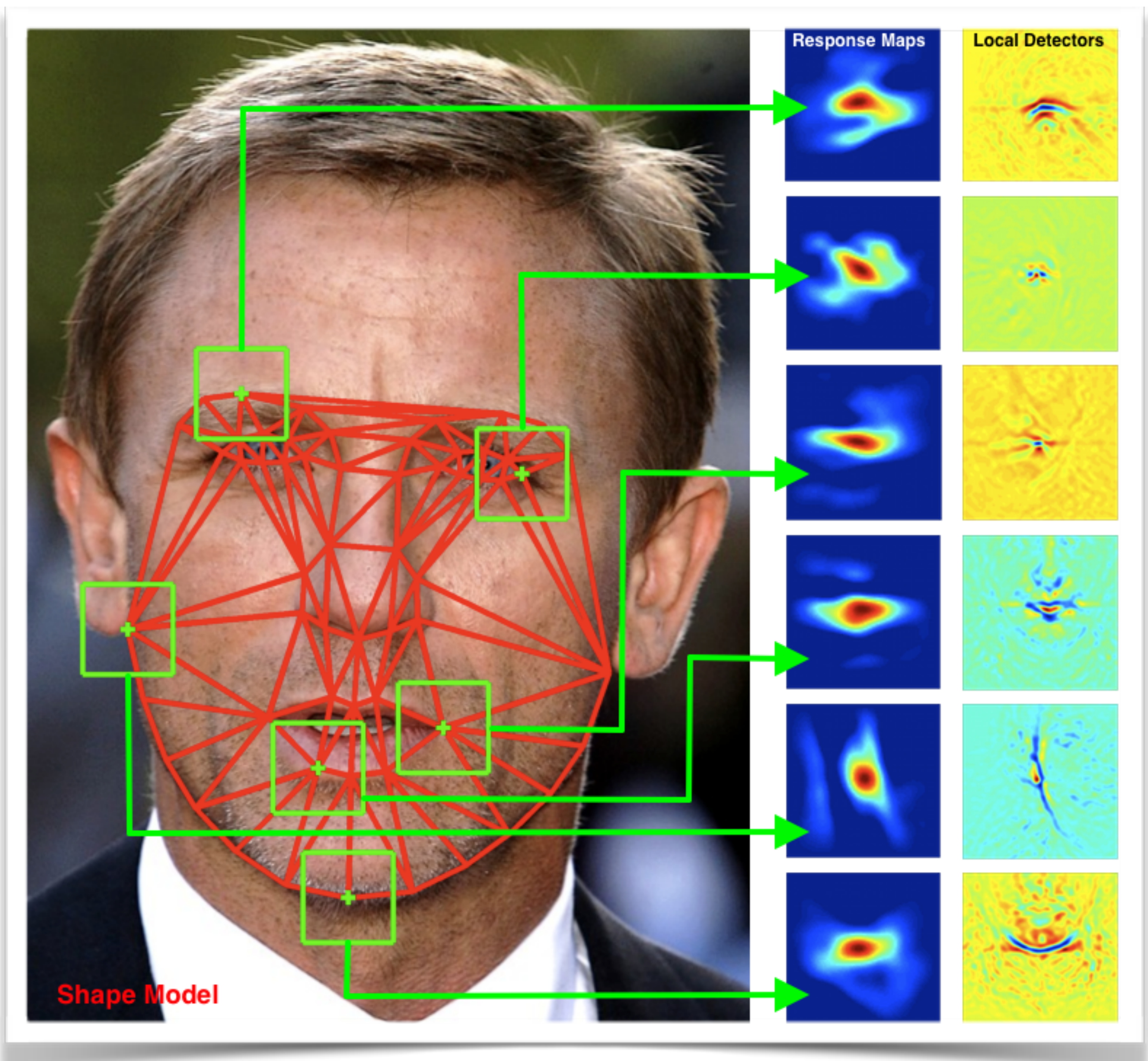




$$\mathcal{D}_i^{\text{linear}}(\mathbf{I}(\mathbf{y}_i)) = \mathbf{w}_i^T \mathbf{I}(\mathbf{y}_i) + b_i$$

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

Local Landmark Detectors (MOSSE Filters)



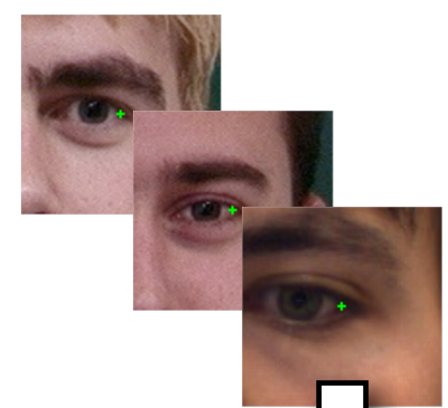
$p(\mathbf{z}_i)$ $\mathbf{h}_i$

Response Map

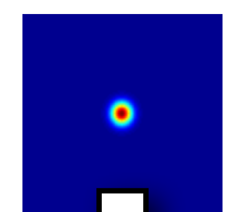
$$p(\mathbf{z}_i) = p(a_i = 1 | \mathcal{D}_i, \mathbf{I}(\mathbf{z}_i)) = \frac{1}{1 + e^{-a_i \beta_1 \mathcal{D}_i(\mathbf{I}(\mathbf{z}_i)) + \beta_0}}$$

Regression Problem

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



Gaussian Target

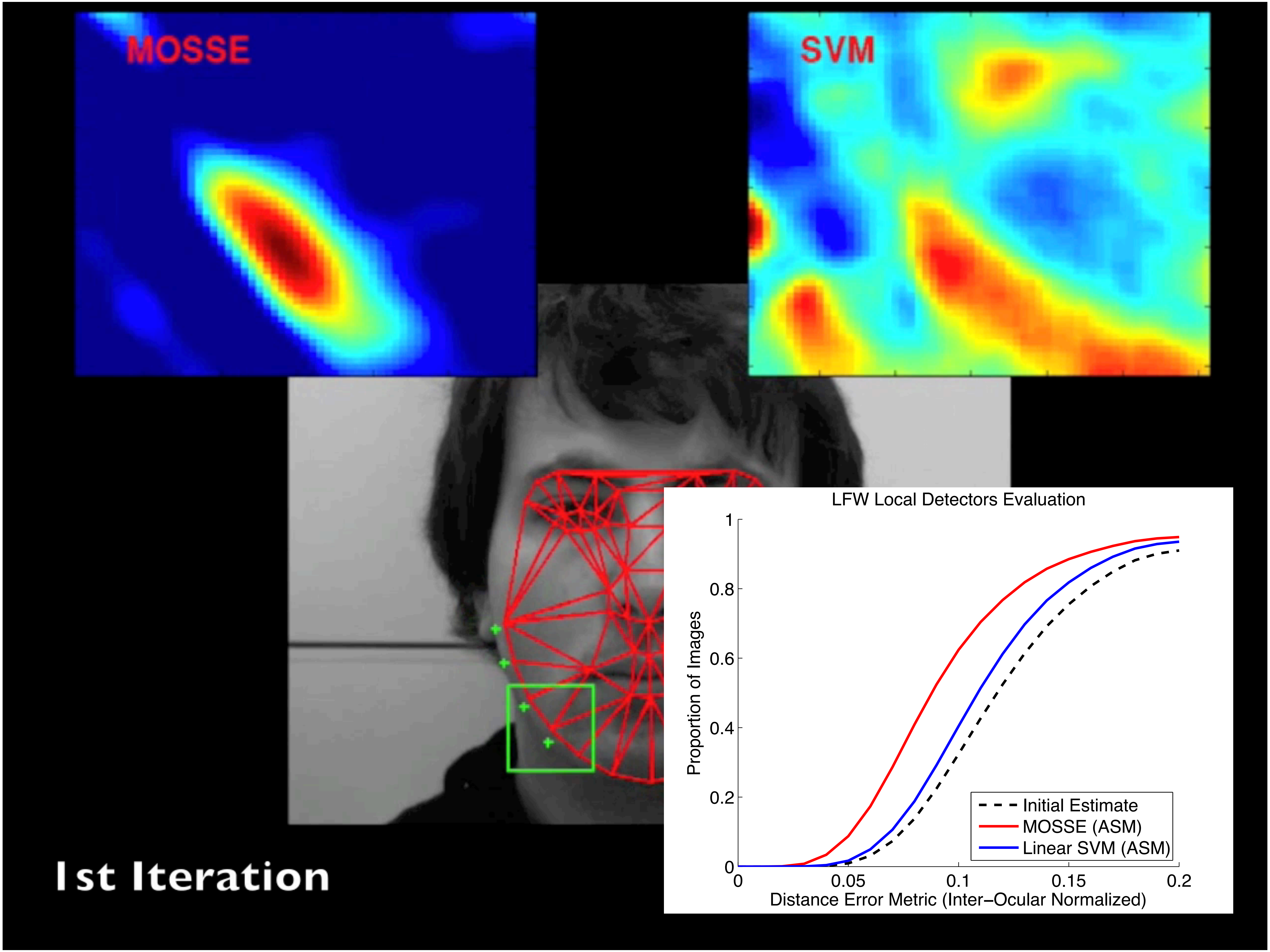


$$\min_{\mathbf{H}^\dagger} \sum_{j=1}^{N} \left(\mathcal{F}\{\mathbf{I}_j\} \odot \mathbf{H}_i^\dagger - \mathcal{F}\{\mathbf{g}_j\}\right)^2 + \lambda ||\mathbf{H}_i||^2$$

