Imperial College London

From Lukas-Kanade to Mnemonic Descent and Deep Dense Shape Regression: A brief history of (deformable) image alignment

Stefanos Zafeiriou

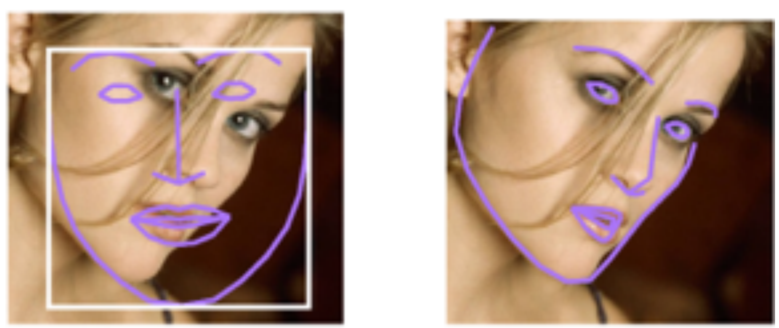
Thursday 19,
January 2017

Image/Object alignment

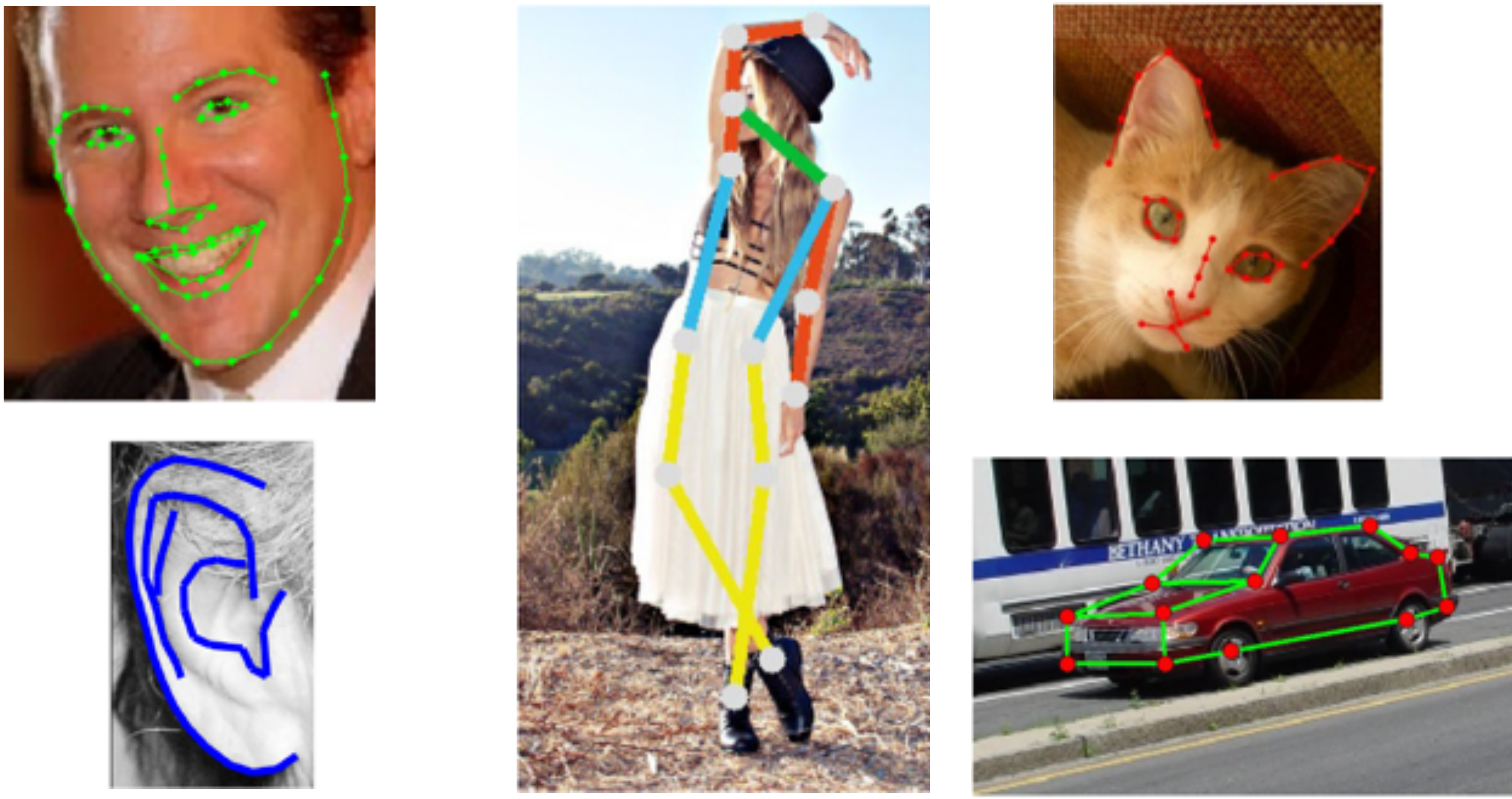
Image/Object alignment (or registration) is the process of transforming different sets of data into one coordinate system

- Global alignment

- Alignment using object parts (i.e., semantically meaningful components)

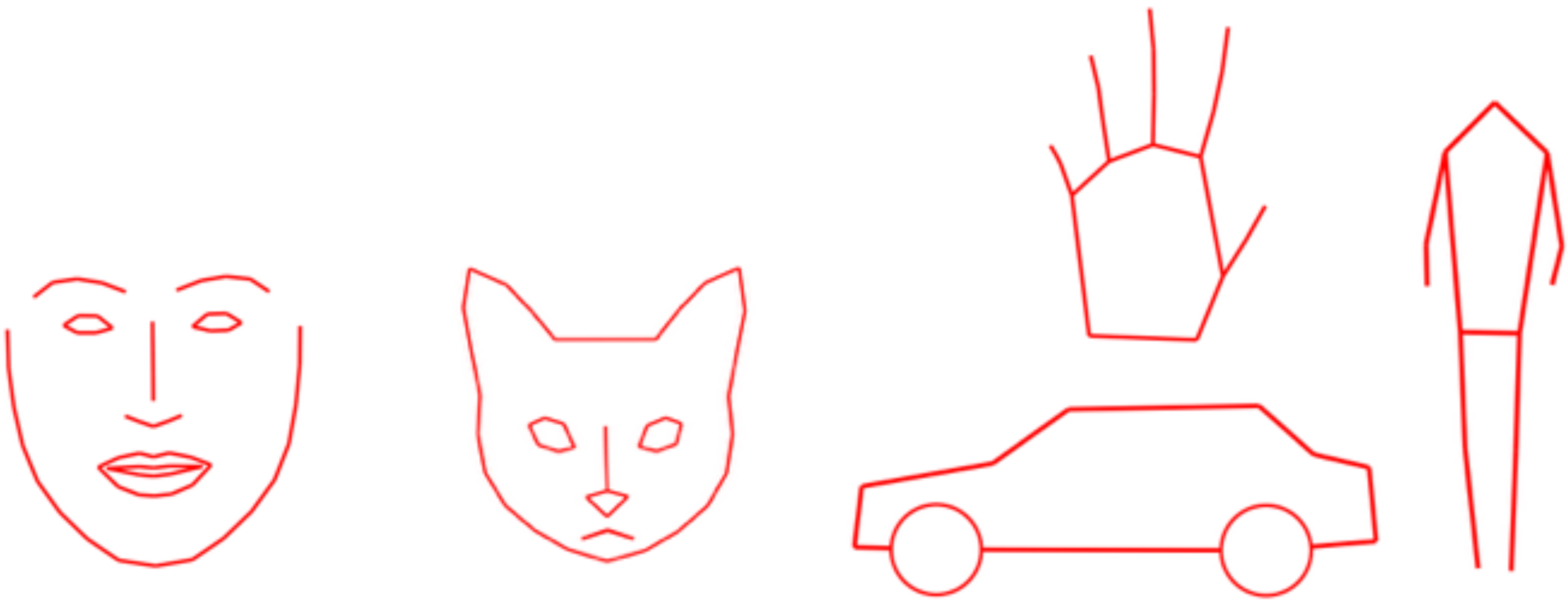


Deformable Object Alignment



What are Deformable Models?

Deformable Models aim to model
the *shape* and *appearance* variations of an object class.

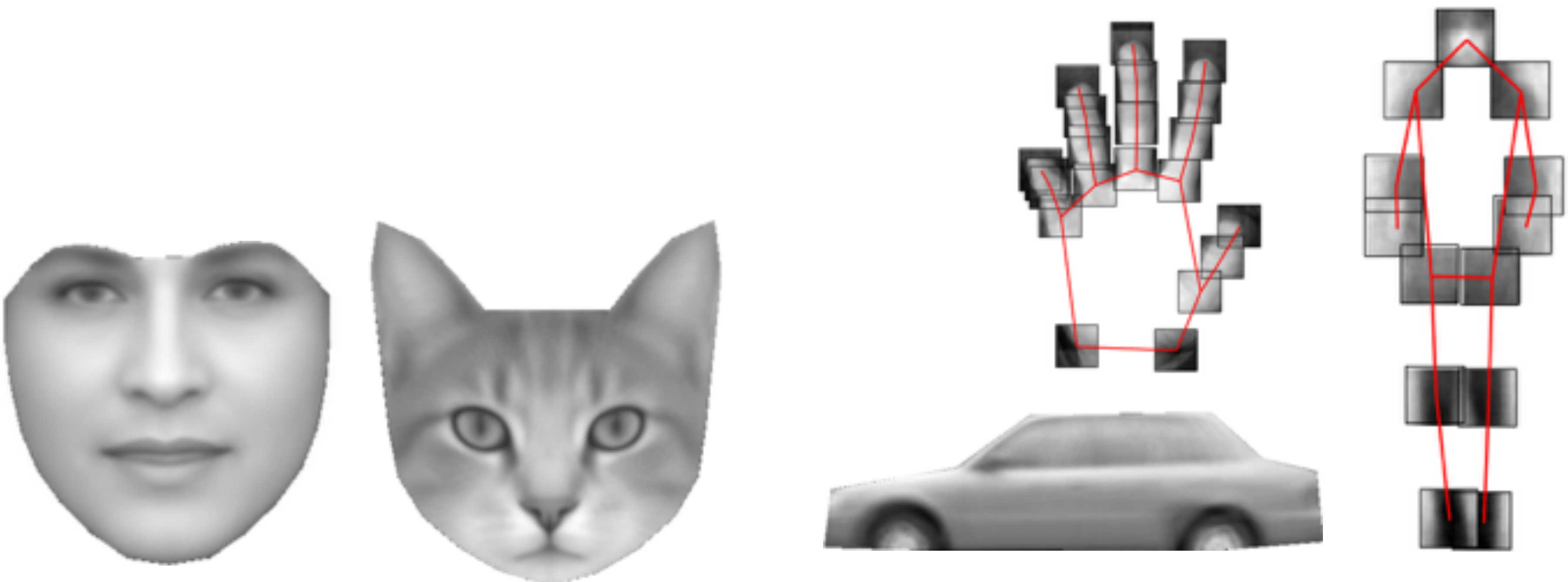


Trained using “*In-the-Wild*” data

(i.e. images captured under totally unconstrained conditions)

What are
Deformable Models?

Deformable Models aim to model
the *shape* and *appearance* variations of an object class.



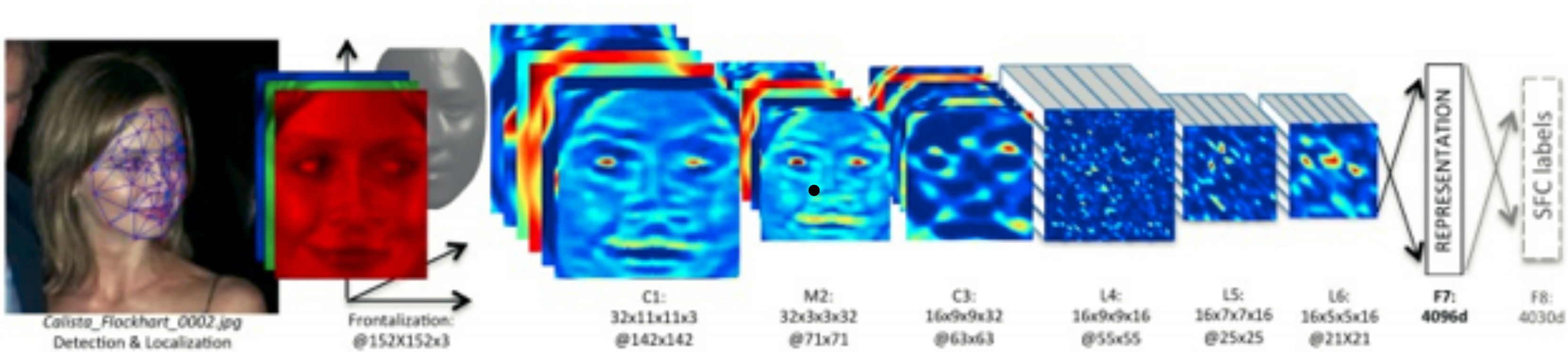
Trained using “*In-the-Wild*” data

(i.e. images captured under totally unconstrained conditions)

Recent successes in Computer Vision

Face recognition/verification ~ 97% in LFW

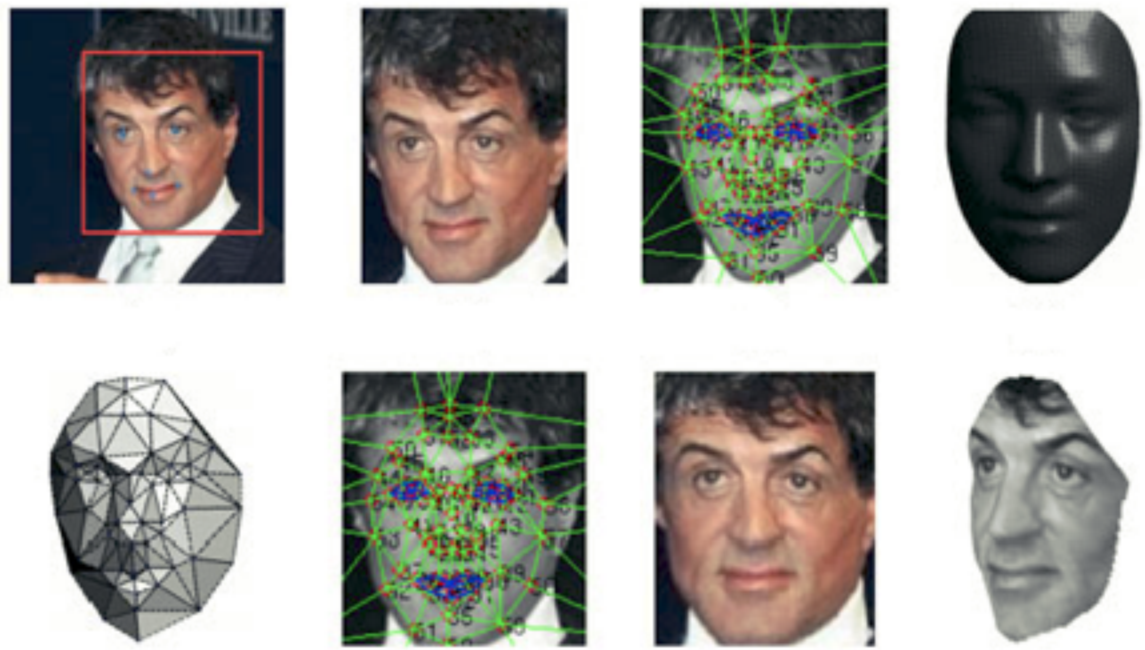
Trained on 4.4 million labeled facial images



- Should this success really attributed in deep learning?
- The network without alignment produces : 87.9%
- A very simple classifier on aligned data : ~ 92%
- In the same year Kitttler's group reported : ~ 96% (MRF alignment and simple features)

Deepface by Facebook, in CVPR 2014

Facebook's Deepface



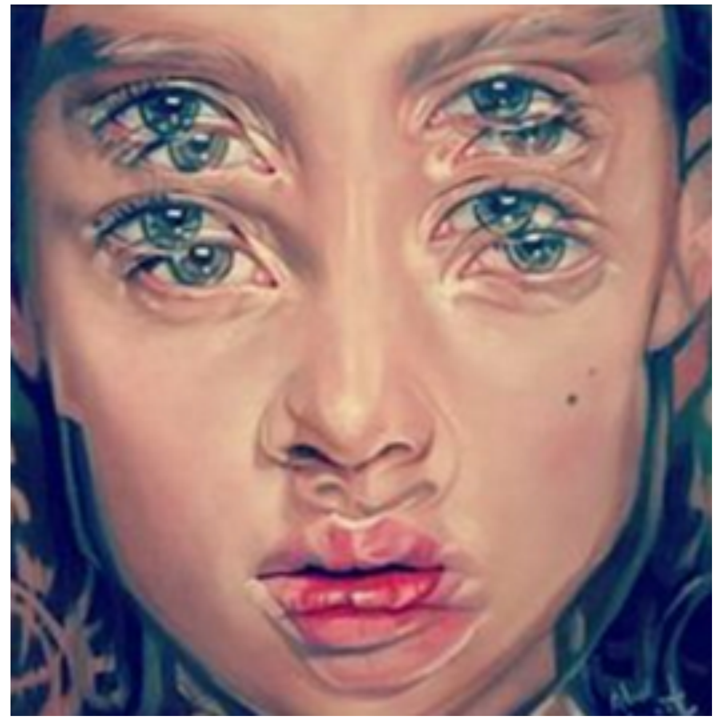
The success is mainly due to an elaborate image alignment and image normalization (frontalization) procedure

