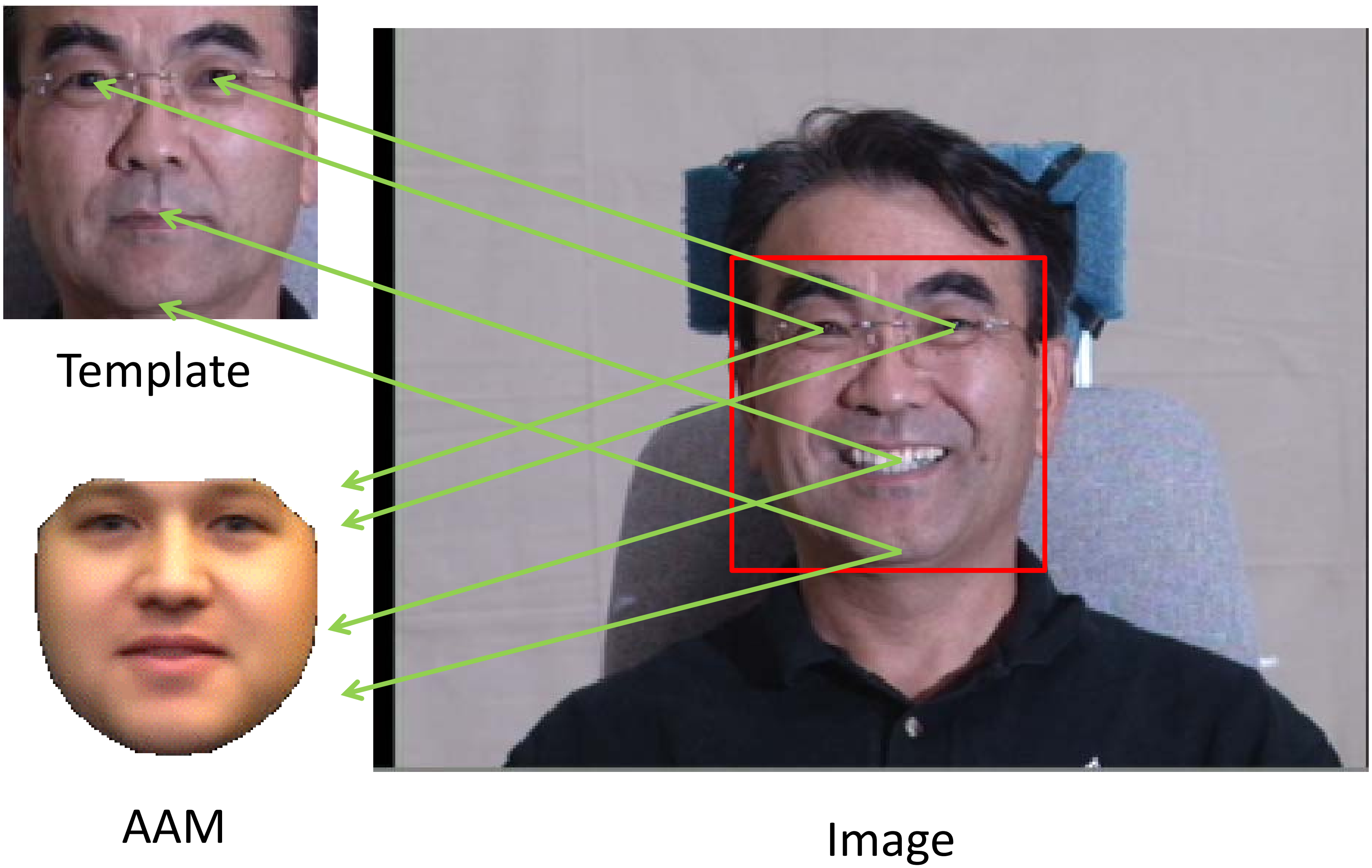
Learning Image Alignment without Local Minima for Face Detection and Tracking