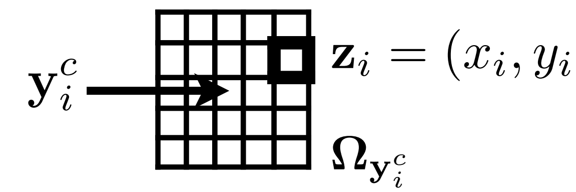
[D. Bolme et.al., CVPR 2010]

Solution

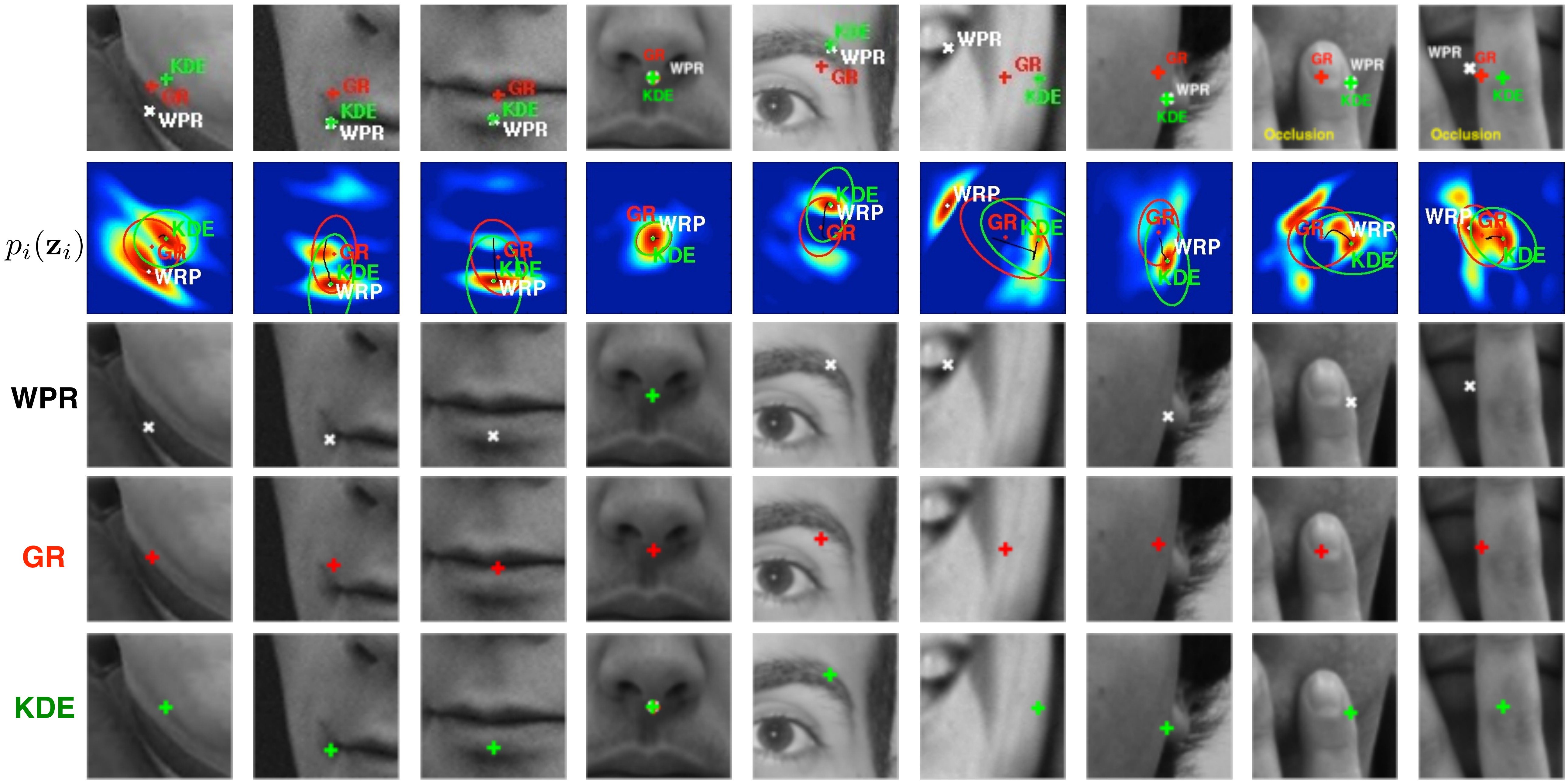
$$\mathbf{h}_i = \mathcal{F}^{-1}\left\{\frac{\sum_{j=1}^{N} \mathcal{F}\{\mathbf{g}_j\} \odot \mathcal{F}\{\mathbf{I}_j\}^\dagger}{\sum_{j=1}^{N} \mathcal{F}\{\mathbf{I}_j\} \odot \mathcal{F}\{\mathbf{I}_j\}^\dagger + \lambda}\right\}^\dagger$$


MOSSE
SVM
LFW Local Detectors Evaluation
Proportion of Images
0
0.2
0.4
0.6
0.8
1
0
0.05
0.1
0.15
0.2
Distance Error Metric (Inter−Ocular Normalized)
Initial Estimate
MOSSE (ASM)
Linear SVM (ASM)
1st Iteration

Local Optimization Strategies



Patches under occlusion



Weighted Peak Response

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

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

Gaussian Response

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

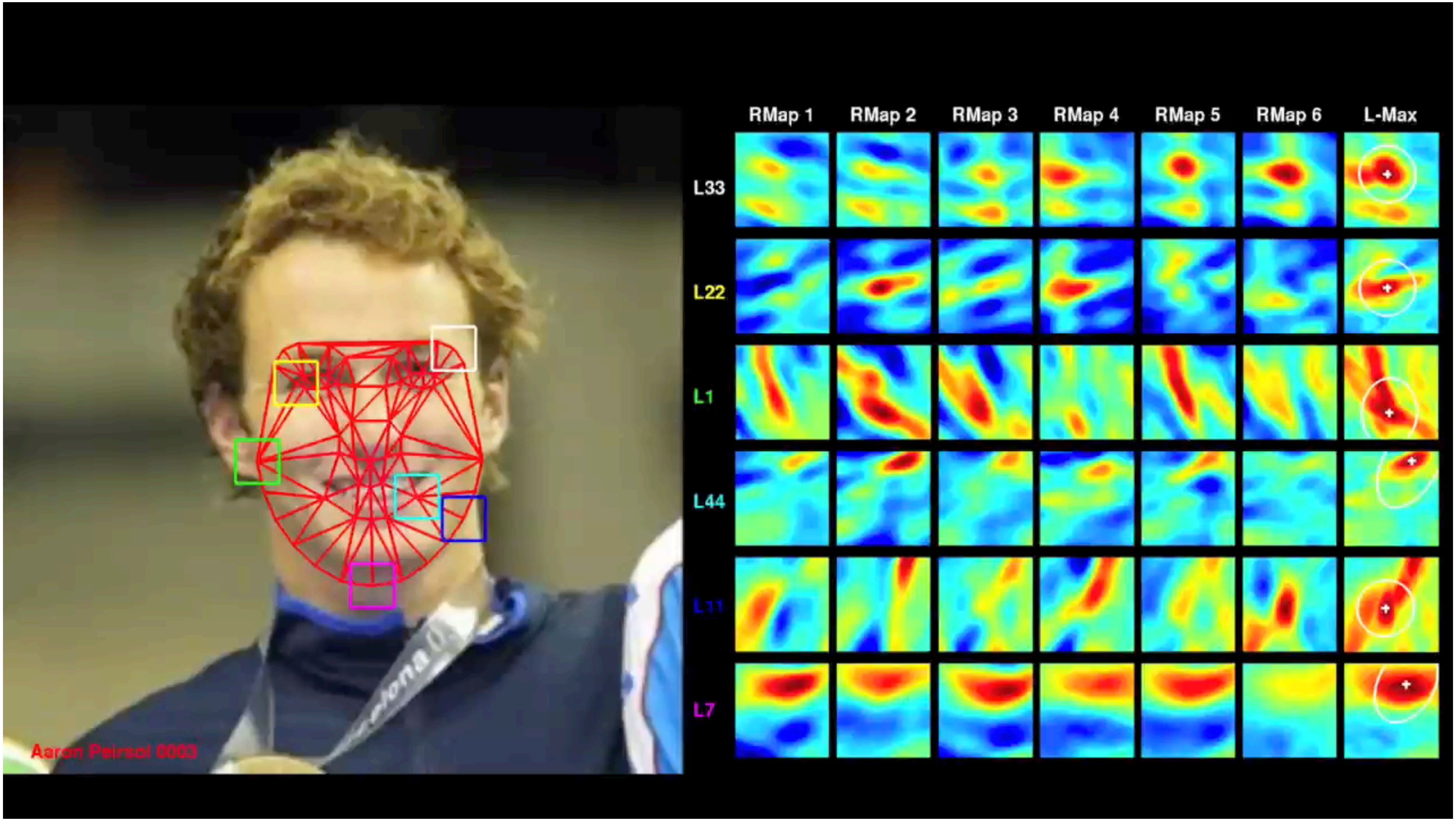
$$\Sigma_{\mathbf{y}_i}^{\text{GR}} = \frac{1}{d-1} \sum_{\mathbf{z}_i \in \mathbf{\Omega}_{\mathbf{y}_i^c}} p_i(\mathbf{z}_i)(\mathbf{z}_i - \mathbf{y}_i^{\text{GR}})(\mathbf{z}_i - \mathbf{y}_i^{\text{GR}})^T$$