Google's FaceNet

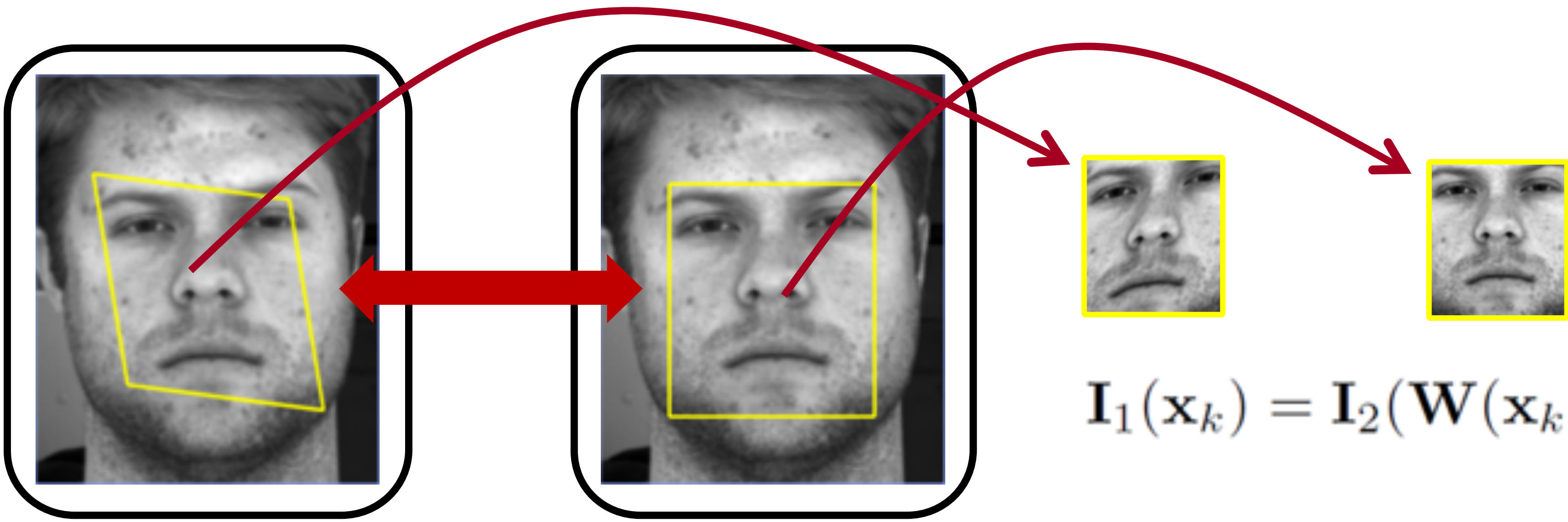
- Better performance without the need to apply such elaborate alignment (still alignment improves performance)

- But it is trained on the record number of 260,000,000 images!!!!

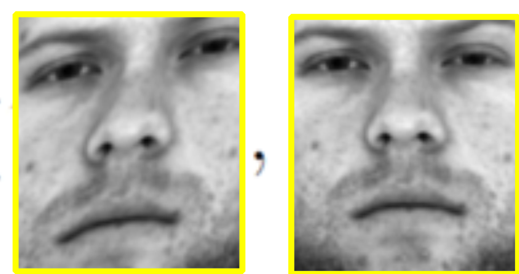
Why are these images disturbing?



Holistic Object alignment



$$\mathbf{I}_1(\mathbf{x}_k) = \mathbf{I}_2(\mathbf{W}(\mathbf{x}_k;\mathbf{p}))$$

Max(minim)ize $q(\mathbf{p}) = q($  $)$

W is the motion model, **p** are the motion parameters

Lukas-Kanade (LK) in a nutshell

- Find motion parameters by solving

$$\mathbf{p}_o = \arg\min_{\mathbf{p}} ||\mathbf{x}(\mathbf{p}) - \mathbf{m}||^2$$

- **x** is the test image and **m** the template (N-pixels)
- Example of warps and parameters (n-params):

Translation (2-parameters): $\mathbf{W}(\mathbf{x};\mathbf{p}) = \begin{pmatrix} x + p_1 \\ y + p_2 \end{pmatrix}$

Affine (6-parameters):

$$\mathbf{W}(\mathbf{x};\mathbf{p}) = \begin{pmatrix} (1+p_1)\cdot x & + & p_3 \cdot y & + & p_5 \\ p_2 \cdot x & + & (1+p_4)\cdot y & + & p_6 \end{pmatrix} = \begin{pmatrix} 1+p_1 & p_3 & p_5 \\ p_2 & 1+p_4 & p_6 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

[1] B. Lucas, and T. Kanade. "An iterative image registration technique with an application to stereo vision." IJCAI. Vol. 81. No. 1. 1981.

LK: Forward-Additive

Step 1: Linearise around the current estimate

$$\mathbf{x}(\mathbf{p}_c + \delta\mathbf{p}) - \mathbf{m} \approx \mathbf{e}(\mathbf{p}_c) + \mathbf{J_x}\delta\mathbf{p}$$

Error image: $\mathbf{e}(\mathbf{p}_c) = \mathbf{x}(\mathbf{p}_c) - \mathbf{m}$

Jacobian: $\mathbf{J_x} = \nabla_W \mathbf{x} \frac{\partial W}{\partial \mathbf{p}}$ Derivative of the warp: $\frac{\partial W}{\partial \mathbf{p}}$

Step 2: Solve

$$\delta\mathbf{p}_o = \arg\max_{\delta\mathbf{p}} ||\mathbf{e}(\mathbf{p}_c) + \mathbf{J_x}\delta\mathbf{p}||^2$$

$$\delta\mathbf{p}_o = -(\mathbf{J_x}^T\mathbf{J_x})^{-1}\mathbf{J_x}^T\mathbf{e}(\mathbf{p}_c)$$

Step 3: Update

$$\mathbf{p}_o \leftarrow \mathbf{p}_c + \delta\mathbf{p}_o$$

LK: Inverse Compositional (IC)

Step 1: Linearise around zero (parameters in the template)

$$\mathbf{x}(\mathbf{p}_c) - \mathbf{m}(\delta\mathbf{p}) \approx \mathbf{e}(\mathbf{p}_c) - \mathbf{J}_\mathbf{m}\delta\mathbf{p}$$

Jacobian and Hessian can be pre-computed

$$\mathbf{J}_\mathbf{m} = \nabla_W \mathbf{m} \frac{\partial W}{\partial \mathbf{p}}$$

Step 2: Update

$$\delta\mathbf{p}_o = (\mathbf{J}_\mathbf{m}^T \mathbf{J}_\mathbf{m})^{-1} \mathbf{J}_\mathbf{m}^T \mathbf{e}(\mathbf{p}_c)$$

Step 3: Update

$$\mathcal{W}(\mathbf{p}_o) \leftarrow \mathcal{W}(\mathbf{p}_c) \circ \mathcal{W}(\delta\mathbf{p}_o)^{-1}$$

Linear Complexity!!!!

[1] Baker, S. and Matthews, I.. "Lucas-kanade 20 years on: A unifying framework." International journal of computer vision 56.3 (2004): 221-255.

From LK to statistical deformable models

Active Appearance Models (AAMs)

Timeline

	Generative Models	Discriminative Models
1998		[Cootes et al., 1998] AAMs with single step regression fitting!
2004	[Matthews & Baker, 2004] Project-Out Inverse Compositional Simultaneous Inverse Compositional	
2008	[Papandreou & Maragos, 2008] Alternating Inverse Compositional	
2009	[Amberg et al., 2009] [Tzimiropoulos et al., 2013] Project-Out Forward Compositional	
2013		[Asthana, Zafeiriou et al., 2013] [Xiong & De la Torre, 2013] [Cao et al., 2014] Cascaded regression fitting framework
2014	[Alabort & Zafeiriou, 2014, 2016] Unified framework for compositional fitting	
2015	[Antonakos et al., 2015] Active Pictorial Structures	
2016		[Trigeorgis et al., 2016] Mnemonic Descent Method

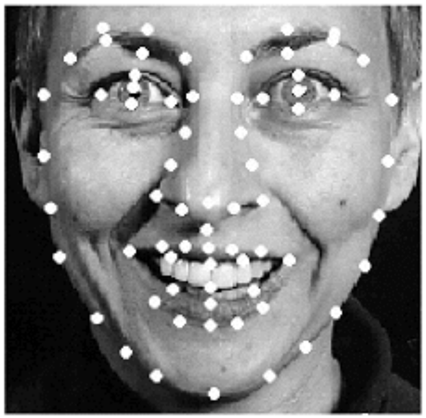
[Antonakos et al., 2016]
Adaptive Cascaded Regression

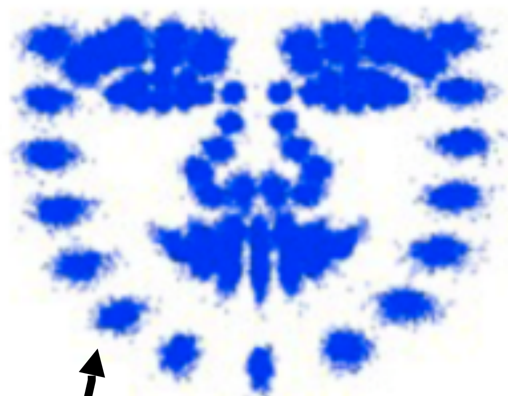
Timeline

	Generative Models	Discriminative Models
1998		[Cootes et al., 1998] AAMs with single step regression fitting!
2004	[Matthews & Baker, 2004] **Project-Out Inverse Compositional** Simultaneous Inverse Compositional	
2008	[Papandreou & Maragos, 2008] **Alternating Inverse Compositional**	
2009	[Amberg et al., 2009] [Tzimiropoulos et al., 2013]	
2013	Project-Out Forward Compositional	[Asthana, Zafeiriou et al., 2013] [Xiong & De la Torre, 2013] [Cao et al., 2014] **Cascaded regression fitting framework**
2014	[Alabort & Zafeiriou, 2014, 2016] Unified framework for compositional fitting	
2015	[Antonakos et al., 2015] **Active Pictorial Structures,**	
2016	**Feature based AAMs**	[Trigeorgis et al., 2016] **Mnemonic Descent Method**

[Antonakos et al., 2016]
Adaptive Cascaded Regression

From LK to AAMs



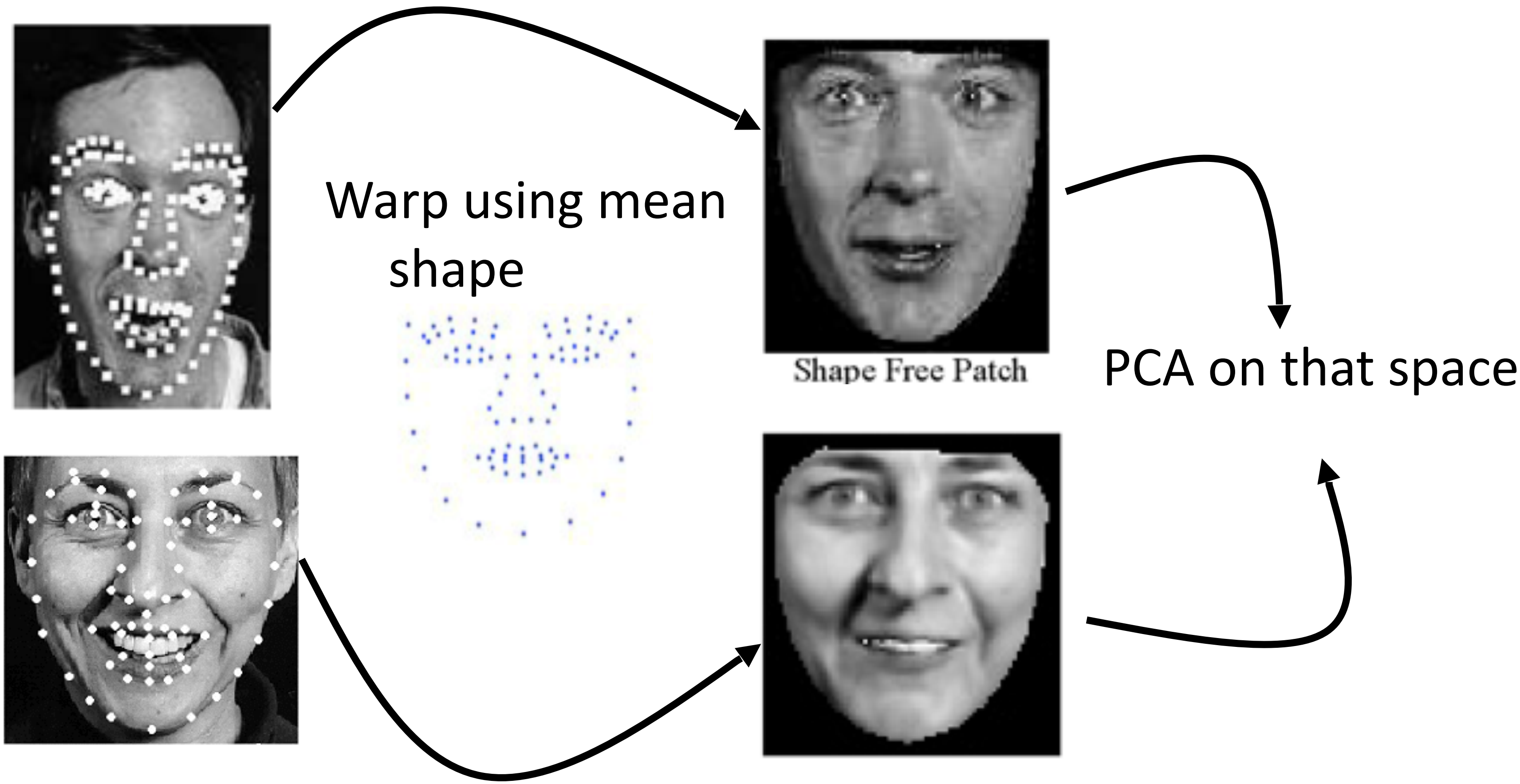


Shapes after procrustes

Mean Shape

PCA on that space

From LK to AAMs



[1] Cootes, T., Edwards, G. and Taylor, C. "Active appearance models." IEEE T-PAMI 23.6 (2001): 681-685.

From LK to AAMs

- The warping function is driven by the shape.
- Instead of the template image we have a linear texture model.
- The parameters that drive the model are
  (a) the weights of the linear shape,
  (b) a global similarity transform and
  (c) the weights of the linear texture model.

From LK to AAMs

Statistical parametric model of the shape and appearance of an object.

Shape PCA

$$\mathbf{s_p} = \mathbf{s}_0 + \mathbf{S}\mathbf{p}$$

Appearance PCA

$$\mathbf{a_c} = \boldsymbol{\mu} + \mathbf{A}\mathbf{c}$$

Recover shape and appearance parameters that minimize the reconstruction error of the warped image.

$$\arg\min_{\mathbf{p},\mathbf{c}} \|\mathbf{I}(\mathbf{p}) - (\boldsymbol{\mu} + \mathbf{A}\mathbf{c})\|_2^2$$

[1] I. Matthews and S. Baker. “Active Appearance Models Revisited”, IJCV, 60(2):135-164, 2004.

Robust LK-AAMs

- Not robust to outliers (occlusions, illumination, cast shadows etc.)
- Replace the least squares function with robust error functions
- Noise model for outliers hard to define - Are they always robust?
- Approximations for efficiency

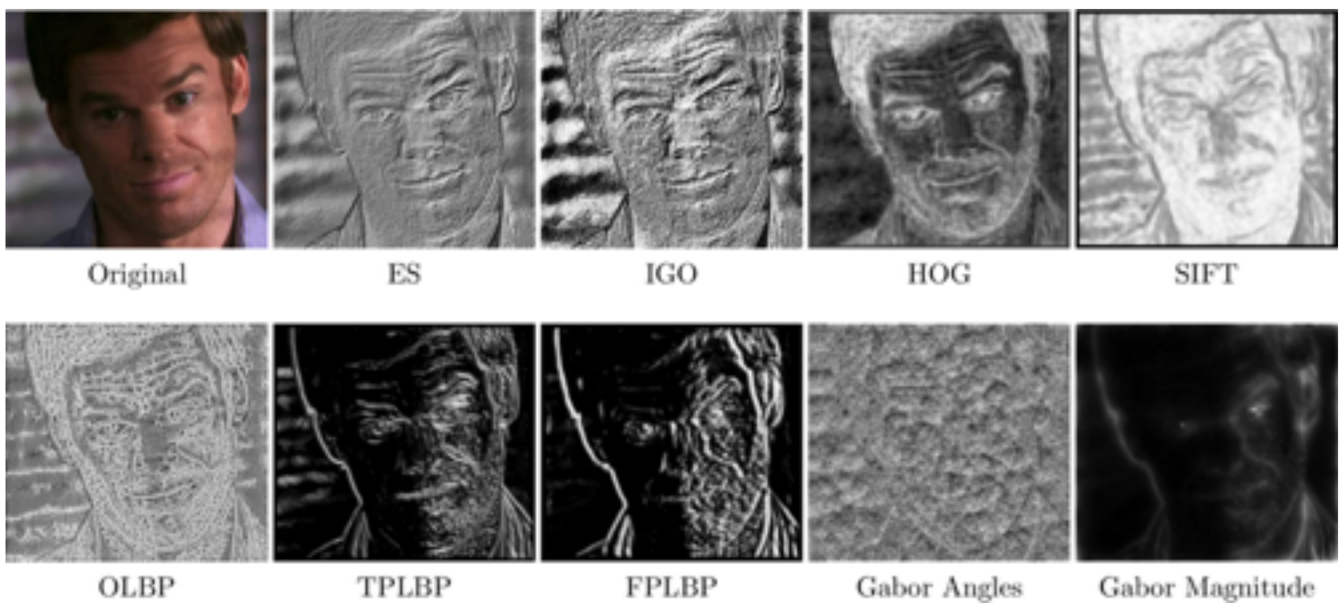
[1] Baker, S. and Matthews, I.. "Lucas-kanade 20 years on: A unifying framework." International journal of computer vision 56.3 (2004): 221-255.
[2] Dowson, N., and R.Bowden. "Mutual information for lucas-kanade tracking (milk): An inverse compositional formulation." IEEE T-PAMI 30.1 (2008): 180-185.
[3] Evangelidis, G., & Psarakis, E. (2008). Parametric image alignment using enhanced correlation coefficient maximization. IEEE T-PAMI, 30(10), 1858-1865.
[4] Lucey, Simon, et al. "Fourier lucas-kanade algorithm." IEEE T-PAMI 35.6 (2013): 1383-1396.

Feature-Based LK-AAMs

Given an image of size *H x W*, the ***dense*** feature extraction function is defined as:

$$\mathcal{F}:\ \mathbb{R}^{H\times W} \longrightarrow \mathbb{R}^{H\times W\times D}$$

where *D* is the number of channels.



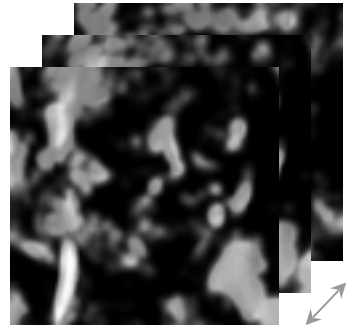
[1] Antonakos, Alabort-i-Medina, Tzimiropoulos and Zafeiriou, “Feature-Based Lucas-Kanade and Active Appearance Models”, IEEE TIP, 2015
[2] Tzimiropoulos, Georgios, Stefanos Zafeiriou, and Maja Pantic. "Robust and efficient parametric face alignment.", ICCV, 2011 (oral)
[3] Tzimiropoulos, G., Argyriou, V., Zafeiriou, S., & Stathaki, T. (2010). Robust FFT-based scale-invariant image registration with image gradients. IEEE T-PAMI, 32(10), 1899-1906.

Feature-Based LK-AAMs

$\mathbf{I}$

$\mathcal{F}(\mathbf{I})$ →

HOG, SIFT, ...



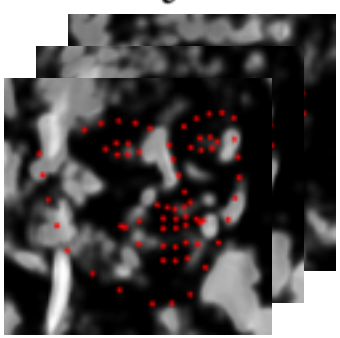
multi-channel

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

Dense image features

Patch-based warping

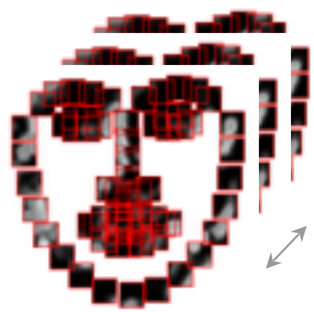
$\mathbf{I}_{\mathcal{F}}$



$\mathbf{s} = [x_1, y_1, x_2, y_2, \ldots]^T$

$\mathbf{I}_{\mathcal{F}}(\mathbf{s})$ →

patches extraction



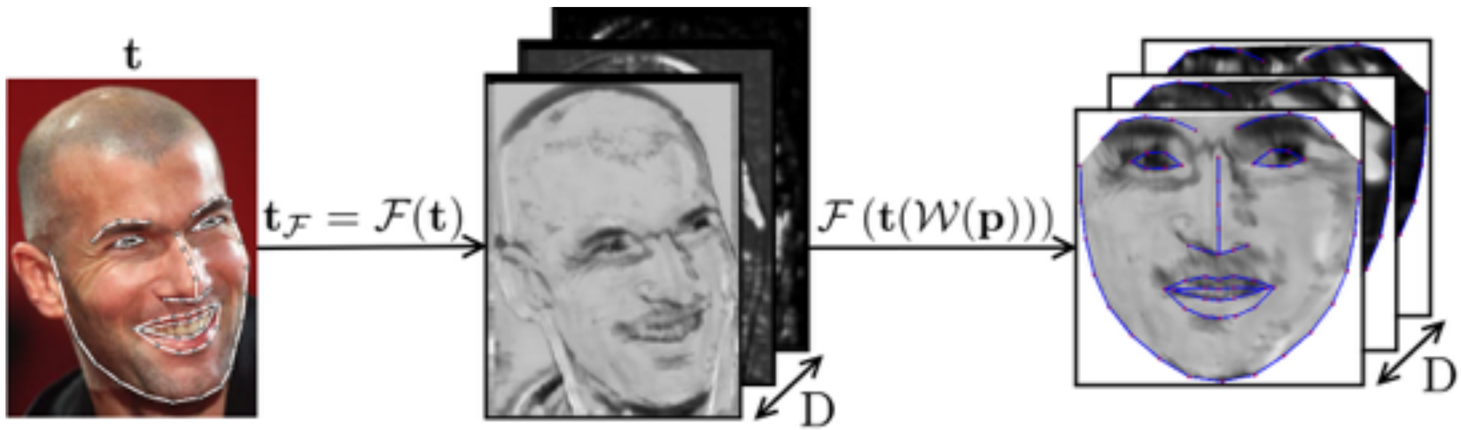
multi-channel patches

$\mathbf{I}_{\mathcal{F}}(\mathbf{s})$

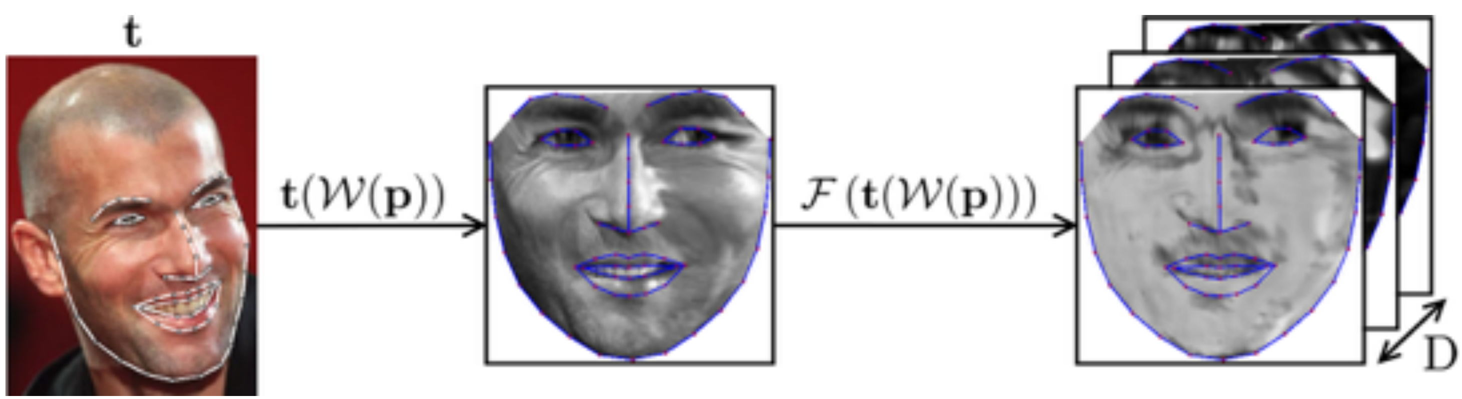
[1] Antonakos, Alabort-i-Medina, Tzimiropoulos and Zafeiriou, “Feature-Based Lucas-Kanade and Active Appearance Models”, IEEE TIP, 2015

Feature-Based AAMs

- Warp the features images



- Extract features on the warped image



AAM: Project-Out Inverse Compositional

$$\arg\min_{\mathbf{p},\mathbf{c}} \|\mathbf{I}(\mathbf{p}) - (\boldsymbol{\mu} + \mathbf{A}\mathbf{c})\|_2^2$$

Project-out

$$\arg\min_{\mathbf{p},\mathbf{c}} \underbrace{\|\mathbf{I}(\mathbf{p}) - (\boldsymbol{\mu} + \mathbf{A}\mathbf{c})\|_{\mathbf{A}\mathbf{A}^T}^2}_{\text{distance within } \mathbf{A} \text{ (0 for all } \mathbf{p})} + \underbrace{\|\mathbf{I}(\mathbf{p}) - (\boldsymbol{\mu} + \mathbf{A}\mathbf{c})\|_{\mathbf{E}-\mathbf{A}\mathbf{A}^T}^2}_{\text{distance to } \mathbf{A} \text{ (i.e. distance within orthogonal complement)}}$$

distance *within* **A**
(0 for all **p**)

distance *to* **A**
(i.e. distance within orthogonal complement)

$$\arg\min_{\mathbf{p}} \|\mathbf{I}(\mathbf{p}) - \boldsymbol{\mu}\|_{\mathbf{E}-\mathbf{A}\mathbf{A}^T}^2$$

Inverse and linearization

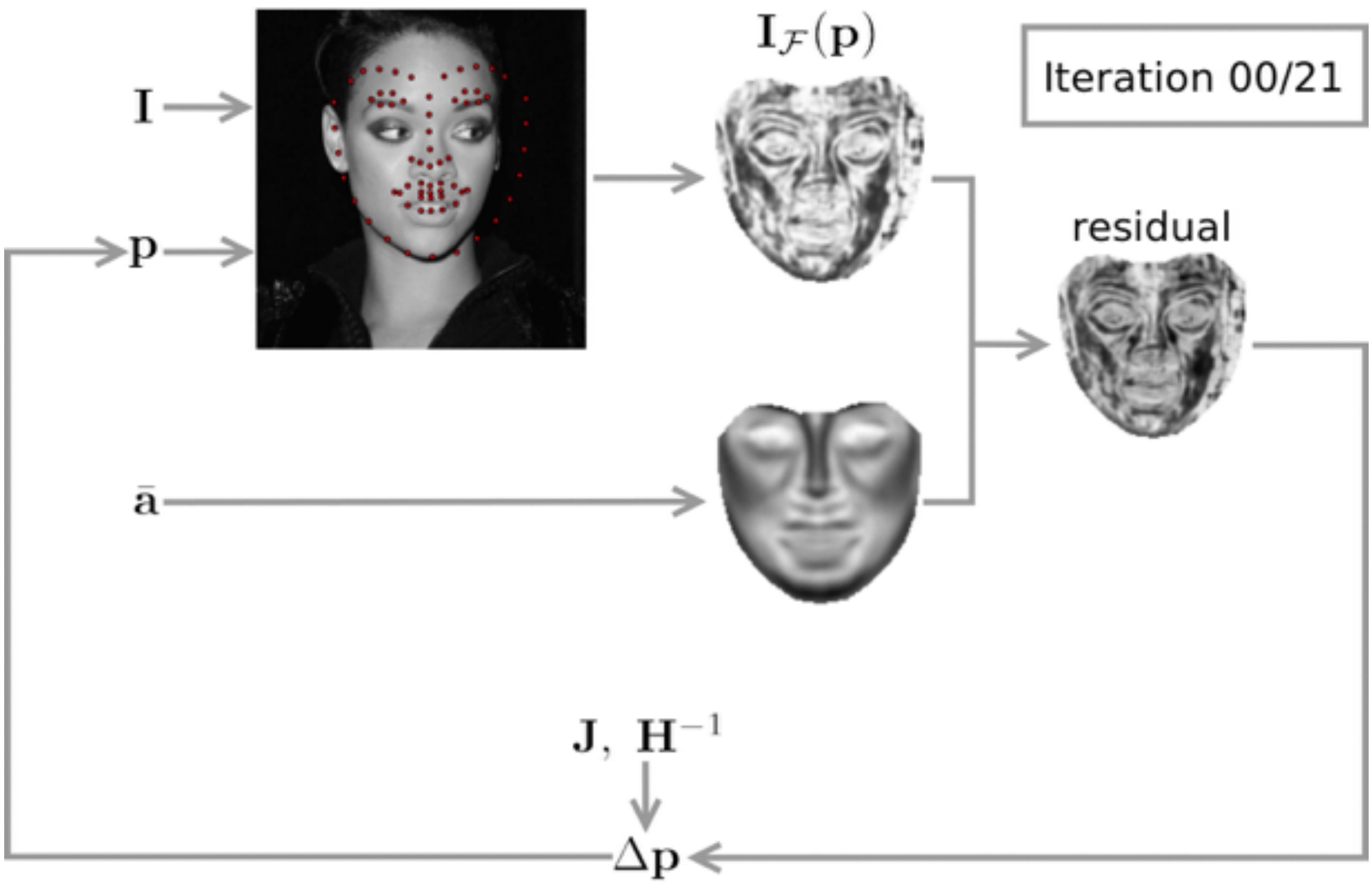
$$\arg\min_{\delta\mathbf{p}} \|\mathbf{I}(\mathbf{p}) - \underbrace{\boldsymbol{\mu}(\delta\mathbf{p})}_{\text{linearize}}\|_{\mathbf{E}-\mathbf{A}\mathbf{A}^T}^2 \longrightarrow \arg\min_{\delta\mathbf{p}} \|\mathbf{I}(\mathbf{p}) - \boldsymbol{\mu} - \mathbf{J}_{\boldsymbol{\mu}}\delta\mathbf{p}\|_{\mathbf{E}-\mathbf{A}\mathbf{A}^T}^2$$

linearize

$\delta\mathbf{p} = \mathbf{H}^{-1}\mathbf{J}^T(\mathbf{I}(\mathbf{p}) - \boldsymbol{\mu})$ with $\mathbf{J} = (\mathbf{E} - \mathbf{A}\mathbf{A}^T)\mathbf{J}_{\boldsymbol{\mu}}$ and $\mathbf{H} = \mathbf{J}^T\mathbf{J}$

[1] I. Matthews and S. Baker. "Active Appearance Models Revisited", IJCV, 60(2):135-164, 2004.

AAM:Project-Out Inverse Compositional



[1] I. Matthews and S. Baker. “Active Appearance Models Revisited”, IJCV, 60(2):135-164, 2004.