Kernel Density Estimator

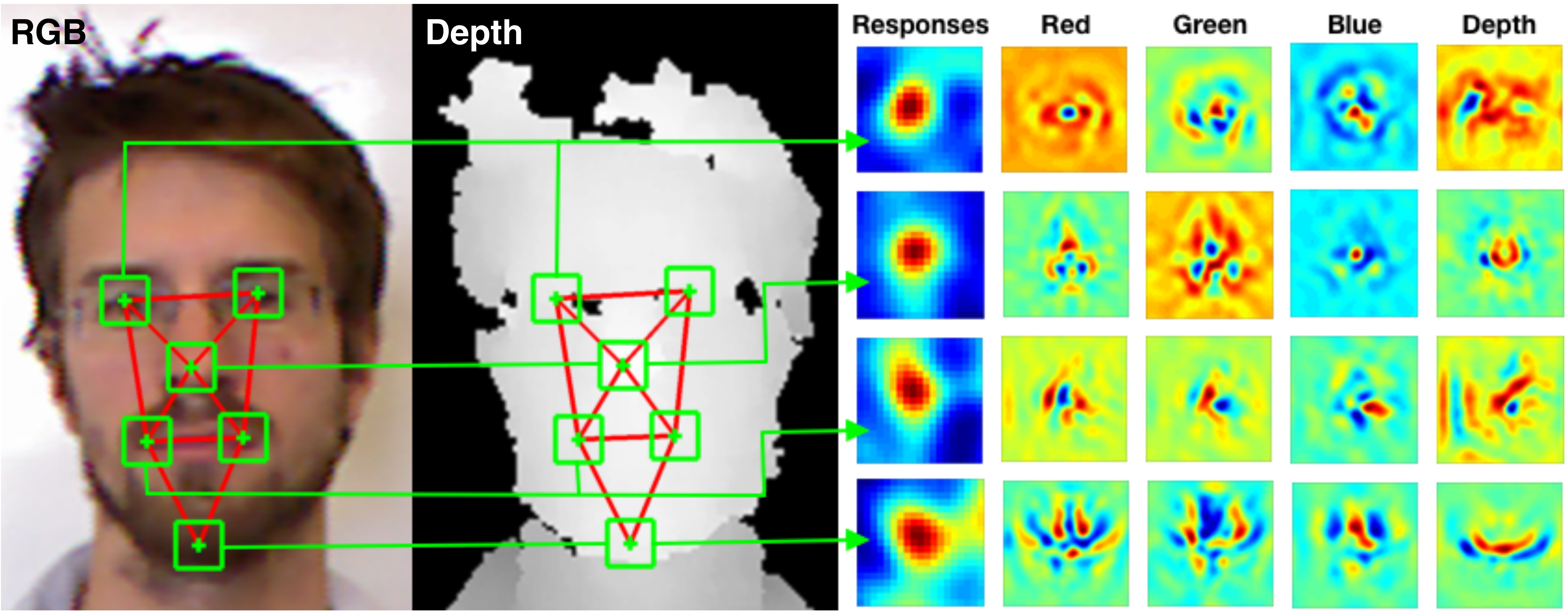
$$\mathbf{y}_i^{\text{KDE}(\tau+1)} \leftarrow \frac{\sum_{\mathbf{z}_i \in \mathbf{\Omega}_{\mathbf{y}_i^c}} \mathbf{z}_i \; p_i(\mathbf{z}_i) \; \mathcal{N}(\mathbf{y}_i^{\text{KDE}(\tau)} | \mathbf{z}_i, \sigma_{h_j}^2 \mathbf{I}_2)}{\sum_{\mathbf{z}_i \in \mathbf{\Omega}_{\mathbf{y}_i^c}} p_i(\mathbf{z}_i)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

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

Multiple Detectors per Landmark (video)

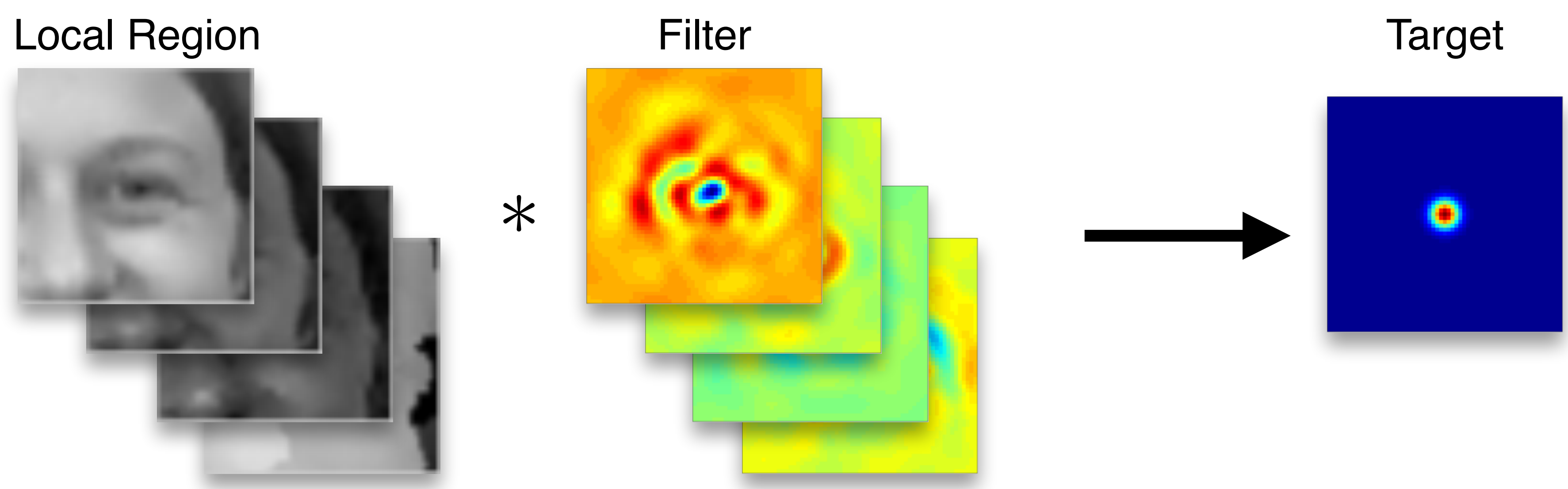


CLM with Depth Data



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

Multiple Channel Correlation Filters



Spatial Domain

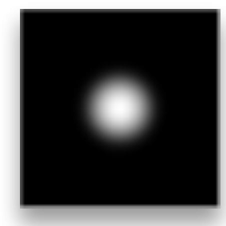
$$\arg \min_{\mathbf{h}_i^{(1)},\ldots,\mathbf{h}_i^{(D)}} \sum_{j=1}^{N} \sum_{k=1}^{D} \left(\mathbf{h}_i^{(k)} * \mathbf{I}_j^{(k)} - \mathbf{g}_j \right)^2 + \lambda \sum_{k=1}^{D} ||\mathbf{h}_i^{(k)}||^2$$

Minimization across all channels

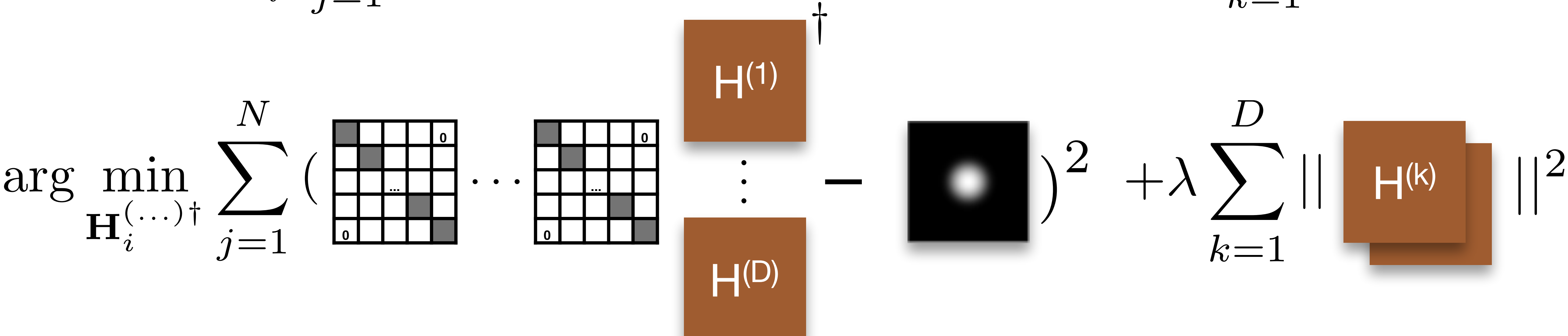
[J. Henriques et.al., ICCV 2013]
[H. Galoogahi et.al., ICCV 2013]
[V. Boddeti et.al., CVPR 2013]