AAM: Alternating Inverse Compositional

$$\arg\min_{\mathbf{p},\mathbf{c}} \|\mathbf{I}(\mathbf{p}) - (\boldsymbol{\mu} + \mathbf{A}\mathbf{c})\|_2^2$$

Inverse and linearization

$$\arg\min_{\delta\mathbf{p},\delta\mathbf{c}} \|\mathbf{I}(\mathbf{p}) - \underbrace{(\boldsymbol{\mu}(\delta\mathbf{p}) + \mathbf{A}(\delta\mathbf{p})(\mathbf{c} + \delta\mathbf{c}))}_{\text{linearize}}\|_2^2$$

$$\arg\min_{\delta\mathbf{p},\delta\mathbf{c}} \|\mathbf{I}(\mathbf{p}) - \boldsymbol{\mu} - \mathbf{A}(\mathbf{c} + \delta\mathbf{c}) - \mathbf{J}_a\delta\mathbf{p}\|_2^2$$

Project-Out

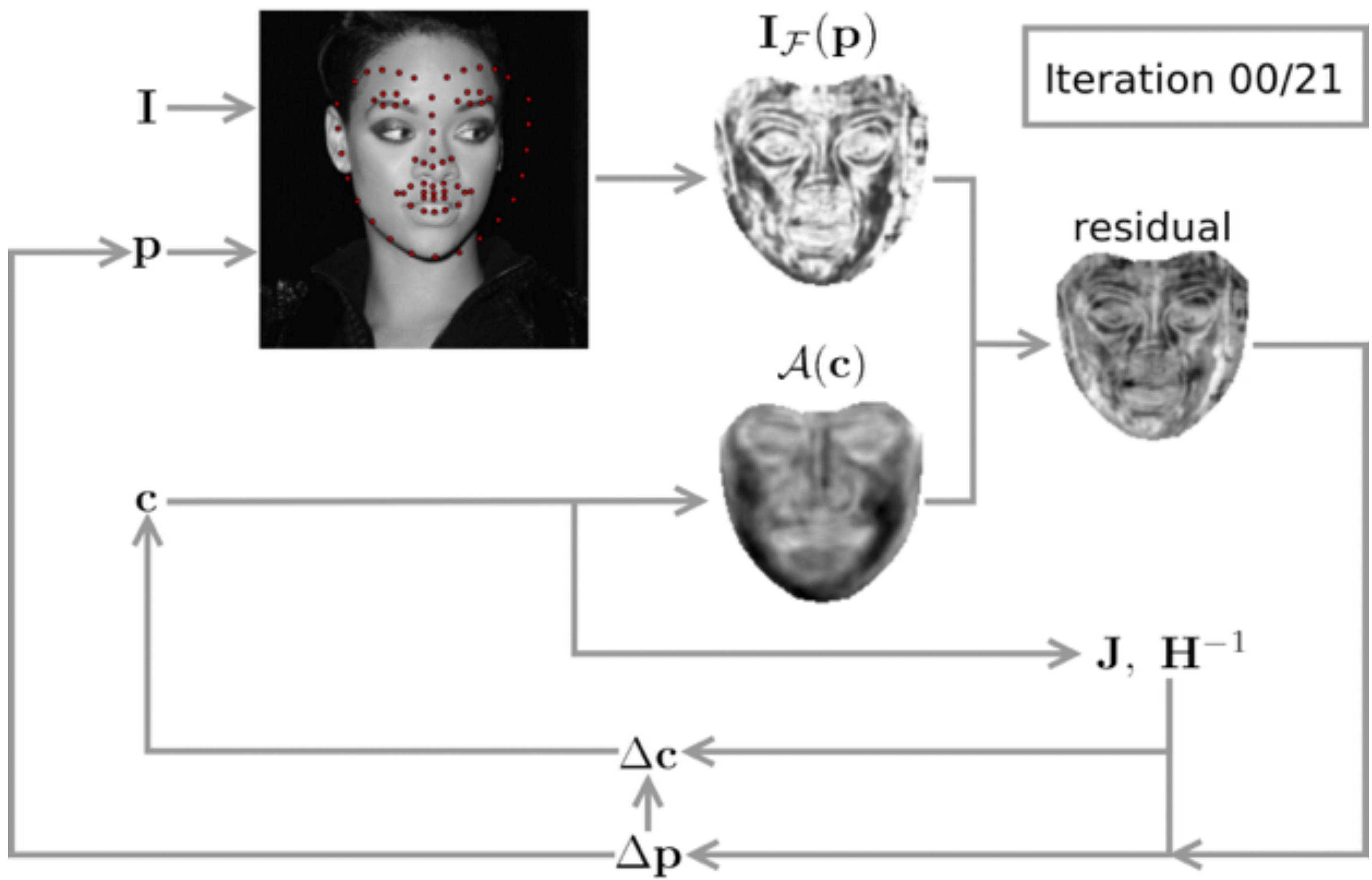
$$\arg\min_{\delta\mathbf{p}} \|\mathbf{I}(\mathbf{p}) - \boldsymbol{\mu} - \mathbf{J}_a\delta\mathbf{p}\|_{\mathbf{E}-\mathbf{A}\mathbf{A}^T}^2$$

$$\delta\mathbf{p} = \mathbf{H}^{-1}\mathbf{J}^T(\mathbf{I}(\mathbf{p}) - \boldsymbol{\mu})$$
$$\delta\mathbf{c} = \mathbf{A}^T(\mathbf{I}(\mathbf{p}) - \boldsymbol{\mu} - \mathbf{A}\mathbf{c} - \mathbf{J}_a\delta\mathbf{p})$$

with $\mathbf{J} = (\mathbf{E} - \mathbf{A}\mathbf{A}^T)\mathbf{J}_a$, $\mathbf{H} = \mathbf{J}^T\mathbf{J}$ and $\mathbf{J}_a = \mathbf{J}_{\boldsymbol{\mu}} + \sum \mathbf{c}_i\mathbf{J}_i$

[1] G. Papandreou and P. Maragos. "Adaptive and constrained algorithms for inverse compositional active appearance model fitting", CVPR, 2008.

AAM: Project-Out Inverse Compositional



[1] G. Papandreou and P. Maragos. “Adaptive and constrained algorithms for inverse compositional active appearance model fitting”, CVPR, 2008.

From LK to AAMs: What’s next?

AAMs: LK algorithm with a shape driven model and a special weighted least-squares cost function (a kind of Mahalanobis)

$$\arg\min_{\mathbf{p}} \|\mathbf{I}(\mathbf{p}) - \boldsymbol{\mu}\|^2_{\mathbf{E}-\mathbf{A}\mathbf{A}^T}$$

What kind of Mahalanobis (why the above weights?)

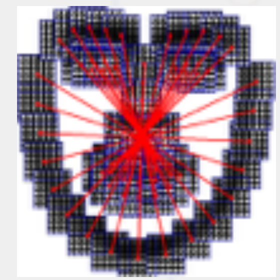
How to achieve real-time performance?

Pictorial Structures

Quick Overview

➜ Gaussian distribution for the patch-based appearance of each landmark

$$\mathcal{A}(\mathbf{x}_i) \sim \mathcal{N}(\bar{\mathbf{a}}_i, \boldsymbol{\Sigma}_i), \ \forall \text{ landmarks } i$$



➜ Gaussian distribution for each pairwise relative location between landmarks based on a tree (acyclic graph).

$$\mathbf{x}_{ij} = \mathbf{x}_i - \mathbf{x}_j \sim \mathcal{N}(\bar{\mathbf{x}}_{ij}, \boldsymbol{\Sigma}_{ij}), \ \forall \text{ edges } i, j$$

➜ Cost function

- Mahalanobis for the appearance of each landmark
- Spring-like deformation cost between pairs of landmarks

$$\arg\min_{\mathbf{x}_i} \sum_i \|\mathcal{A}(\mathbf{x}_i) - \bar{\mathbf{a}}_i\|^2_{\boldsymbol{\Sigma}_i^{-1}} + \sum_{i,j \in edges} \|\mathbf{x}_i - \mathbf{x}_j - \bar{\mathbf{x}}_{ij}\|^2_{\boldsymbol{\Sigma}_{ij}^{-1}}$$

➜ Global optimum using an efficient dynamic programming algorithm.

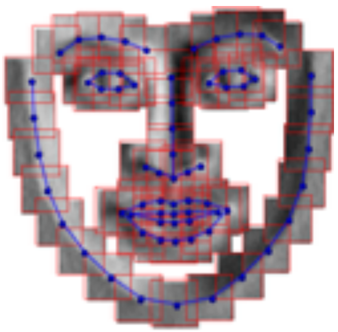
[1] P. Felzenszwalb and D. Huttenlocher. "Pictorial Structures for object recognition", IJCV 2005.
[2] P. Felzenszwalb and D. Huttenlocher. "Distance transforms of sampled functions", 2004.
[3] P. Felzenszwalb, et al. "Object detection with discriminatively trained part-based models", IEEE T-PAMI 2010.
[4] X. Zhu and D. Ramanan. "Face detection, landmark localization in the wild", CVPR 2012.
[5] Fischler, M.A. and Elschlager, R.A. 1973. The representation and 1376 matching of pictorial structures. IEEE Transactions on Computer, 1377 22(1):67–92.

Active Pictorial Structures

Appearance GMRF

$$[\mathcal{F}(\mathbf{x}_i), \mathcal{F}(\mathbf{x}_j)] \sim \mathcal{N}(\boldsymbol{\mu}^a_{ij}, \boldsymbol{\Sigma}^a_{ij})$$

$$\forall i, j \in \text{edges}$$



$$\mathbf{Q}_a = \boldsymbol{\Sigma}_a^{-1}$$

Block sparse precision matrix

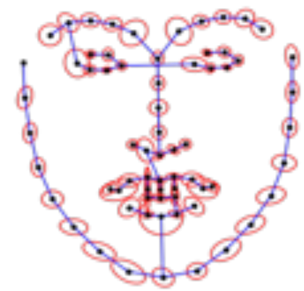
Shape PCA

$$\mathcal{S}(\mathbf{p}) = \bar{\mathbf{s}} + \mathbf{U}_s\mathbf{p}$$

Deformation GMRF

$$\mathbf{x}_i - \mathbf{x}_j \sim \mathcal{N}(\boldsymbol{\mu}^d_{ij}, \boldsymbol{\Sigma}^d_{ij})$$

$$\forall i, j \in \text{edges}$$



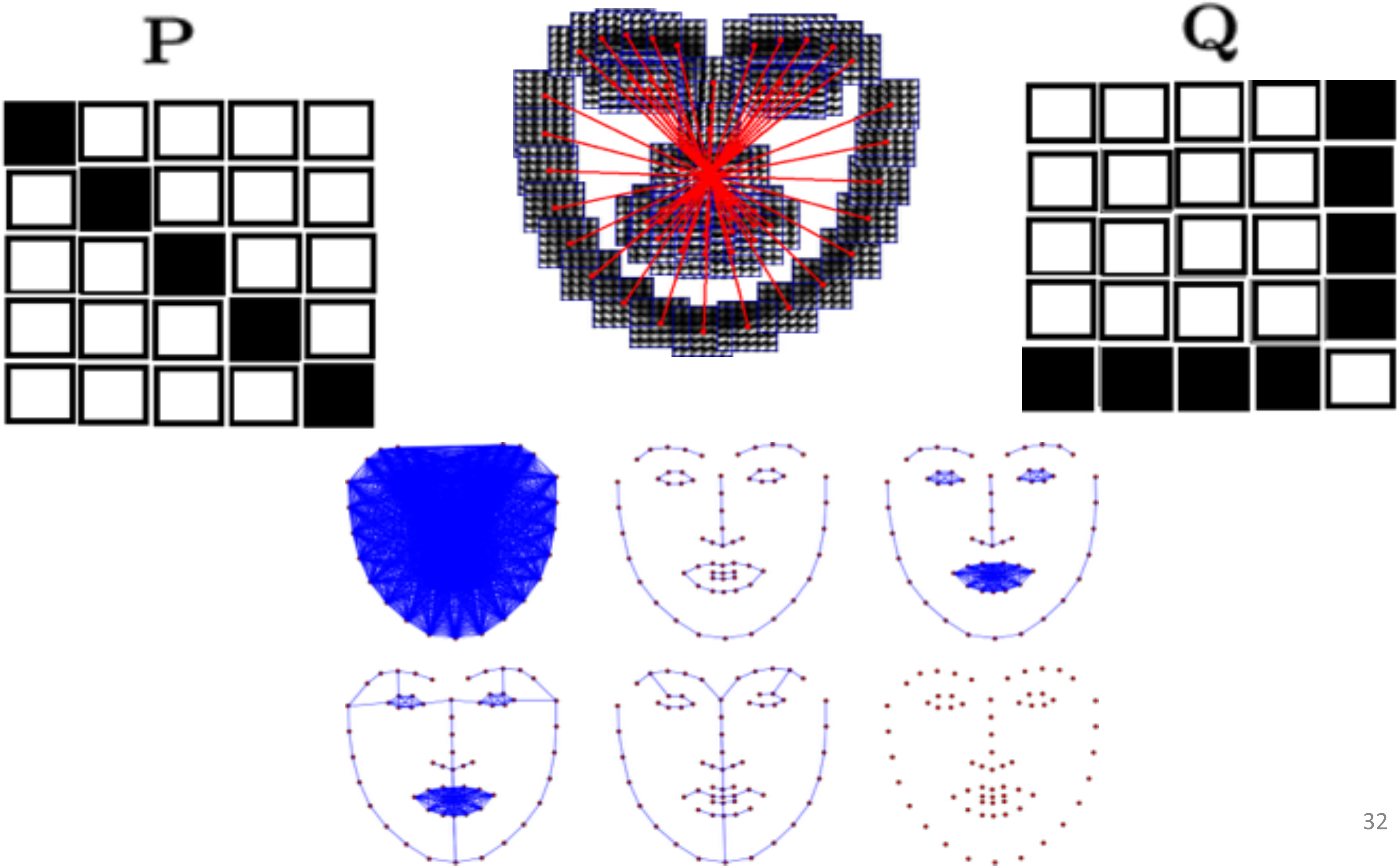
$$\mathbf{Q}_d = \boldsymbol{\Sigma}_d^{-1}$$

Block sparse precision matrix

The cost function consists of the appearance Mahalanobis distance, the deformation prior and a weight hyperparameter between them.

[1] **E. Antonakos**, J. Alabort-i-Medina, S. Zafeiriou. "Active Pictorial Structures", CVPR 2015.

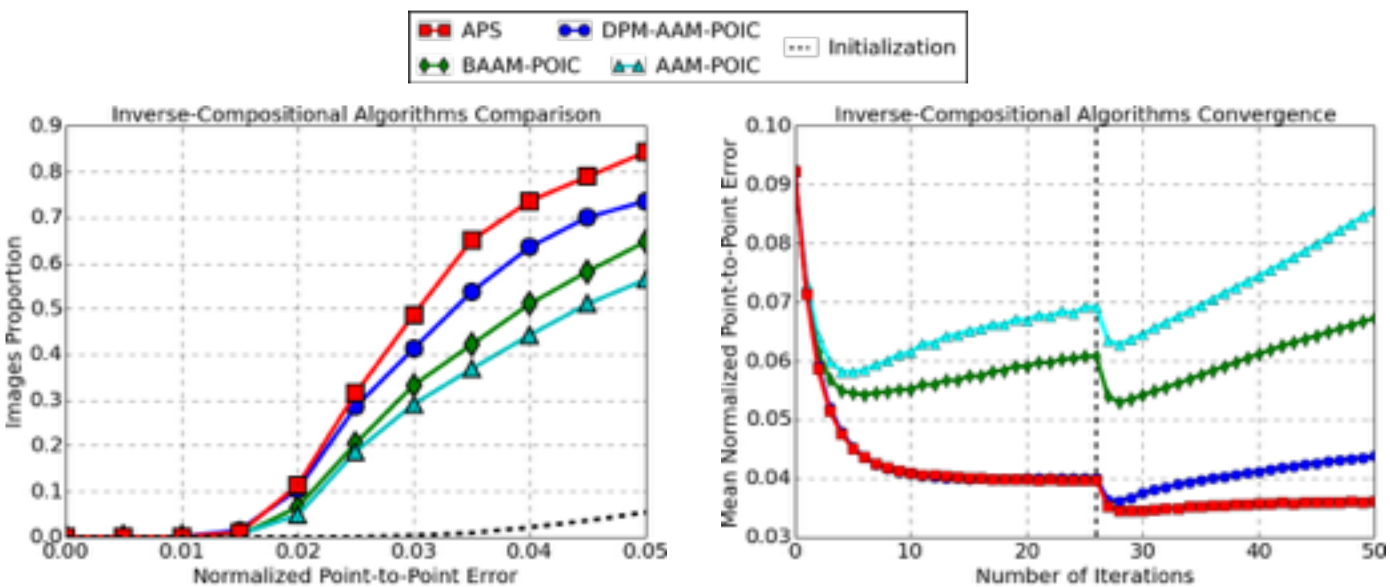
Active Pictorial Structures

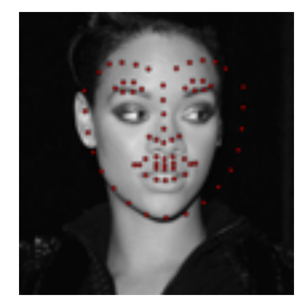


Active Pictorial Structures

$$\Delta \mathbf{p} = \underbrace{(\mathbf{H}_a + \lambda \mathbf{H}_s)^{-1}}_{\text{pre-computed}} \left[\underbrace{\mathbf{J}^T \mathbf{Q}_a}_{\text{pre-computed}} \overbrace{(\mathbf{I}_{\mathcal{F}}(\mathcal{S}(\mathbf{p})) - \bar{\mathbf{a}})}^{\text{residual}} + \underbrace{\lambda \mathbf{H}_s \mathbf{p}}_{\text{pre-computed}}\right]$$

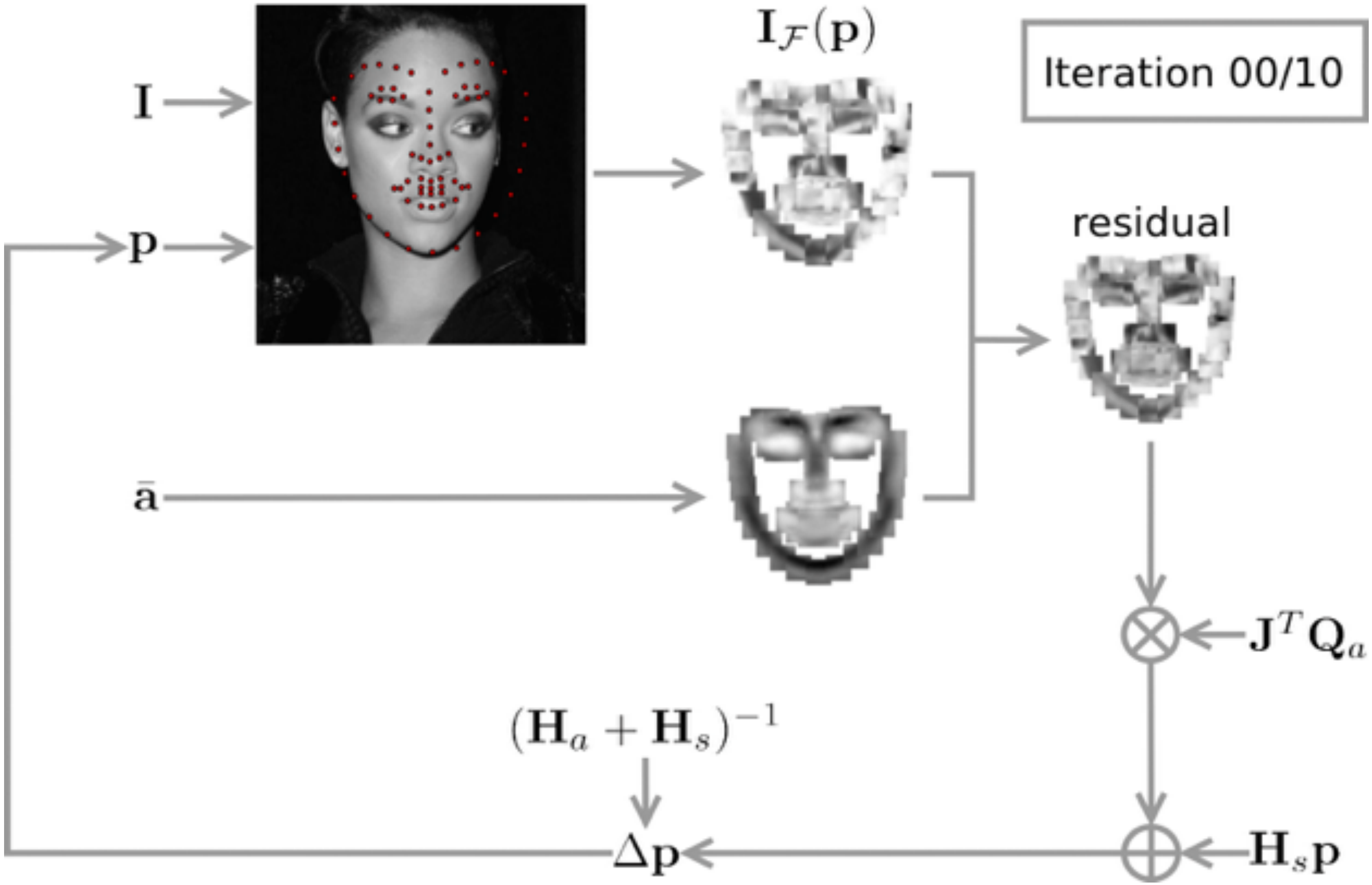
- Weighted Inverse Compositional Gauss-Newton with fixed Jacobian and Hessian
- Fastest inverse compositional fitting with fastest convergence rate
- Deformation prior makes model very robust





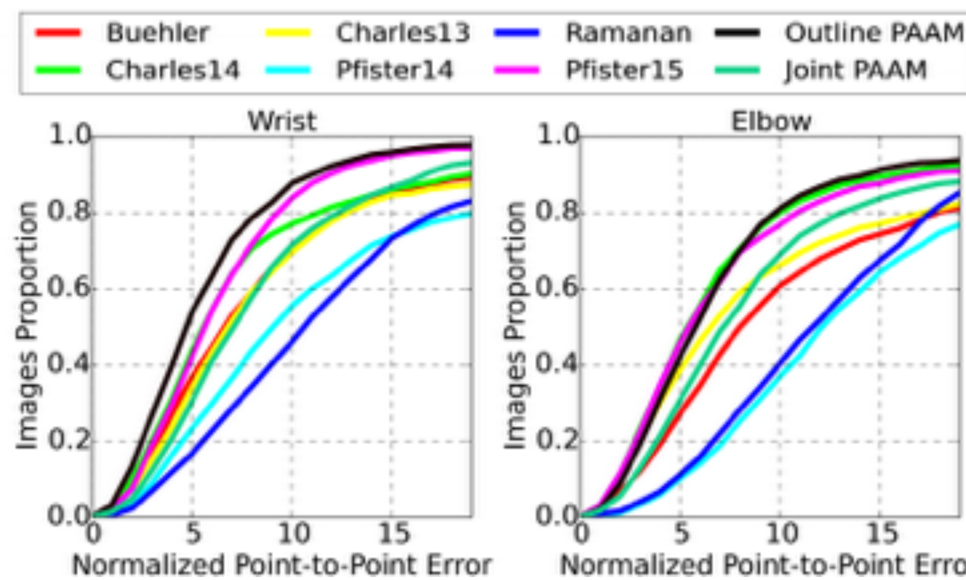
[1] **E. Antonakos** and S. Zafeiriou. “Active Pictorial Structures”, CVPR 2015.

Active Pictorial Structures



[1] E. Antonakos, J. Alabort-i-Medina and S. Zafeiriou. "Active Pictorial Structures", CVPR, 2015.

Active Pictorial Structures



Discriminative Alignment

Learning the Descent Directions

Gauss-Newton $\mathbf{p}_o \leftarrow \mathbf{p}_c - (\mathbf{J}_\mathbf{x}^T \mathbf{J}_\mathbf{x})^{-1} \mathbf{J}_\mathbf{x}^T \mathbf{e}(\mathbf{p}_c)$

Newton $\mathbf{p}_o \leftarrow \mathbf{p}_c - \mathbf{H} \mathbf{J}^T \mathbf{e}(\mathbf{p}_c)$

General Descent Direction $\mathbf{p}_o \leftarrow \mathbf{p}_c + \mathbf{W}\mathbf{x}(\mathbf{p}_c) + \mathbf{b}$

Instead of computing the Jacobian from the image at each iteration why not pre-compute it?

[1] Xiong, X., and F. De la Torre. "Supervised descent method and its applications to face alignment." CVPR, 2013.
[2] Asthana, A., Zafeiriou, S., Cheng, S., & Pantic, M. Robust discriminative response map fitting with constrained local models, CVPR, 2013
[3] Asthana, A., Zafeiriou, S., Cheng, S., & Pantic, M. Incremental face alignment in the wild. In CVPR 2014
[4] Asthana, A., Zafeiriou, S., Tzimiropoulos, G., Cheng, S., & Pantic, M. From pixels to response maps: Discriminative image filtering for face alignment in the wild. IEEE T-PAMI, 37(6), 1312-1320. 2015

Discriminative Alignment

Ground truth

Perturbation

Solve: $$\mathbf{W}_o, \mathbf{b}_o = \arg\min_{\mathbf{W},\mathbf{b}} \sum_i ||\Delta\mathbf{p}_c^i - (\mathbf{W}\mathbf{x}(\mathbf{p}_c^i) + \mathbf{b})||^2$$

Learn the mapping

$$\mathbf{W}_o^{(1)}, \mathbf{b}_o^{(1)} = \arg\min_{\mathbf{W}^{(1)},\mathbf{b}^{(1)}} \sum_i ||\Delta \mathbf{p}_1^i - (\mathbf{W}^{(1)}\mathbf{x}(\mathbf{p}_1^i) + \mathbf{b}^{(1)})||^2$$

$$\mathbf{W}_o^{(2)}, \mathbf{b}_o^{(2)} = \arg\min_{\mathbf{W}^{(2)},\mathbf{b}^{(2)}} \sum_i ||\Delta \mathbf{p}_2^i - (\mathbf{W}^{(2)}\mathbf{x}(\mathbf{p}_2^i) + \mathbf{b}^{(2)})||^2$$

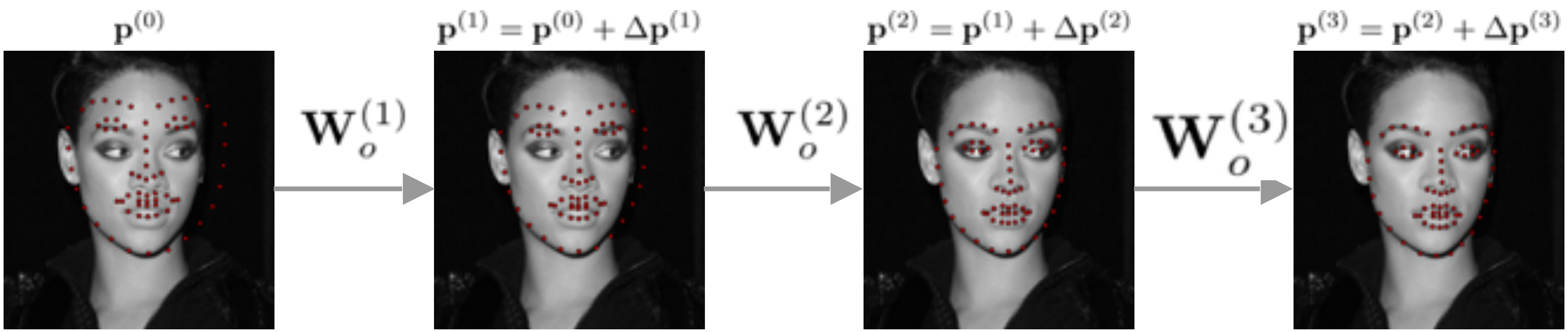
$$\mathbf{W}_o^{(3)}, \mathbf{b}_o^{(3)} = \arg\min_{\mathbf{W}^{(3)},\mathbf{b}^{(3)}} \sum_i ||\Delta \mathbf{p}_3^i - (\mathbf{W}^{(3)}\mathbf{x}(\mathbf{p}_3^i) + \mathbf{b}^{(3)})||^2$$

$$\mathbf{p}_2^i = \mathbf{p}_1^i + \mathbf{W}_o^{(1)}\mathbf{x}(\mathbf{p}_1^i) + \mathbf{b}^{(1)}$$

$$\mathbf{p}_3^i = \mathbf{p}_2^i + \mathbf{W}_o^{(2)}\mathbf{x}(\mathbf{p}_2^i) + \mathbf{b}^{(2)}$$

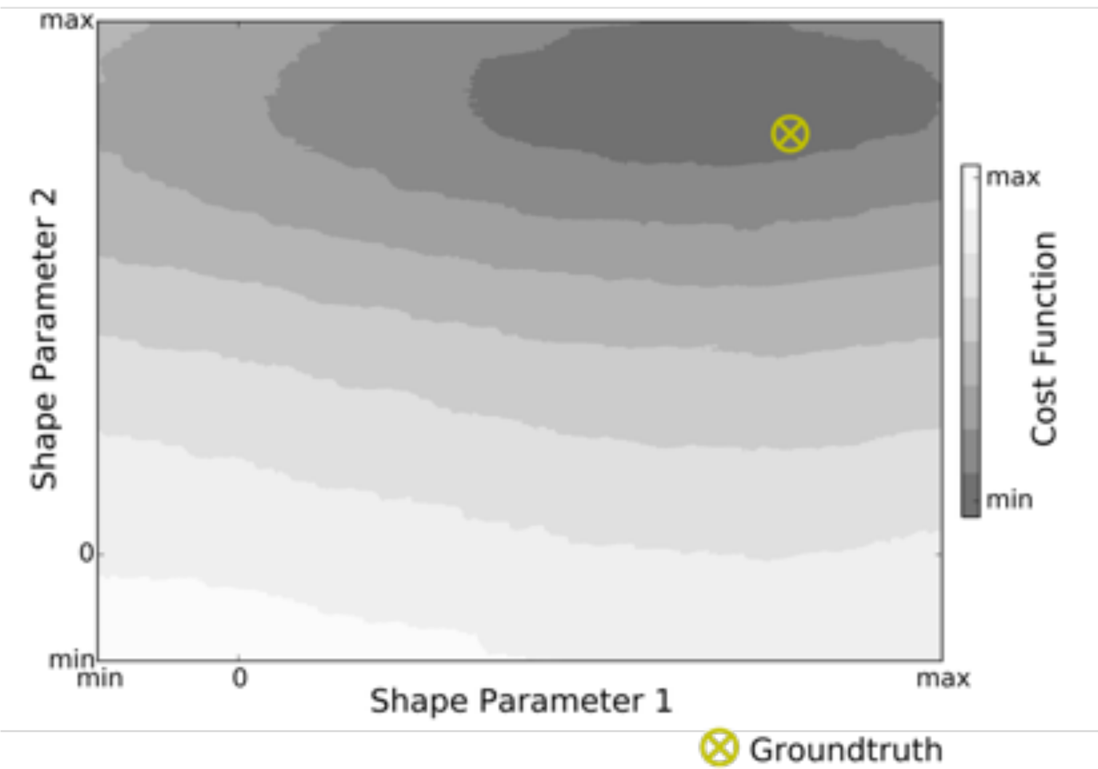
$$\mathbf{p}_4^i = \mathbf{p}_3 + \mathbf{W}_o^{(3)}\mathbf{x}(\mathbf{p}_3^i) + \mathbf{b}^{(3)}$$

Applying the mapping



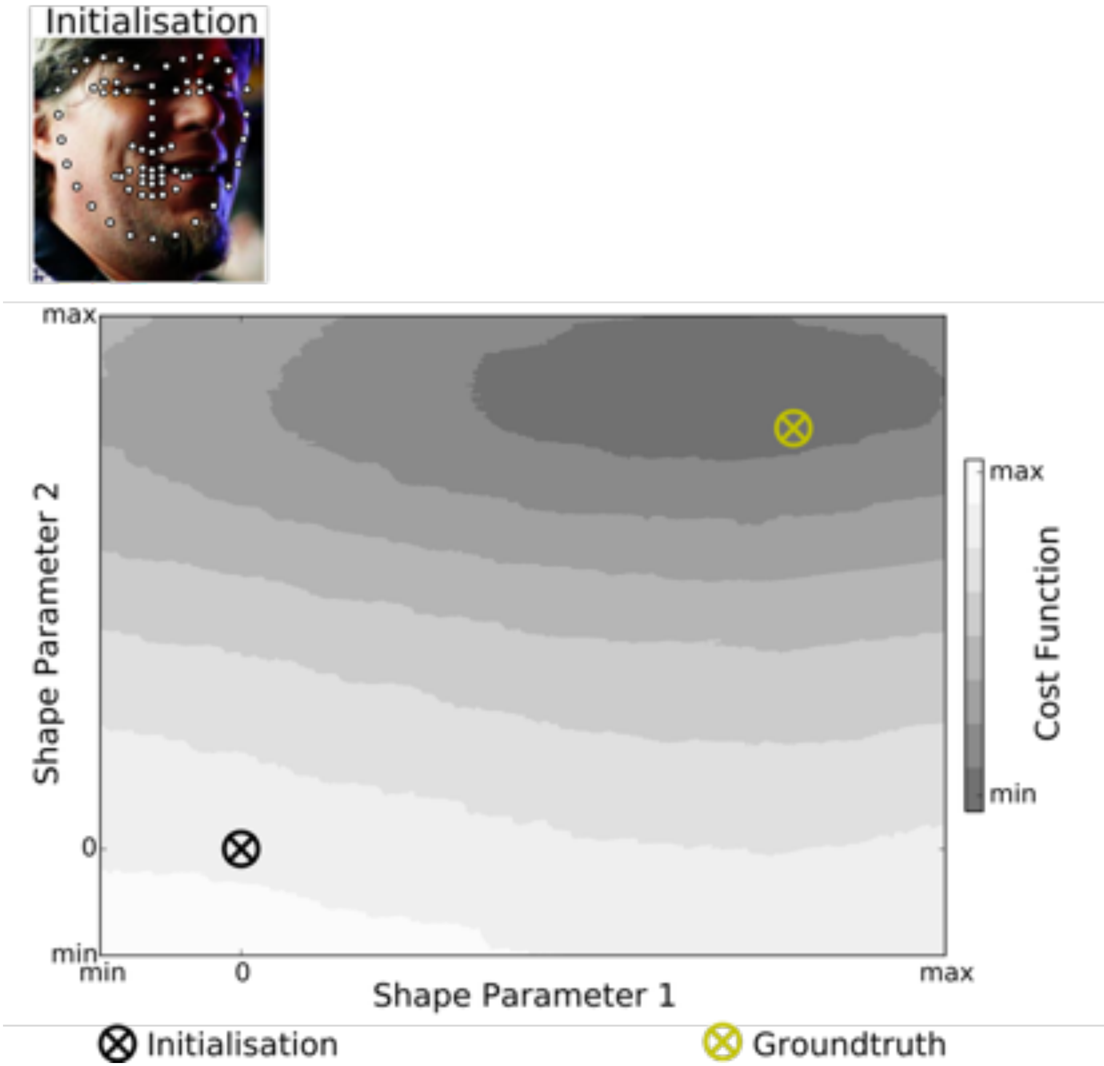
Adaptive Cascaded Regression

Motivation



Cost function with respect to parametric shape model.