Multiple Channel Correlation Filters

Frequency Domain



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



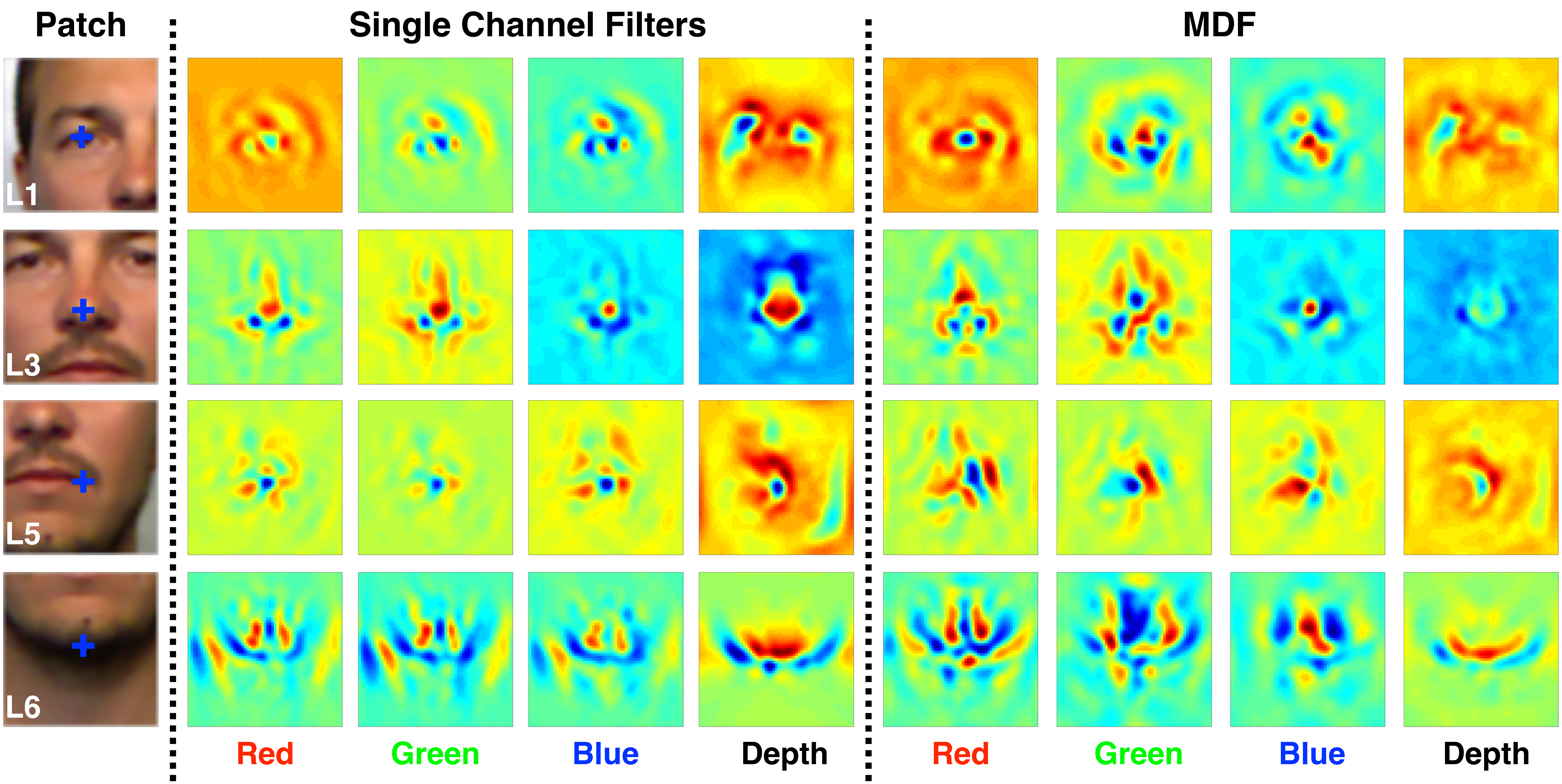
Efficient Solution (w/ variable re-ordering)

$$\{\mathbf{h}_{i(l)}^{(k)}\} = \mathcal{F}^{-1}\left\{\left(\sum_{j=1}^{N} \nu\left(\mathcal{F}\{\mathbf{I}_j\}_l\right)^H \nu\left(\mathcal{F}\{\mathbf{I}_j\}_l\right) + \lambda\mathbf{I}\right)^{-1} \sum_{j=1}^{N} \nu\left(\mathcal{F}\{\mathbf{I}_j\}_l\right)^H \nu\left(\mathcal{F}\{\mathbf{g}_j\}_l\right)\right\}^{\dagger}$$

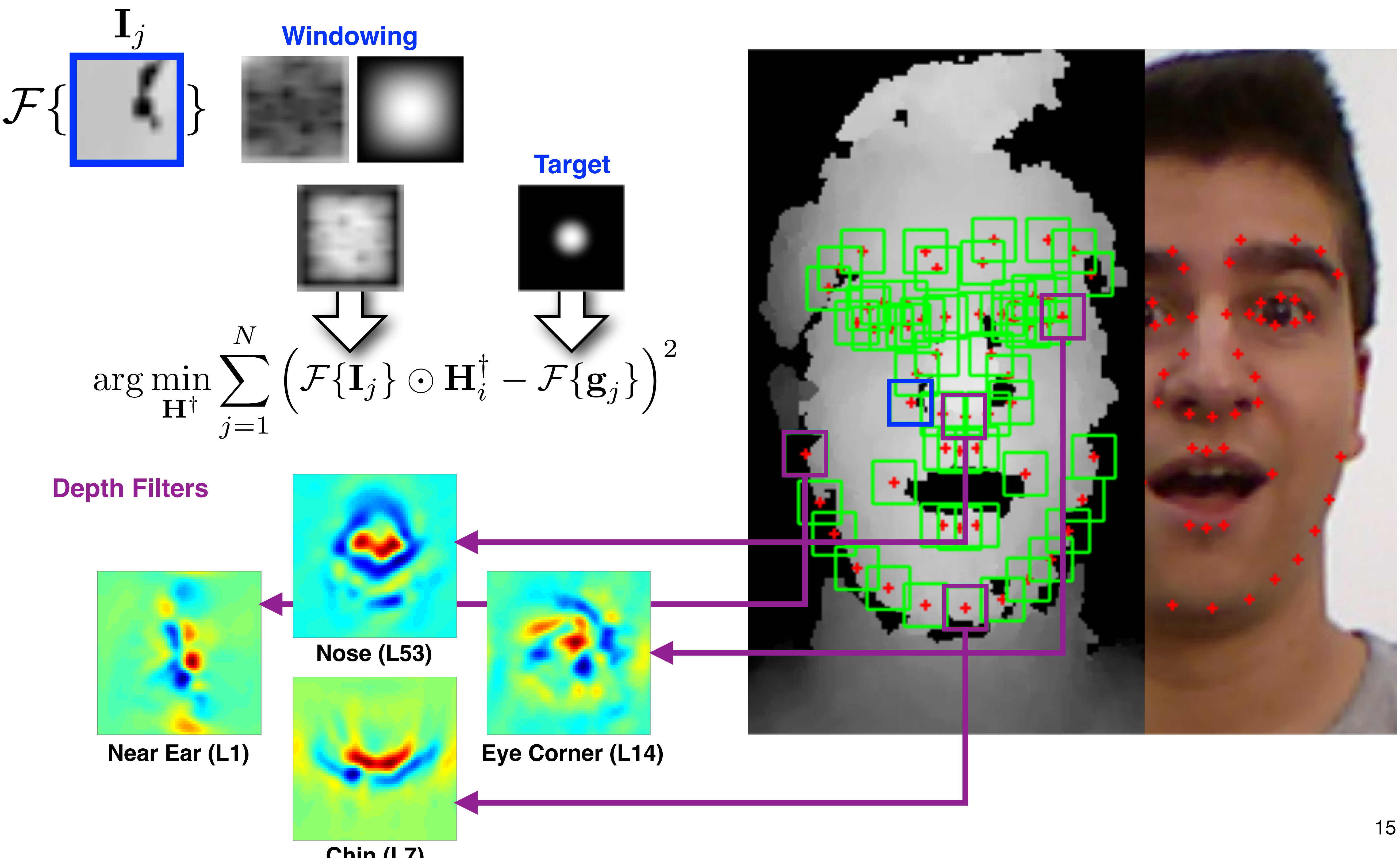
$$\nu\left(\mathbf{A}_l\right) = \left[\begin{array}{ccc} \mathbf{A}_l^{(1)} & \cdots & \mathbf{A}_l^{(D)} \end{array}\right]^T$$

[J. Henriques et.al., ICCV 2013]
[H. Galoogahi et.al., ICCV 2013]

Single Channel vs Multiple Channels Filters



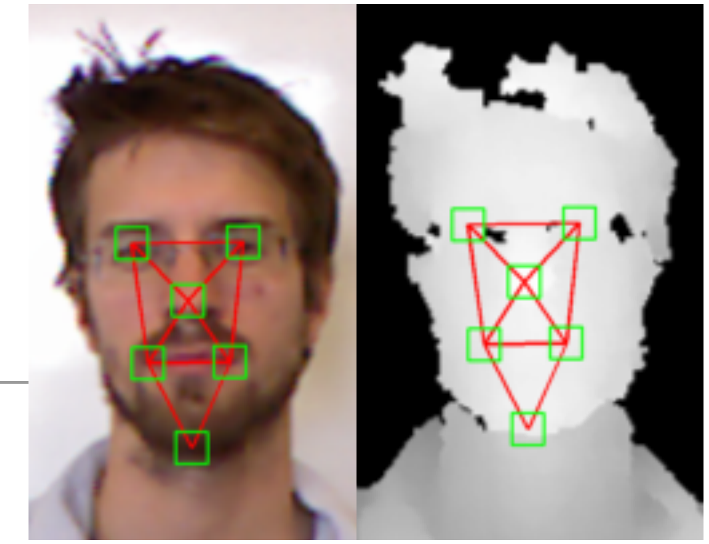
Depth Data (Noise Removal)

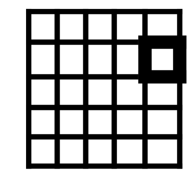


$$\mathcal{F}\{\mathbf{I}_j\}$$

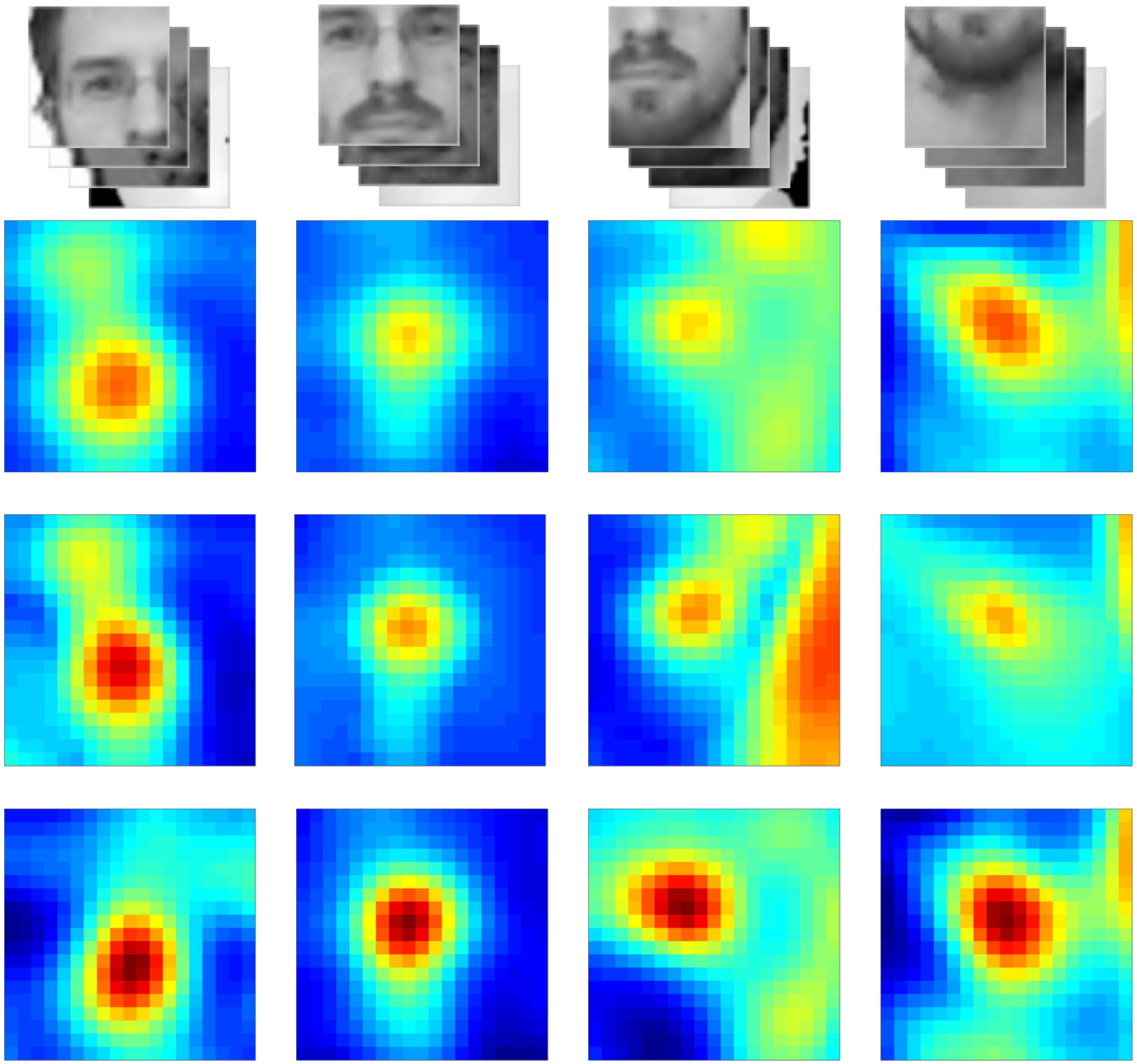
$$\arg\min_{\mathbf{H}^\dagger} \sum_{j=1}^{N} \left(\mathcal{F}\{\mathbf{I}_j\} \odot \mathbf{H}_i^\dagger - \mathcal{F}\{\mathbf{g}_j\} \right)^2$$

Response Map Fusion Strategies





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



AVG

$$p(\mathbf{z}_i)^{\mathrm{AVG}} = \frac{1}{D}\sum_{k=1}^{D} p(a_i|\mathcal{D}_i^{(k)}, \mathbf{I}(\mathbf{z}_i)^{(k)})$$

[T. Baltrusaitis et.al., CVPR 2012]
[S. Cheng et.al. ICIP 2014]

MAX

$$p(\mathbf{z}_i)^{\mathrm{MAX}} = \max_{\mathbf{z}_i}\; p(a_i|\mathcal{D}_i^{(k)}, \mathbf{I}(\mathbf{z}_i)^{(k)})$$

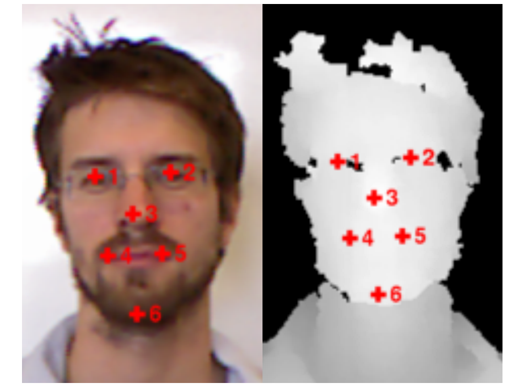
[P. Martins et.al., ICIP 2014]

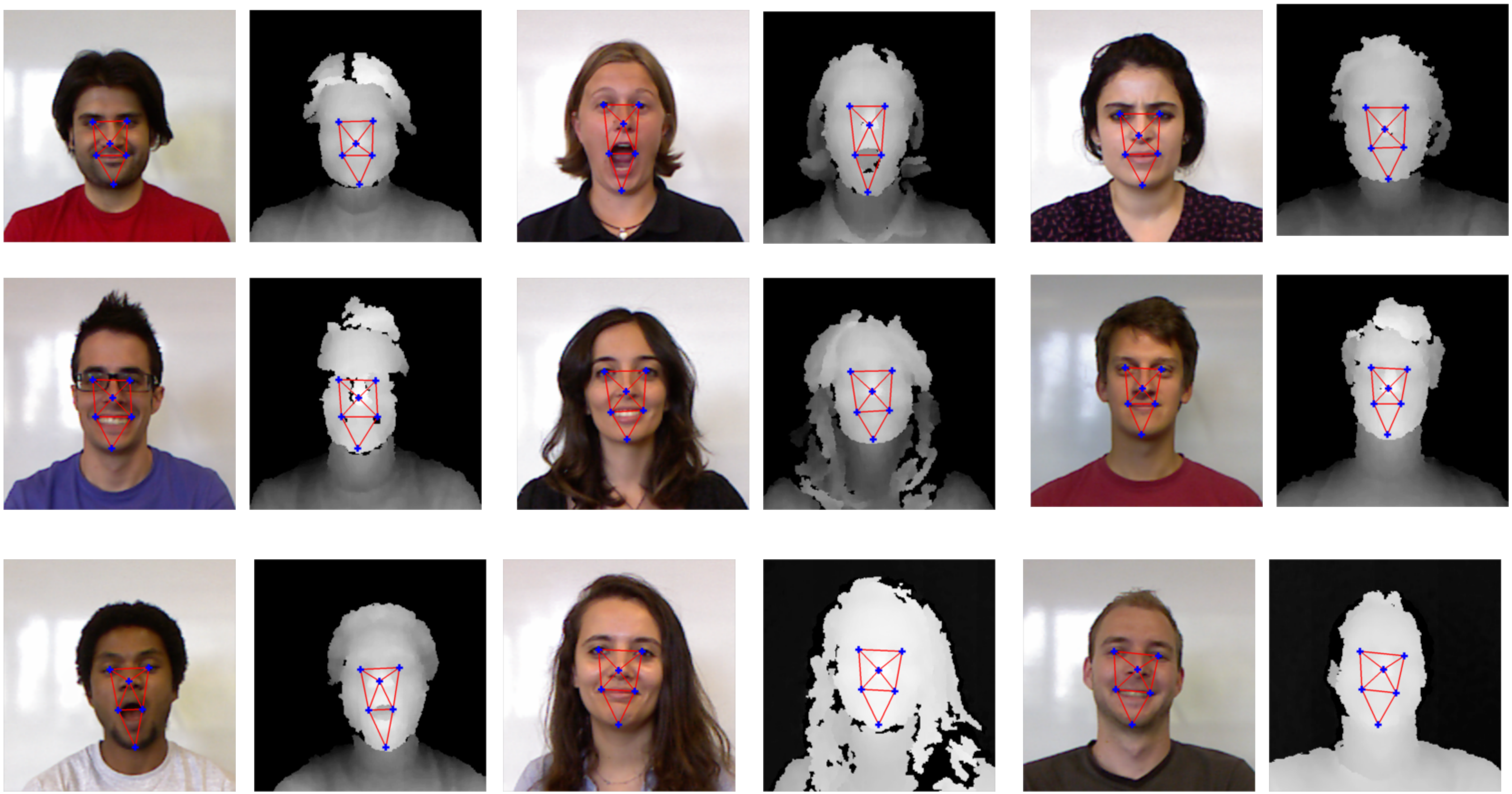
MDF

$$\mathcal{D}_i^{\mathrm{MDF}} = \sum_{k=1}^{D} \mathbf{h}_i^{(k)} \mathbf{I}(\mathbf{z}_i)^{(k)}$$