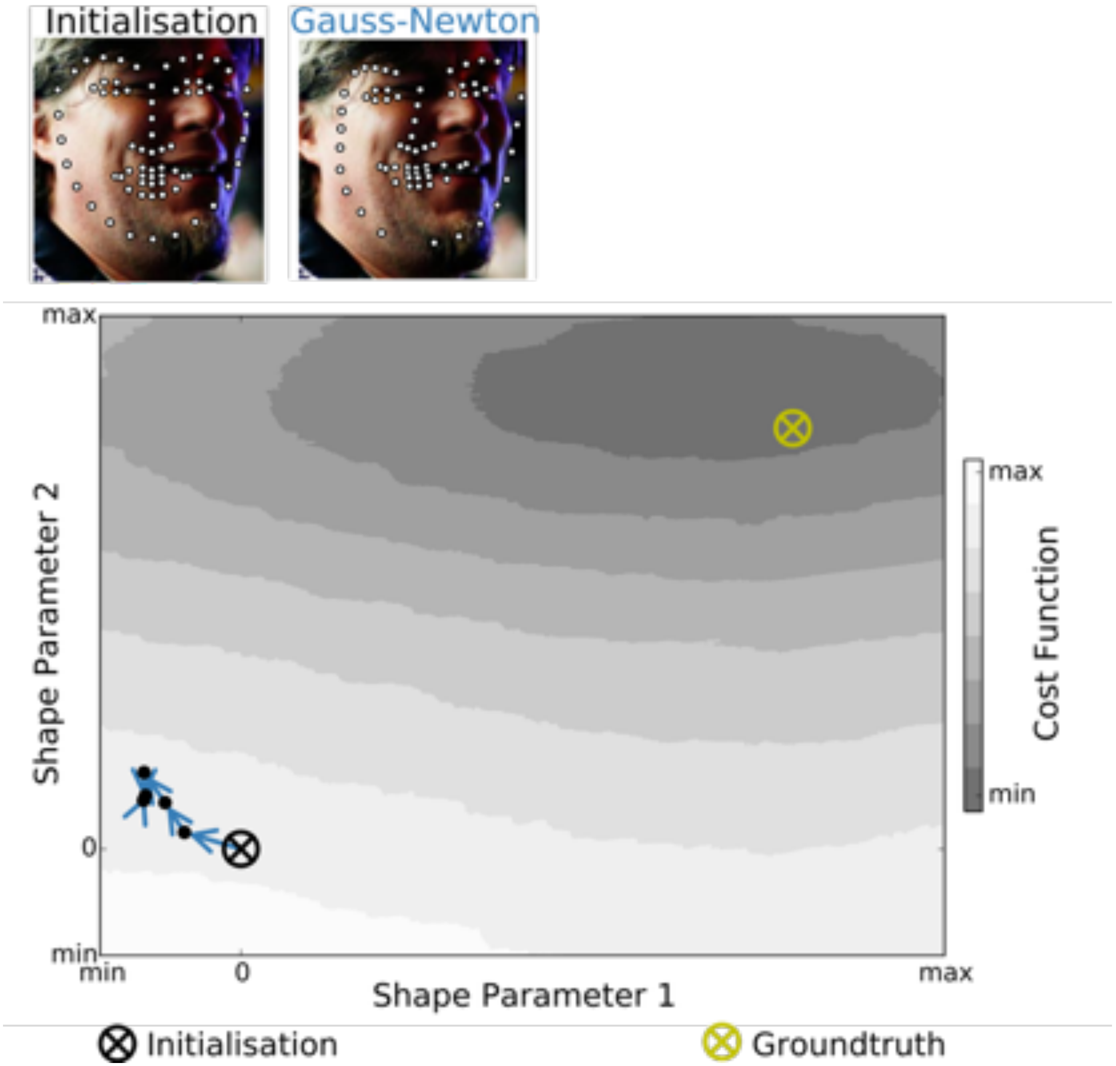
Motivation



Cost function with respect to parametric shape model.

Realistic initialization from face detector.

Motivation

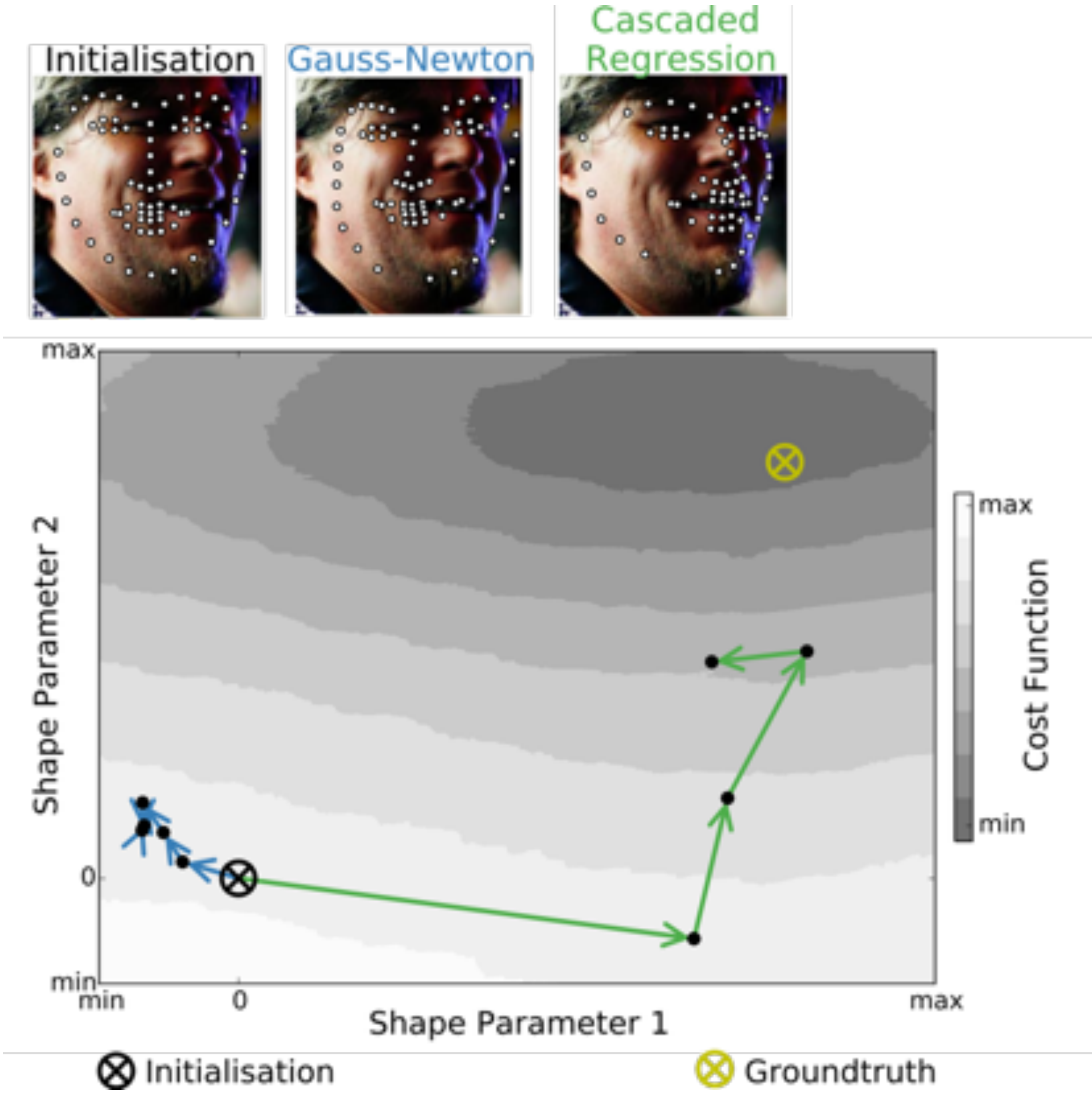


Cost function with respect to parametric shape model.

Realistic initialization from face detector.

Gauss-Newton fails due to bad initialization.

Motivation



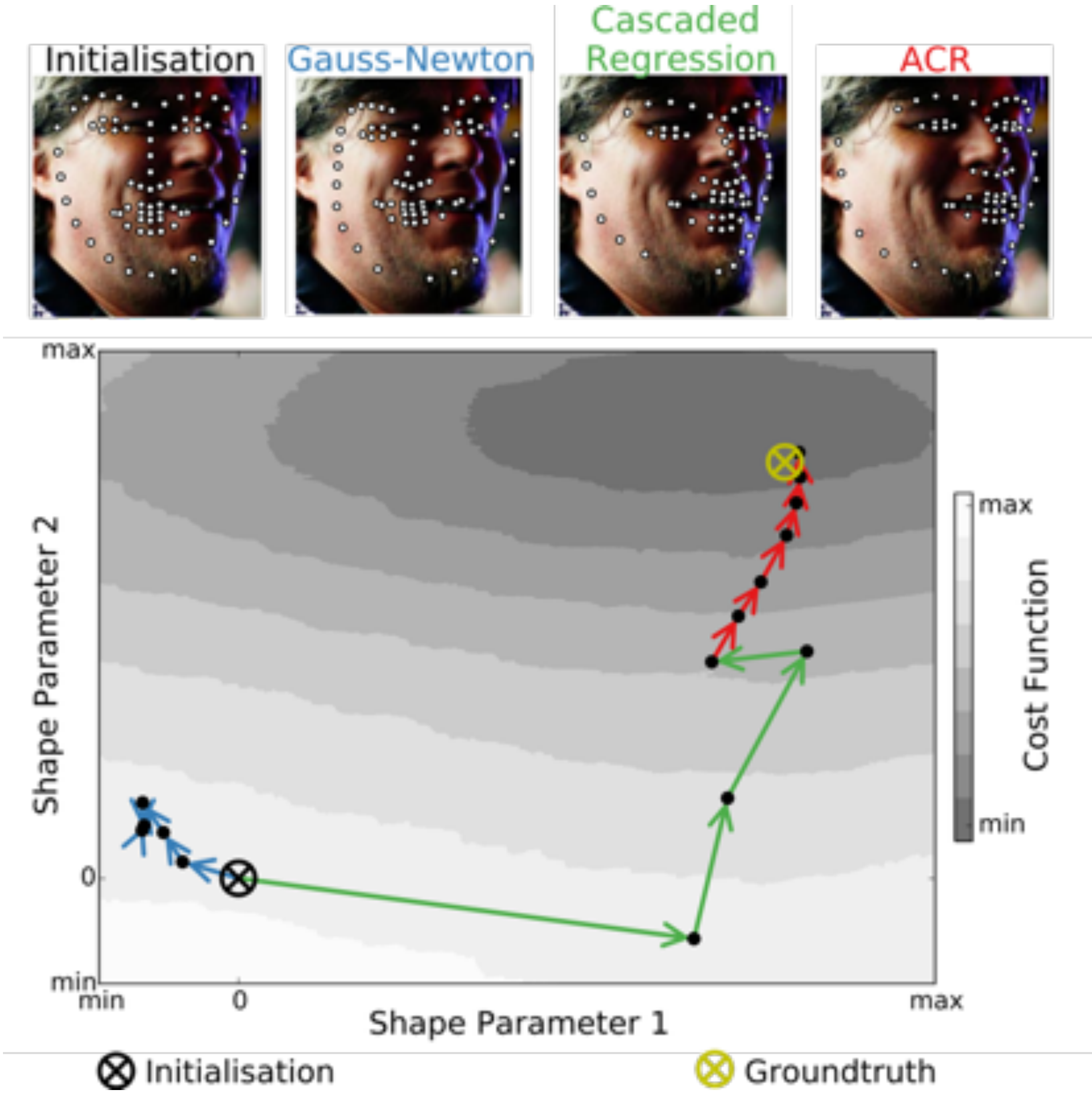
Cost function with respect to parametric shape model.

Realistic initialization from face detector.

Gauss-Newton fails due to bad initialization.

Cascaded regression moves towards the correct direction but not close enough.

Motivation



Cost function with respect to parametric shape model.

Realistic initialization from face detector.

Gauss-Newton fails due to bad initialization.

Cascaded regression moves towards the correct direction but not close enough.

Applying Gauss-Newton descent directions right after cascaded regression directions reaches optimum!

Adaptive Cascaded Regression

Projected-Out Residual

$\hat{\phi}(\mathbf{p}) = \hat{\mathbf{P}}\left(\mathbf{I}_{\mathcal{F}}(\mathbf{p}) - \bar{\mathbf{a}}\right)$ with $\hat{\mathbf{P}} = \mathbf{E} - \mathbf{U}_a {\mathbf{U}_a}^T$

Discriminative

Cascaded Regression descent directions

$\Delta\mathbf{p}^{(k)} = \mathbf{R}^{(k)}\hat{\phi}(\mathbf{p}^{(k)})$

Generative

Gauss-Newton descent directions

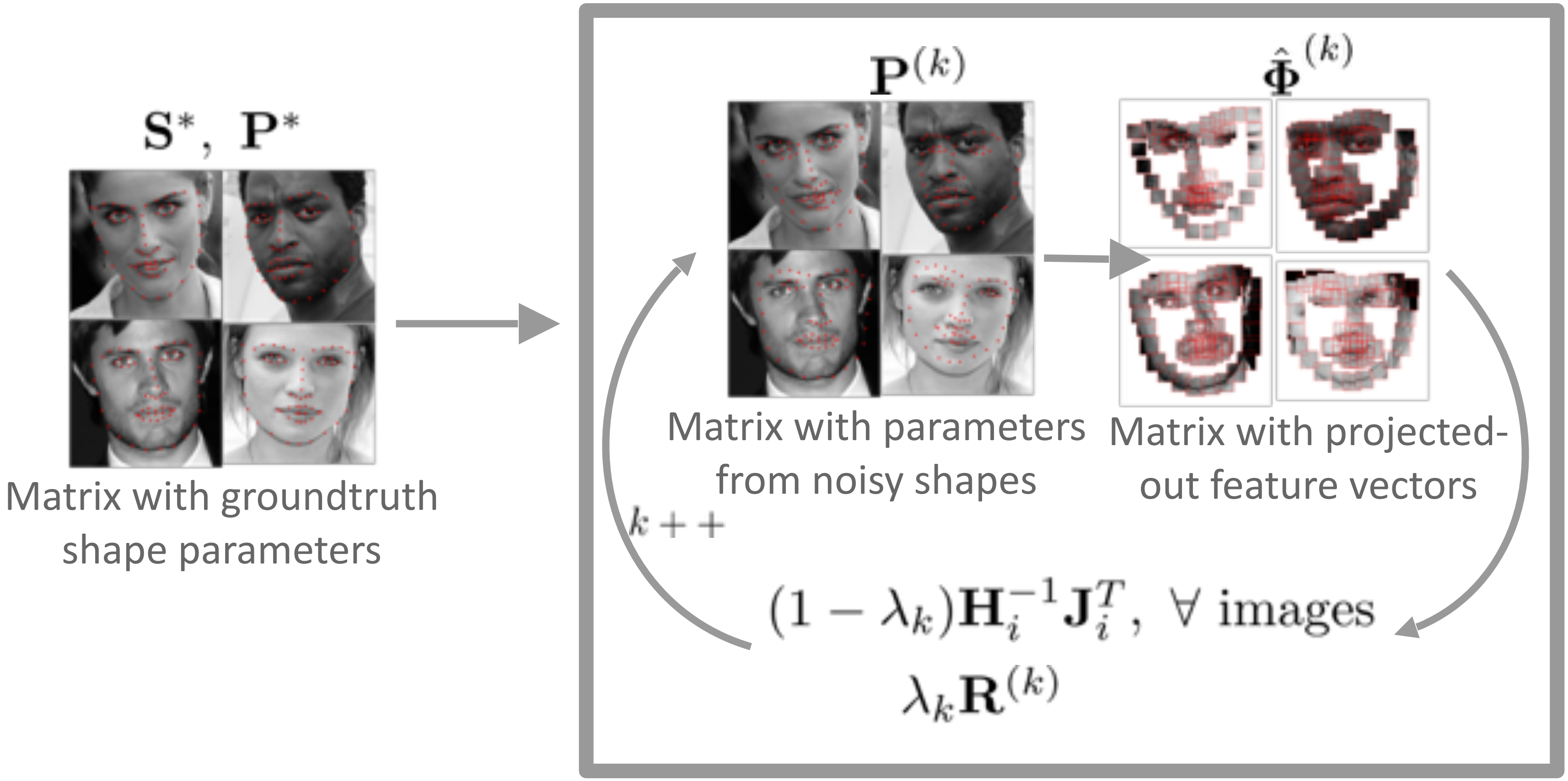
$\Delta\mathbf{p} = \mathbf{H}^{-1}\mathbf{J}^T\hat{\phi}(\mathbf{p})$