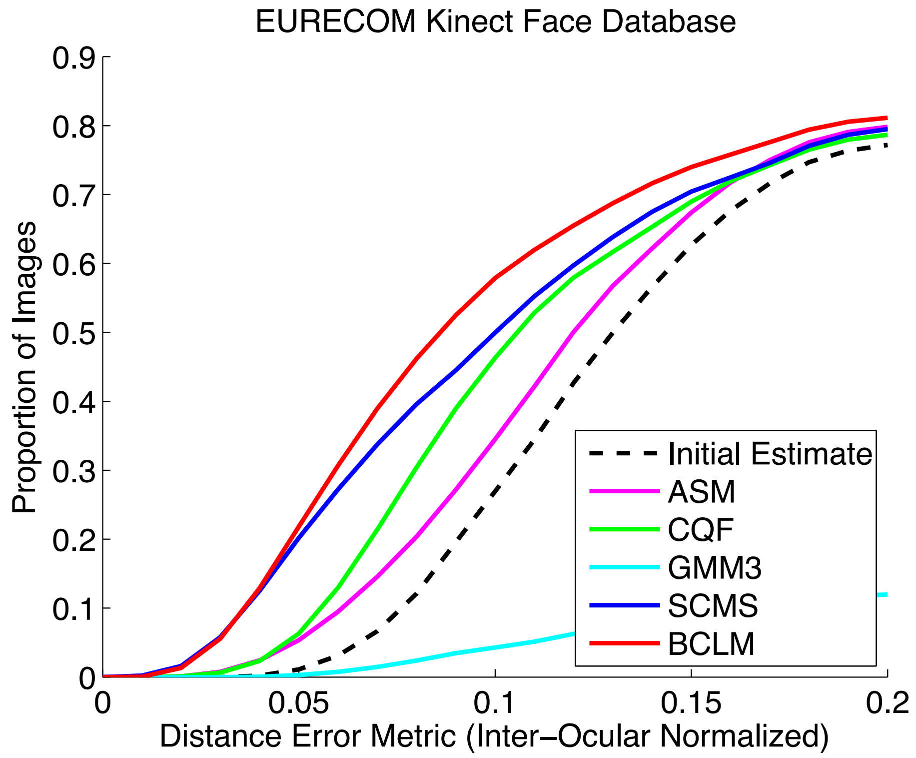
$$p(\mathbf{z}_i)^{\mathrm{MDF}} = p(a_i|\mathcal{D}_i^{\mathrm{MDF}}, \mathbf{I}(\mathbf{z}_i)^{(k)})$$

Qualitative Results EURECOM Database



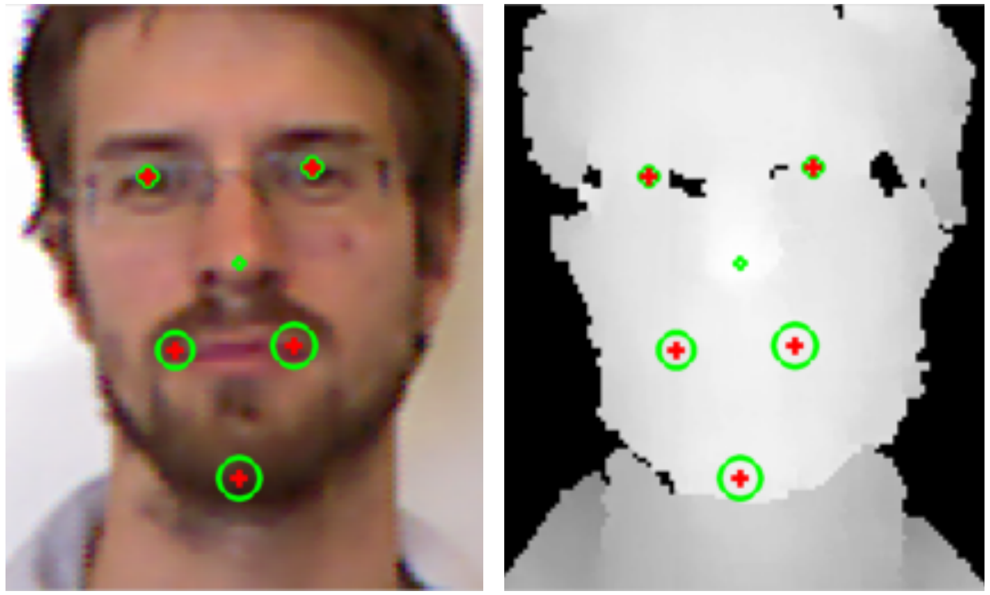


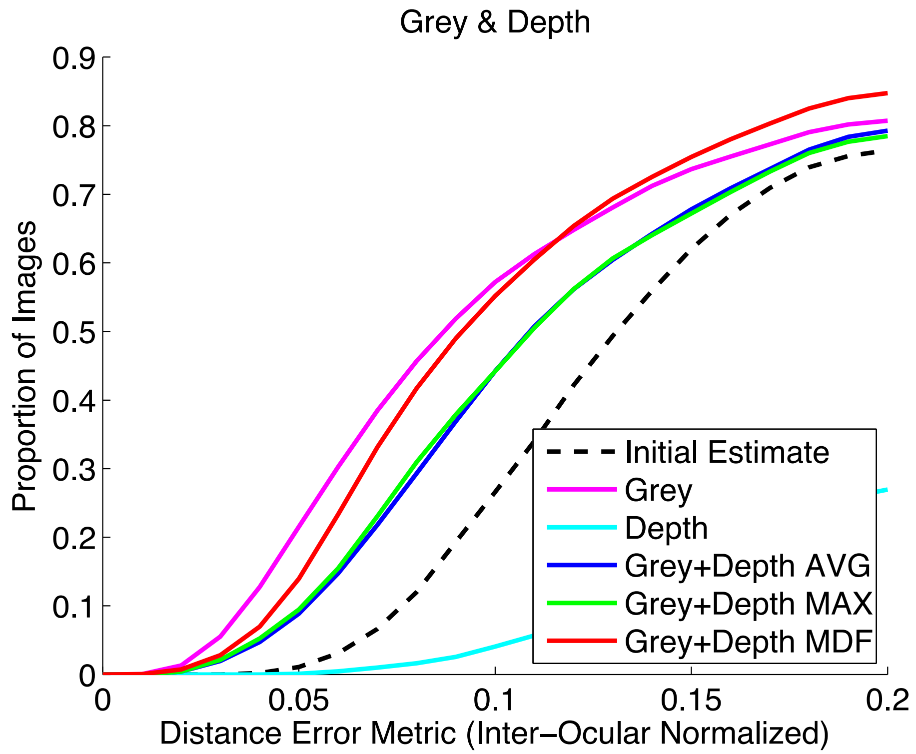
Results EURECOM Kinect Face Database

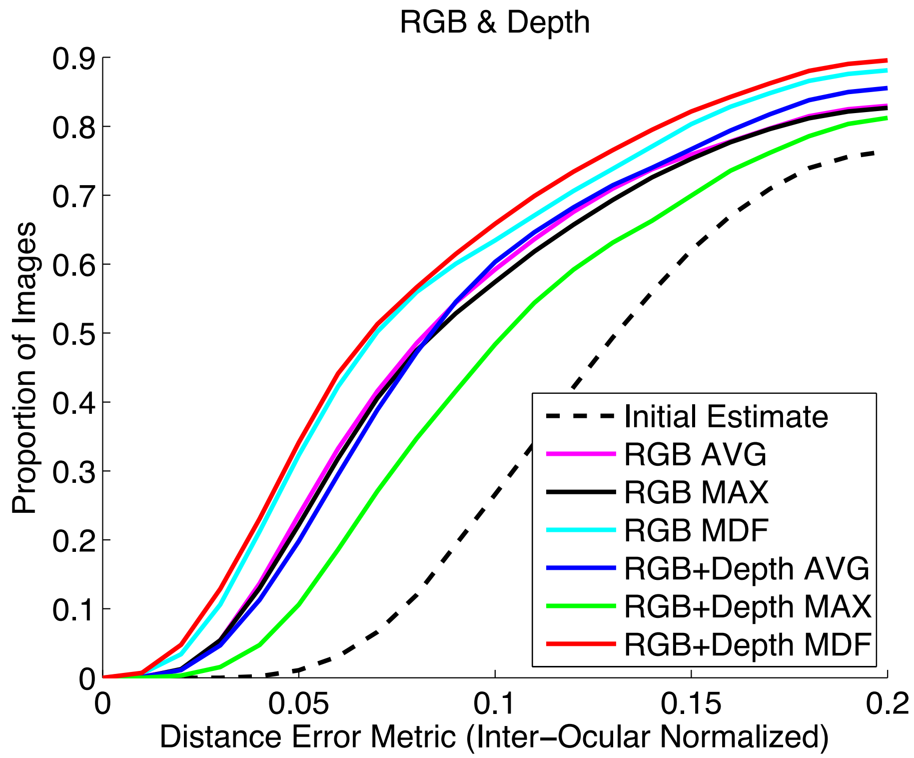


EURECOM Database
624 Images
52 Individuals (14F, 38M)
6 Landmarks

Error Standard Deviation

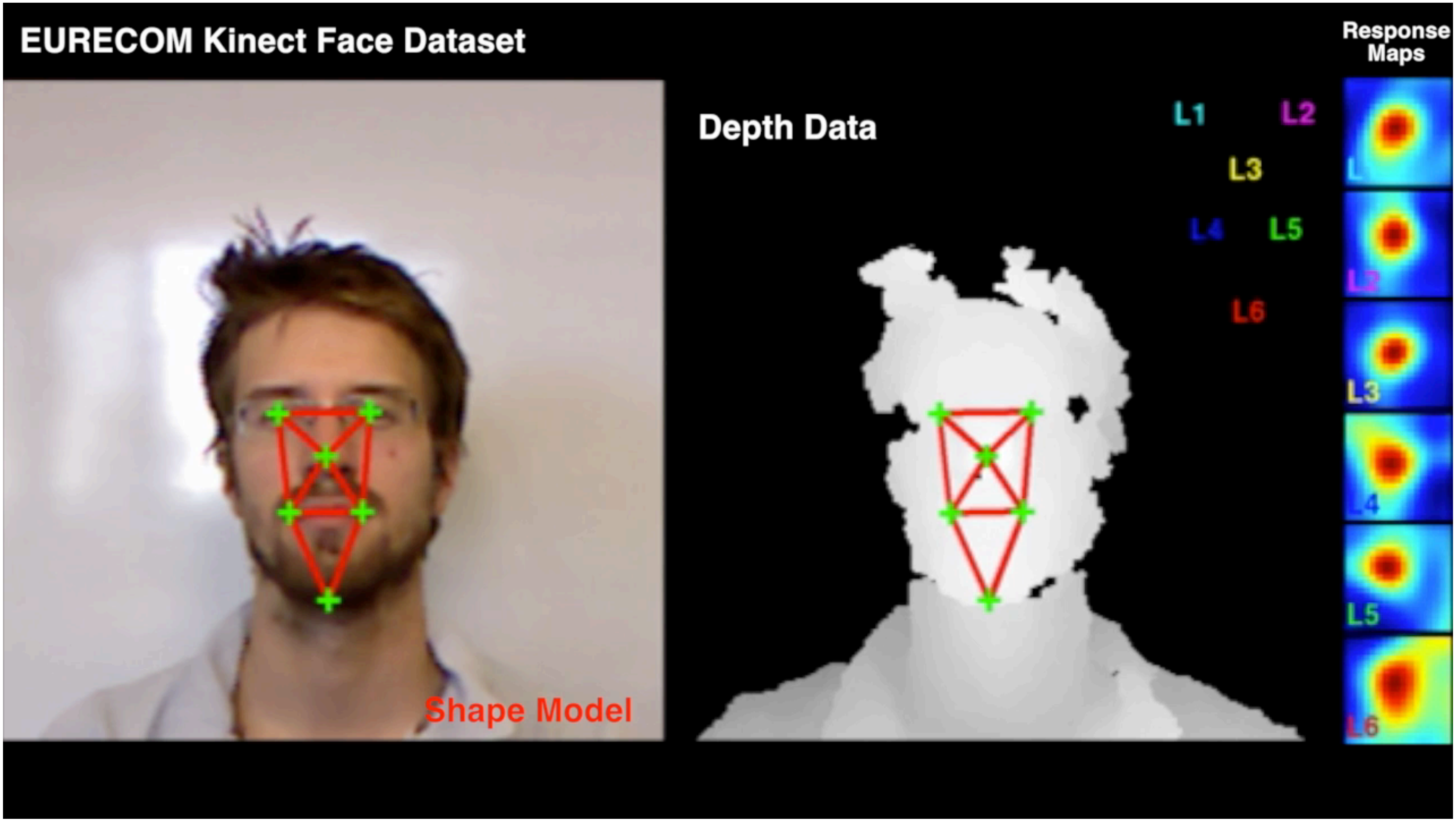




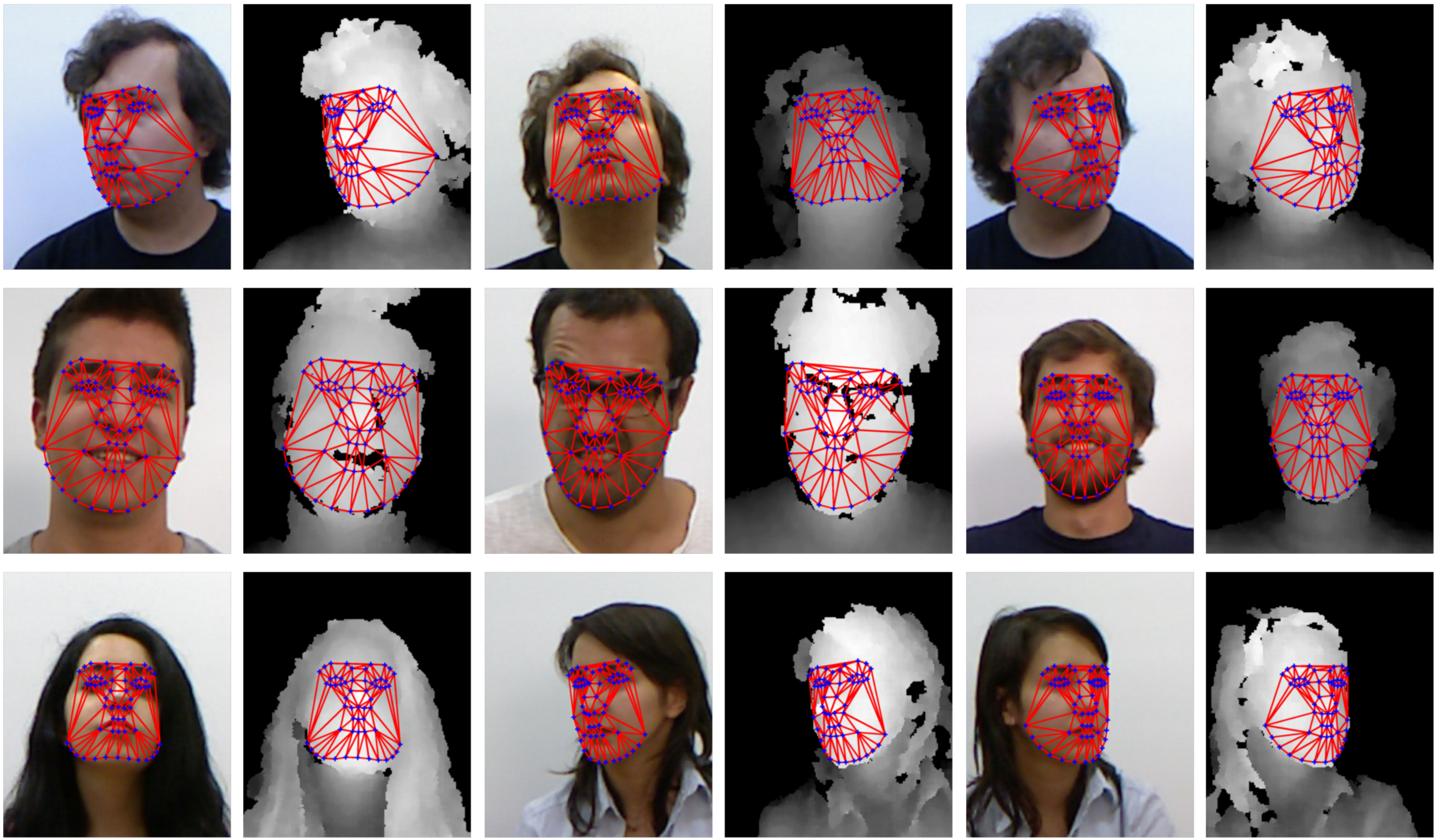


BCLM (MDF)	Area Below CDF
Grey	48.2
Grey+Depth	48.4
RGB	54.7
RGBD	56.4

BCLM MDF EURECOM Fitting (video)