Adaptive Cascaded Regression

Linear combination of descent directions with **adaptive weights**:

$$\Delta\mathbf{p}^{(k)} = \underbrace{[\lambda_k\mathbf{R}^{(k)} - (1-\lambda_k)\mathbf{H}^{-1}\mathbf{J}^T]}_{\text{Adaptive descent steps}}\hat{\phi}(\mathbf{p}^{(k-1)})$$

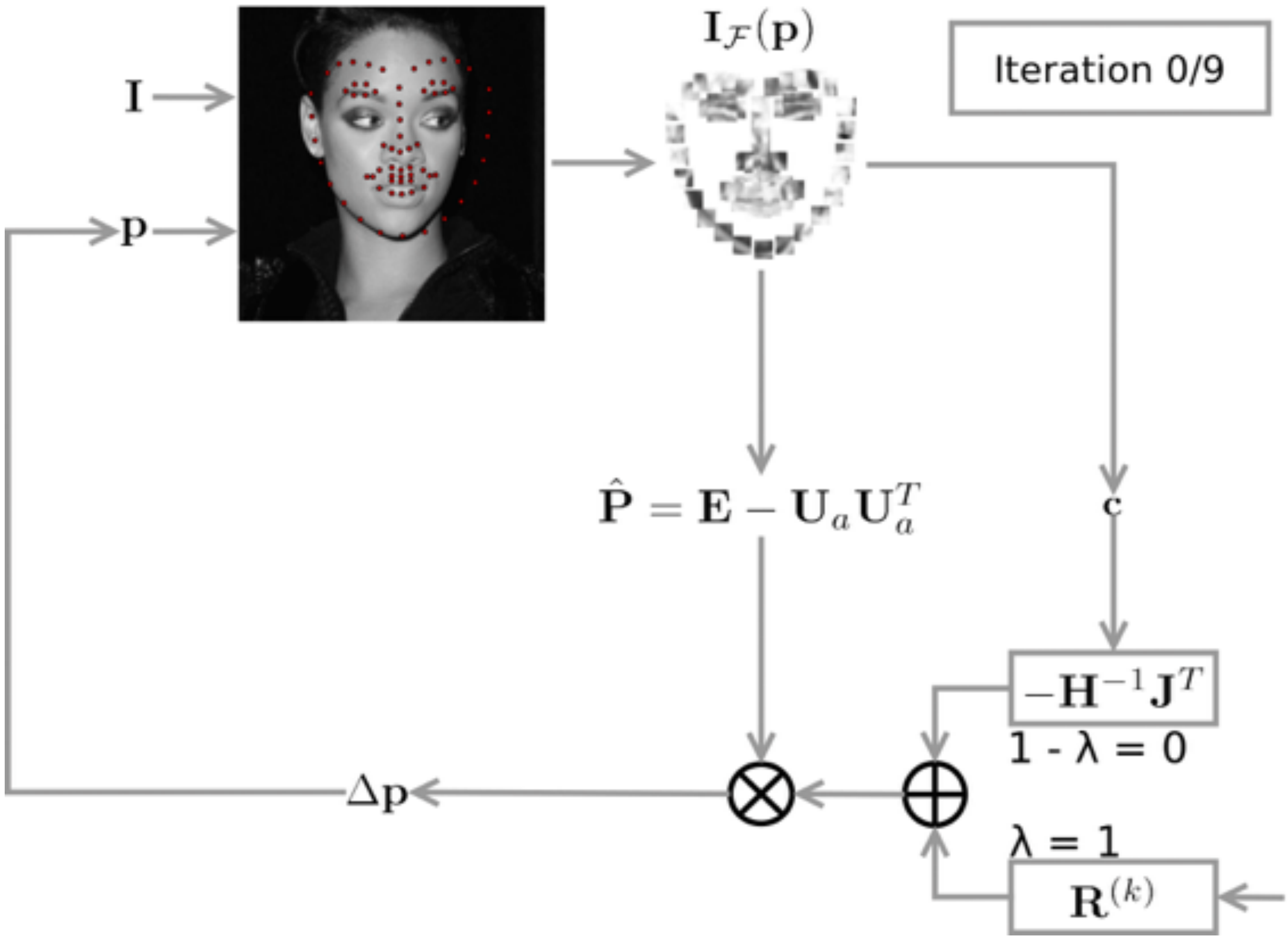
Adaptive Cascaded Regression



$$\boldsymbol{\lambda} = [1, 0.75, 0.5, 0.25, 0, 0, \ldots]$$

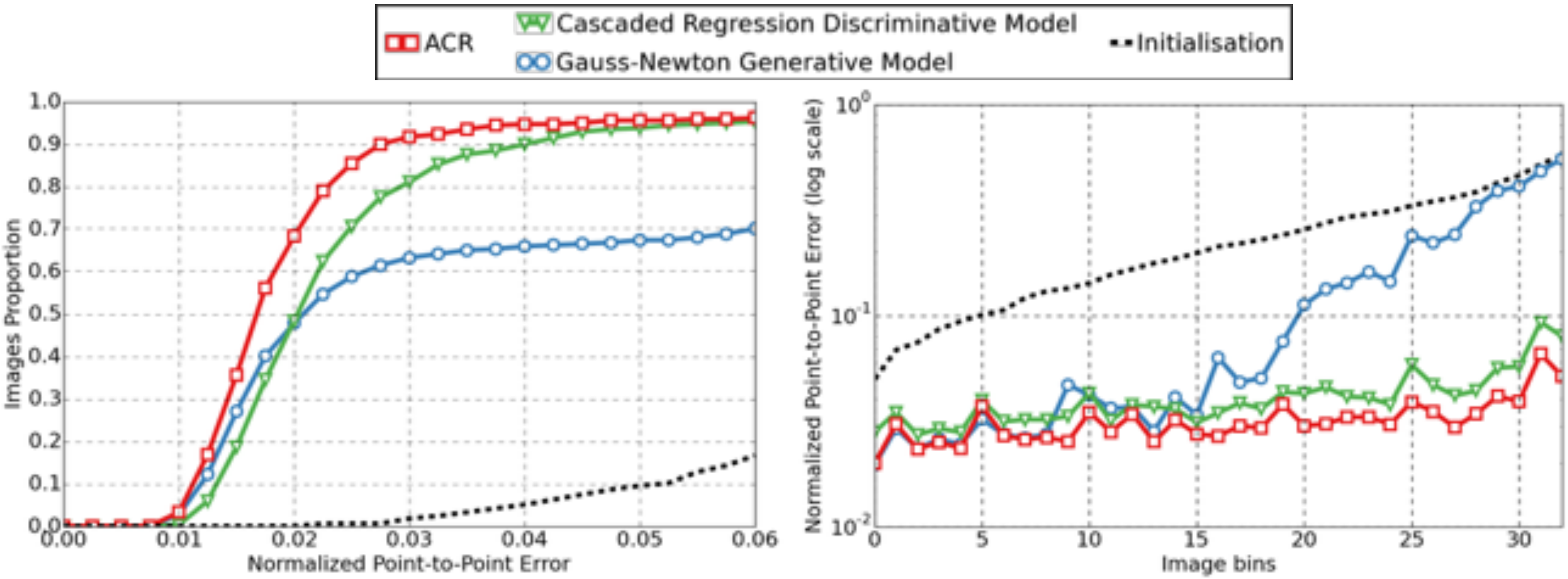
[1] **E. Antonakos***, P. Snape*, G. Trigeorgis and S. Zafeiriou. "Adaptive Cascaded Regression", *oral*, ICIP 2016.

Adaptive Cascaded Regression



[1] **E. Antonakos***, P. Snape*, G. Trigeorgis and S. Zafeiriou. "Adaptive Cascaded Regression", *oral*, ICIP 2016.

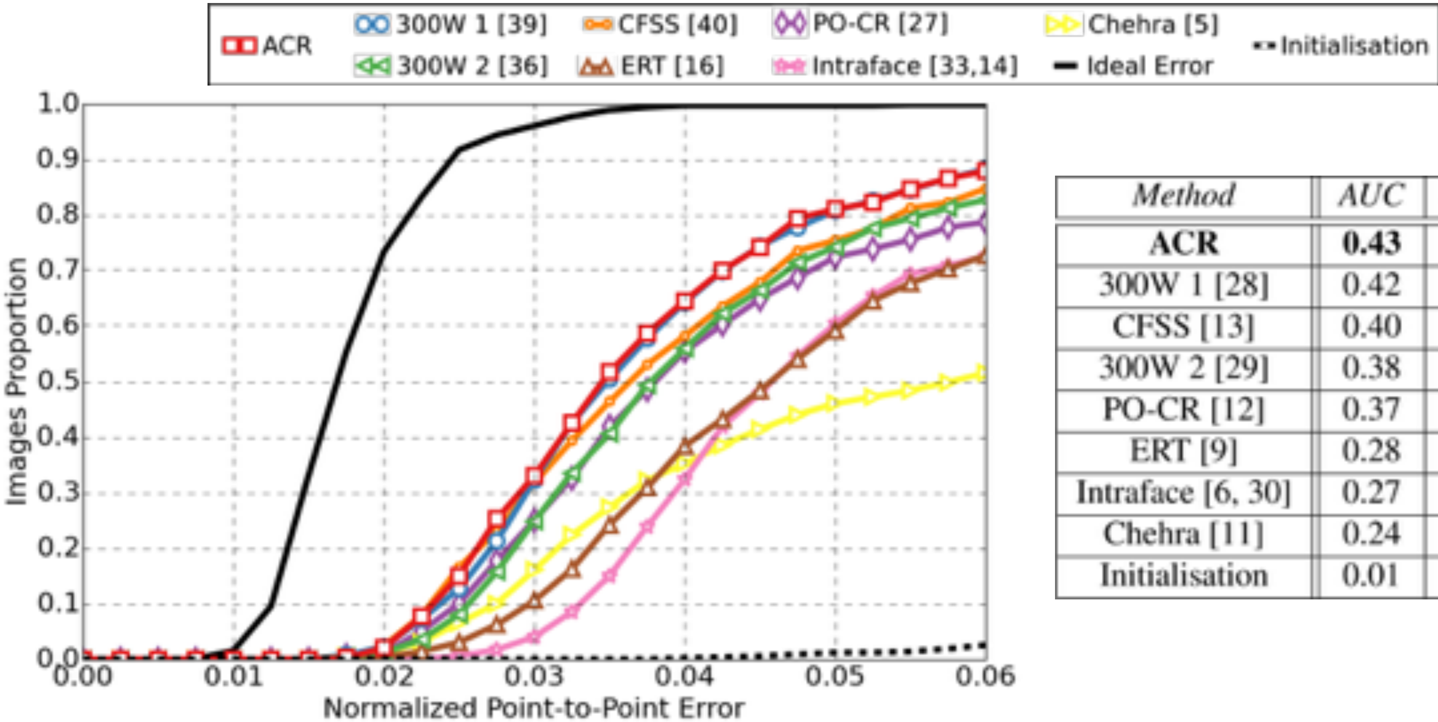
Adaptive Cascaded Regression



Consistently lower error compared to cascaded regression discriminative and to Gauss-Newton generative model.

[1] **E. Antonakos***, P. Snape*, G. Trigeorgis and S. Zafeiriou. "Adaptive Cascaded Regression", *oral*, ICIP 2016.

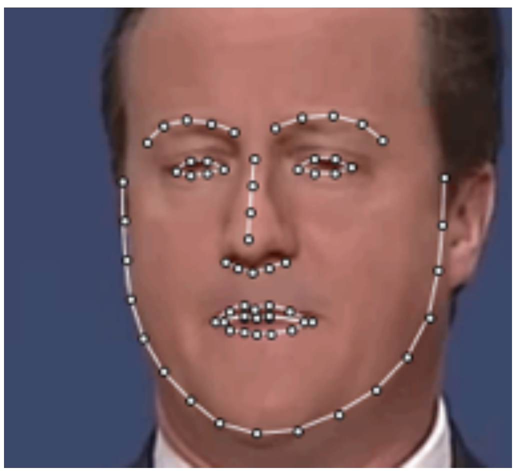
Adaptive Cascaded Regression

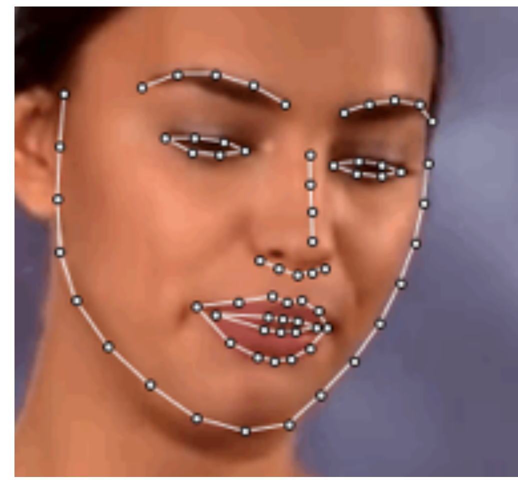


Method	*AUC*	*Failure rate (%)*
ACR	**0.43**	11.0
300W 1 [28]	0.42	**9.3**
CFSS [13]	0.40	13.5
300W 2 [29]	0.38	14.2
PO-CR [12]	0.37	17.7
ERT [9]	0.28	23.7
Intraface [6, 30]	0.27	23.8
Chehra [11]	0.24	46.8
Initialisation	0.01	96.8

[ACR] **E. Antonakos***, P. Snape*, G. Trigeorgis and S. Zafeiriou. “Adaptive Cascaded Regression”, *oral*, ICIP 2016.
[300W 1] E. Zhou et al. “Extensive facial landmark localization with coarse-to-fine convolutional network cascade”, ICCV-W 2013.
[CFSS] S. Zhu et al. “Face alignment by coarse-to-fine shape searching”, CVPR 2015.
[PO-CR] G. Tzimiropoulos. “Project-Out cascaded regression with an application to face alignment”, CVPR 2015.
[300W 2] J. Yanet al. “Learn to combine multiple hypotheses for accurate face alignment”, ICCV-W 2013.
[ERT] V. Kazemi and J. Sullivan. “One millisecond face alignment with an ensemble of regression trees”, CVPR 2014.
[Intraface] X. Xiong and F. De la Torre. “Supervised Descent Method and its Applications to Face Alignment”, CVPR 2013.
[Chehra] A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic. “Incremental face alignment in the wild”, CVPR 2014.

Adaptive Cascaded Regression





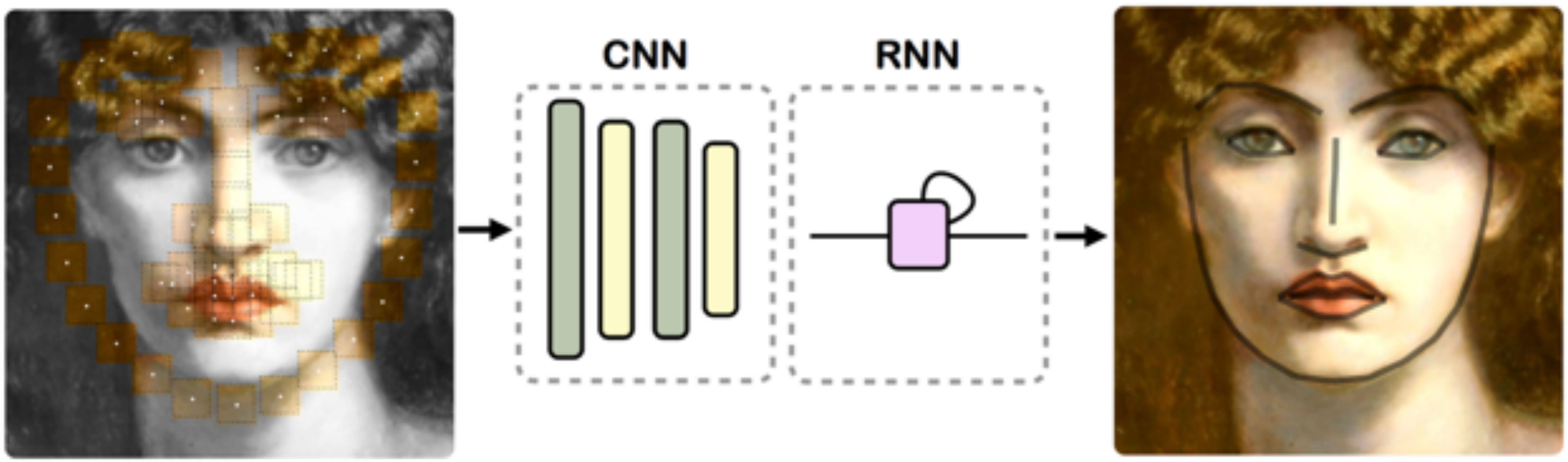
Mnemonic Descent

End-to-End Training of an non-linear cascade regression method

- Why HoGs and not any other features? (Learn a non-linear transform for features)

- The process bears similarities with a dynamical system. Why not model it explicitly model? (Learn a non-linear dynamical system that models the cascade).