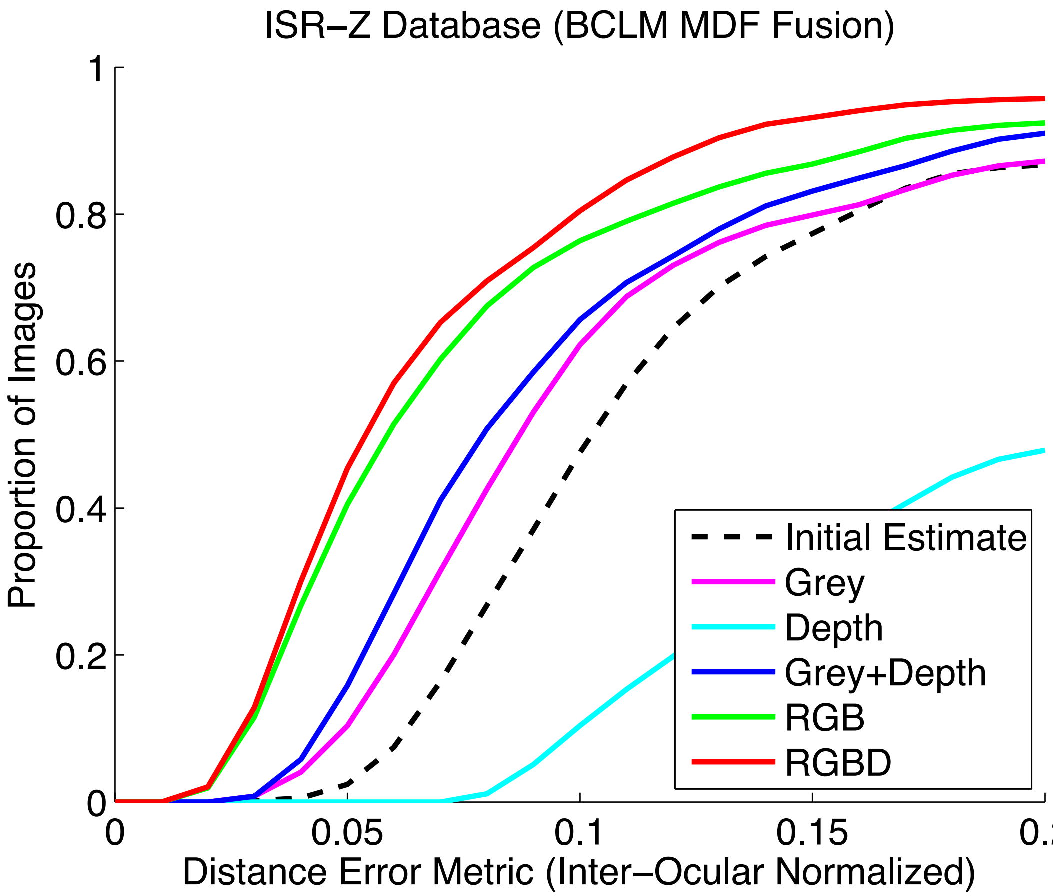


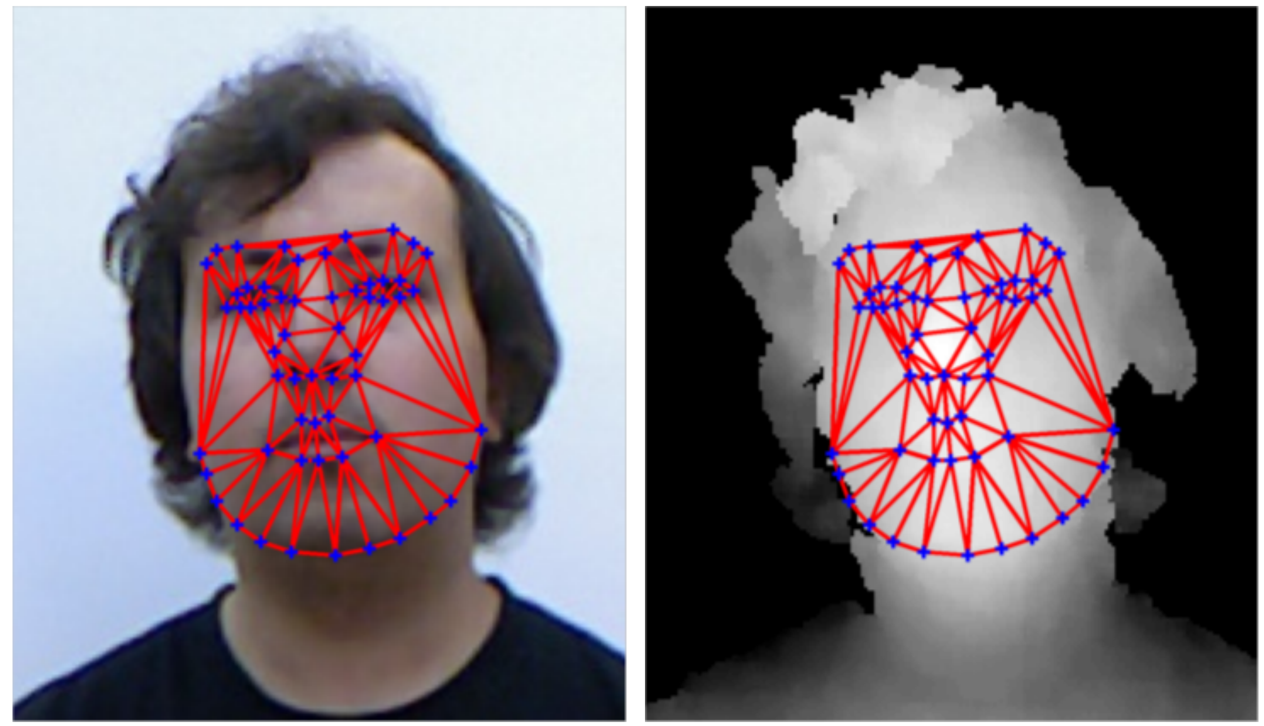
Qualitative Results ISR-Z Dataset



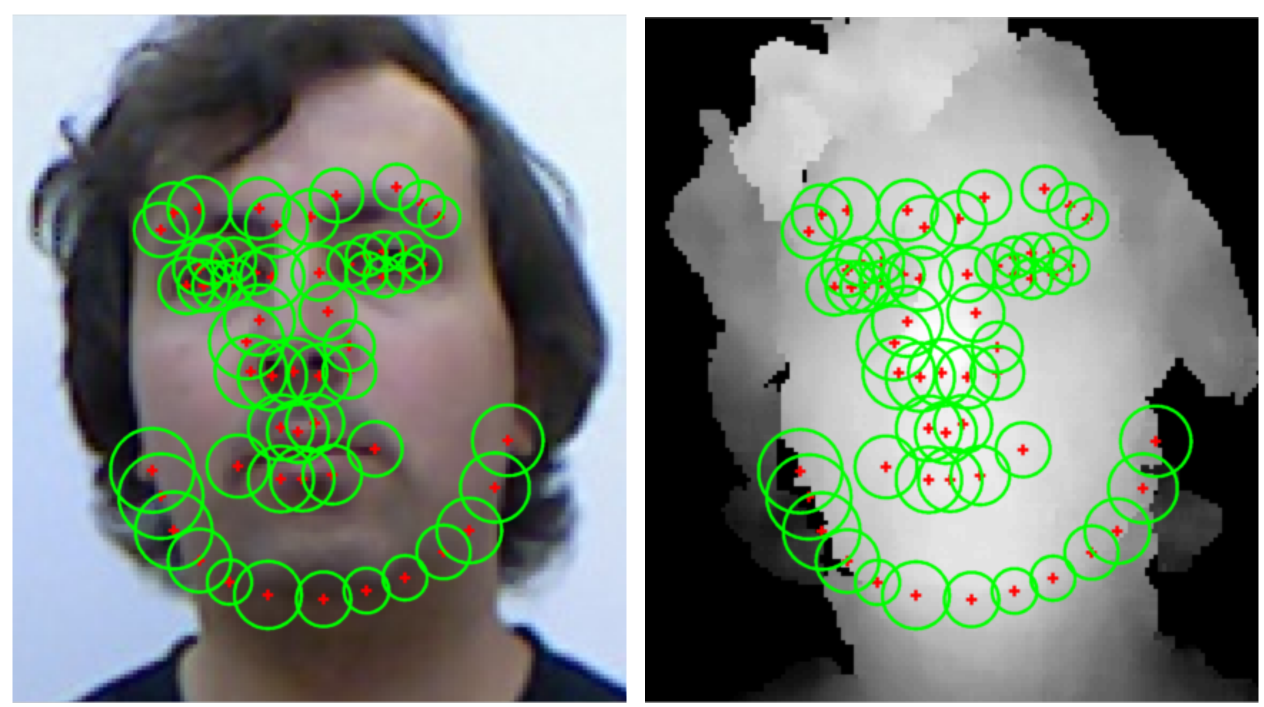
Results ISR-Z Face Database

ISR-Z Database
~250 Images
19 Individuals
58 Landmarks





Error Standard Deviation



BCLM (MDF)	Area Below CDF
Grey	49.0
Depth	16.6
Grey+Depth	52.5
RGB	61.7
RGBD	65.8

Conclusions

- Constrained Local Model (CLM) Extension
- Seamless Fitting of RGB and Depth Data
- Multiple Channel Correlation Filters (Local Detector)
- Evaluation Results (EURECOM and ISR-Z datasets)
- Acknowledgements
    - Work supported by the Portuguese Science Foundation (Fundação para a Ciência e Tecnologia - FCT) through the grant SFRH/BDP/90200/2012.

FCT
Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA EDUCAÇÃO E CIÊNCIA

COMPETE
PROGRAMA OPERACIONAL FACTORES DE COMPETITIVIDADE

QREN
QUADRO DE REFERÊNCIA ESTRATÉGICO NACIONAL
PORTUGAL 2007.2013

UNIÃO EUROPEIA
Fundo Europeu de Desenvolvimento Regional

BCLM MDF ISR-Z Fitting (video)

Questions?

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