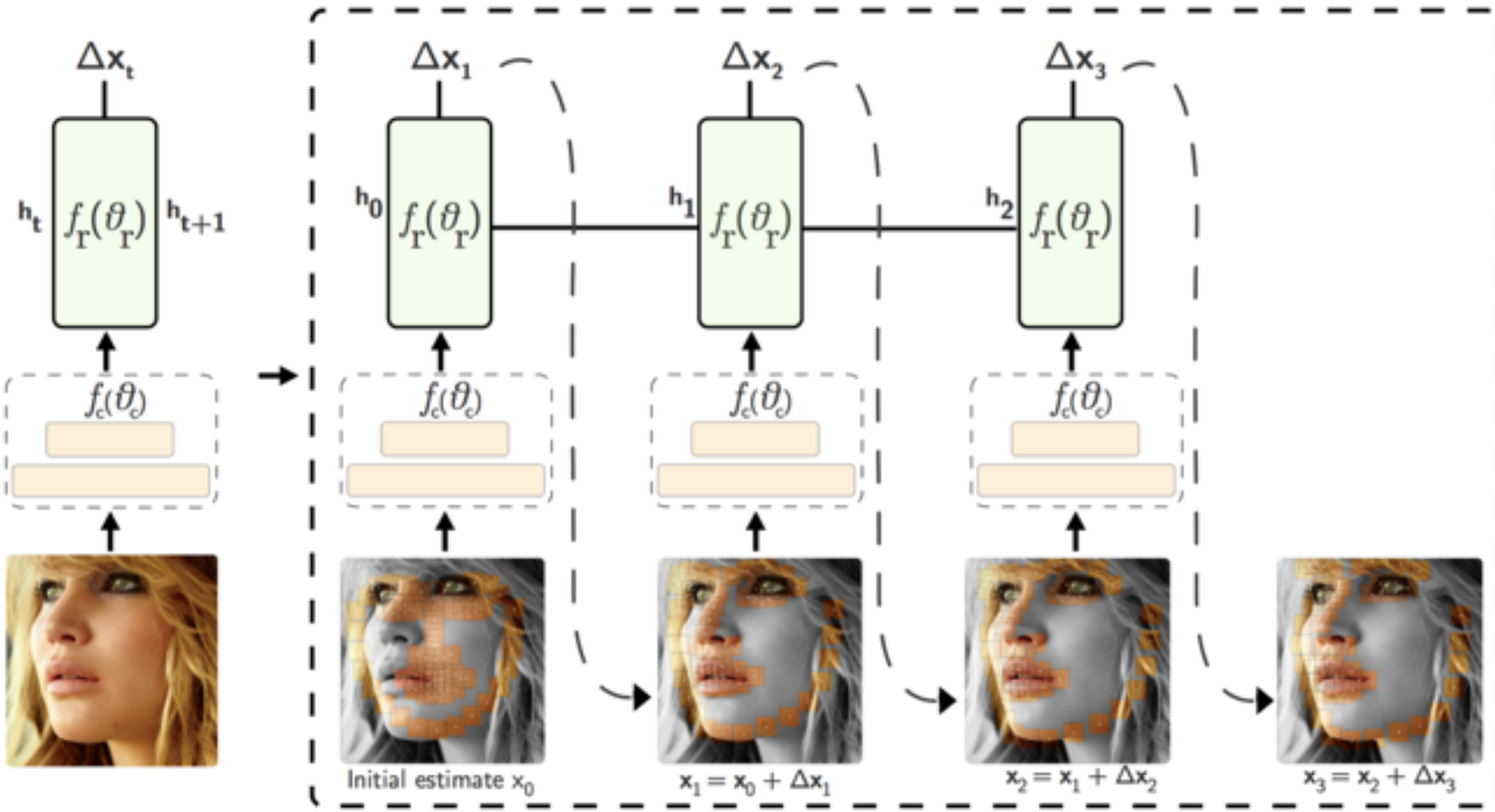
Mnemonic Descent Method



Input image
aligned with the mean face
Final shape estimate

Trigeorgis, G., Snape, P., Nicolaou, M. A., Antonakos, E., & Zafeiriou, S. (2016, June). Mnemonic Descent Method: A recurrent process applied for end-to-end face alignment. CVPR'16, Las Vegas, NV, USA.

Mnemonic Descent Method



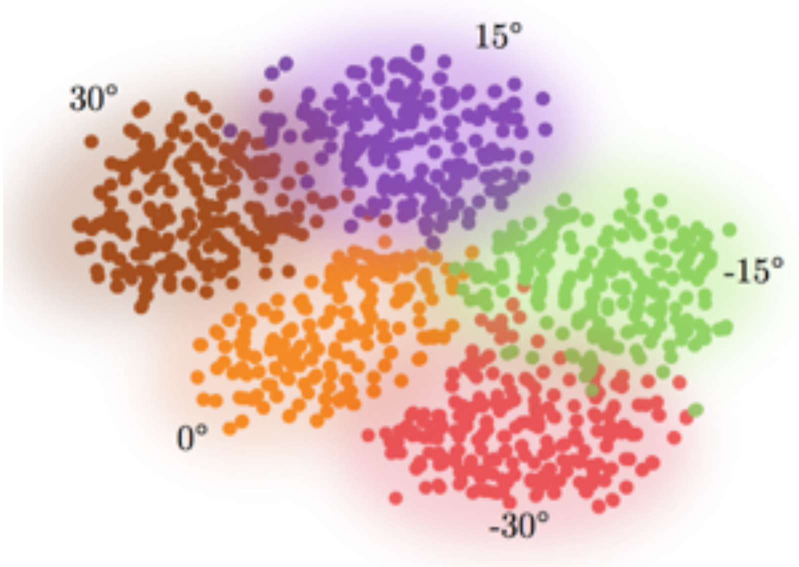
Mnemonic Descent Method

- Updates

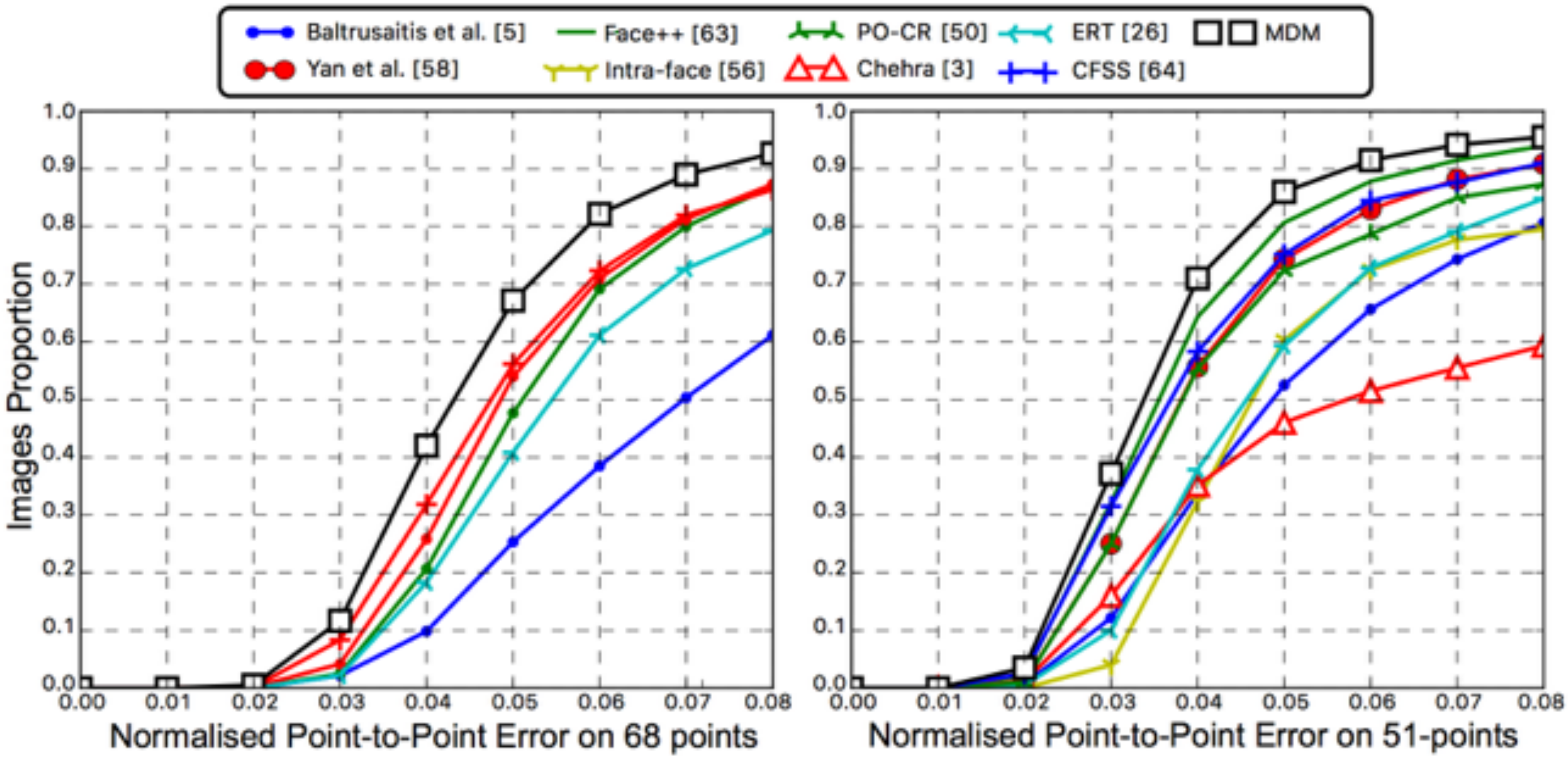
$$\mathbf{h}^{(t+1)} = \tanh(\mathbf{W}_{hi}\phi(\mathbf{z};\mathbf{x}^{(t)}) + \mathbf{W}_{hh}\mathbf{h}^{(t)})$$

$$\Delta\mathbf{x}^{(t+1)} = \mathbf{x}^{(t)} + \mathbf{W}_{ho}\mathbf{h}^{(t)}$$

Clustering of the state

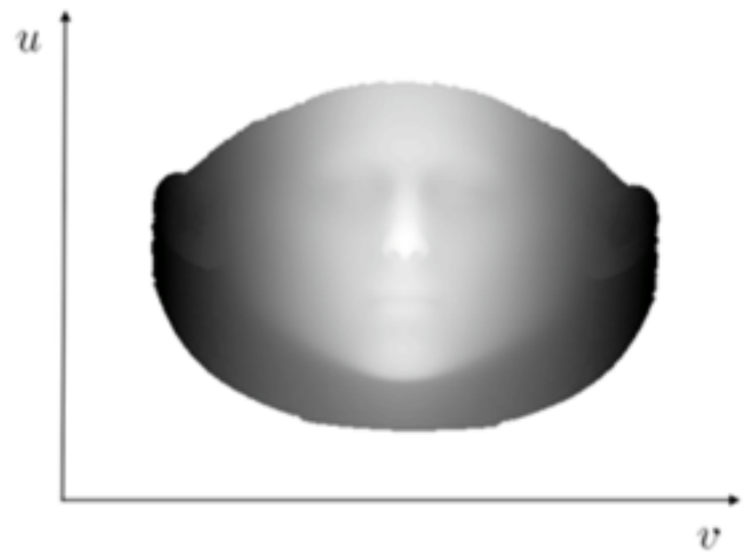


Mnemonic Descent Method



3DMM

3DMM



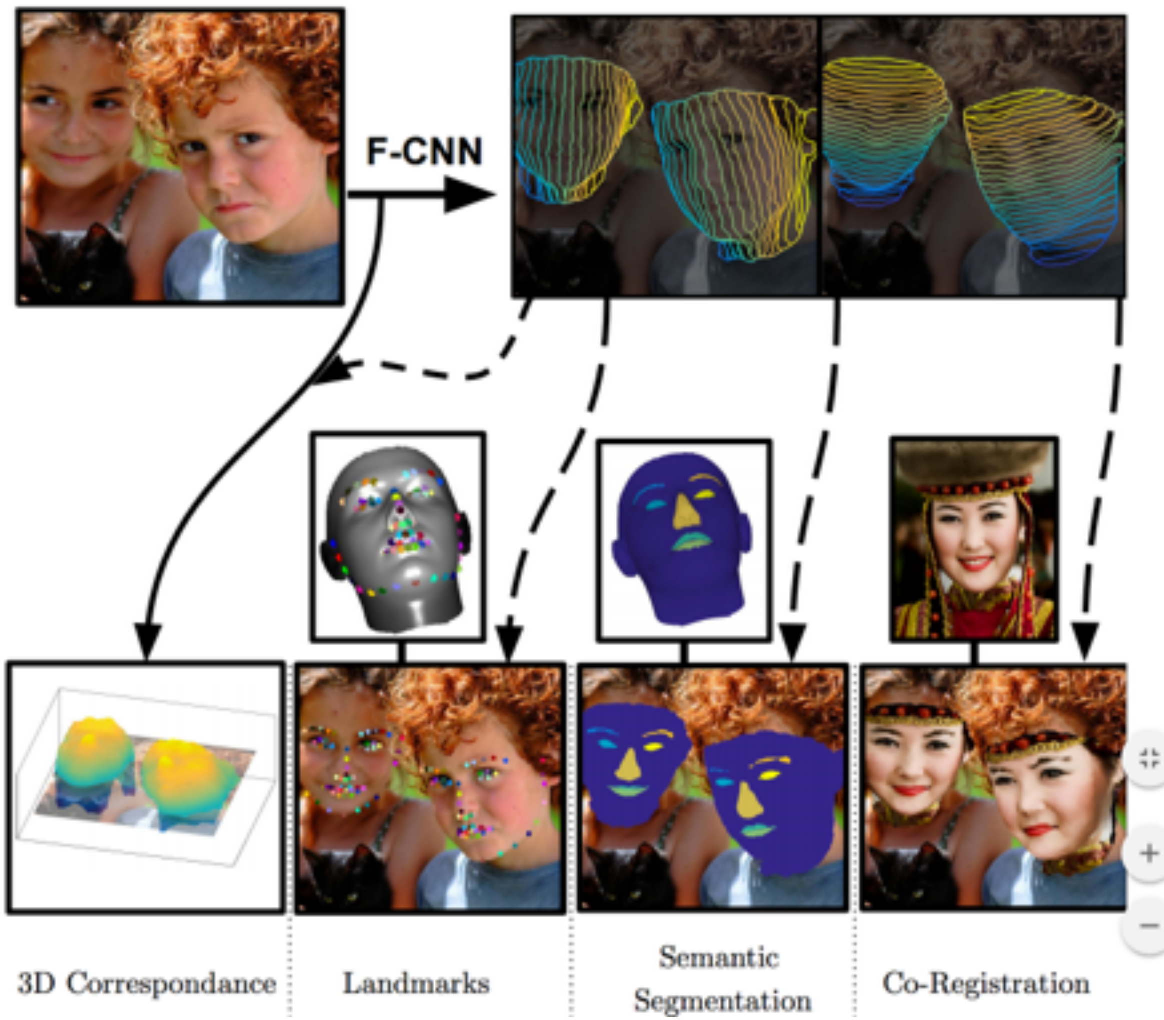
3DMM “in-the-wild”

Fitting a 3DMM

$$\arg\min_{\mathbf{p},\mathbf{c},\boldsymbol{\lambda}} \; \|\mathbf{F}\left(\mathcal{W}(\mathbf{p},\mathbf{c})\right) - \mathcal{T}(\boldsymbol{\lambda})\|^2 + c_l \|\mathcal{W}_l(\mathbf{p},\mathbf{c}) - \mathbf{s}_l\|^2 +$$

$$+ c_s \|\mathbf{p}\|^2_{\boldsymbol{\Sigma}_s^{-1}} + c_t \|\boldsymbol{\lambda}\|^2_{\boldsymbol{\Sigma}_t^{-1}}$$

Deep Learning



Introducing Menpo

- Open Source Python package for generative and discriminative modelling
- Unified framework for AAMs, CLMs, & SDMs
- Under active development, Morphable Models coming
- www.menpo.org
- github.com/menpo

Thank You