Minh Hoai Nguyen & Fernando de la Torre

Robotics Institute, Carnegie Mellon University

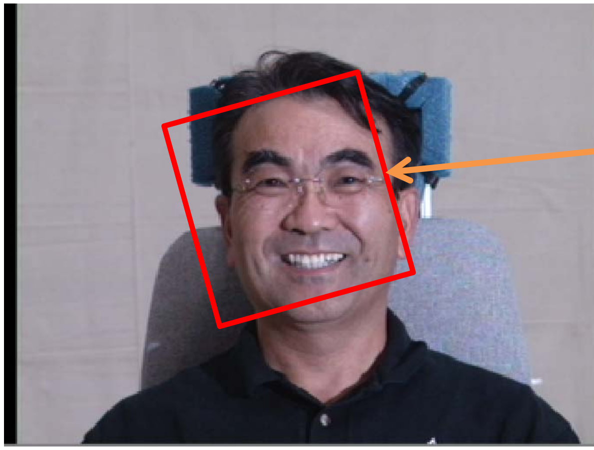
Image alignment



Some work in image alignment

- **Direct methods (pixel domain)**
  - Hierarchical motion estimation [Quam84, Anandan89, Bergen et al 92]
  - Fourier-based [Kuglin & Hines 75, De Castro & Morandi 87, Brown 92, Fleet and Jepson 90, Oppenheim et al 99]
  - Parametric motions [Lucas-Kanade 81, Rehg & Witkin 91, Fuh & Maragos 91, Bergen et al 92, Baker & Matthews 04]
  - Robust metrics [Black & Anandan 96 , Black & Rangarajan 96, Stewart 99, Wimmer et al 06]
- **Feature based**
  - Distinctive features [Hannah 74, Moravec 83, Zoghlami et al 97, Capel & Zisserman 98, Cham & Cipolla 98, Badra et al 98, Mclauchlan & Jaenicke 02, Brown & Lowe 03, Brown et al 05]
- **Model-based methods**
  - Generative
    - Active Appearance Models [Cootes et al 98, Matthews & Baker 04]
    - Morphable models [Jones & Poggio 98, Blanz & Vetter 99]
    - 3D [Movellan 03, Marks 06]
    - Kernel PCA [de la Torre & Nguyen 08]
  - Discriminative
    - SVM [Saragih & Goecke 07]
    - Boosting [Liu 07, Wu et al 08, Whitehill & Movellan 08]

Image alignment as an optimization problem



pixels to align

Template

Image

$$\underset{\mathbf{p}}{\text{minimize}} \, ||\mathbf{d}(\mathbf{f}(\mathbf{x}, \mathbf{p})) - \mathbf{d}_{ref}||_2^2$$

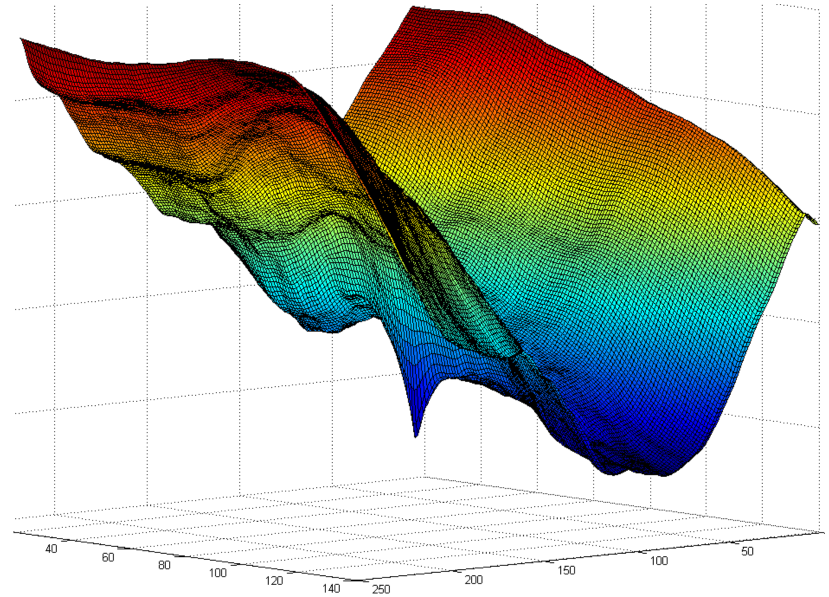
Geometric transformation

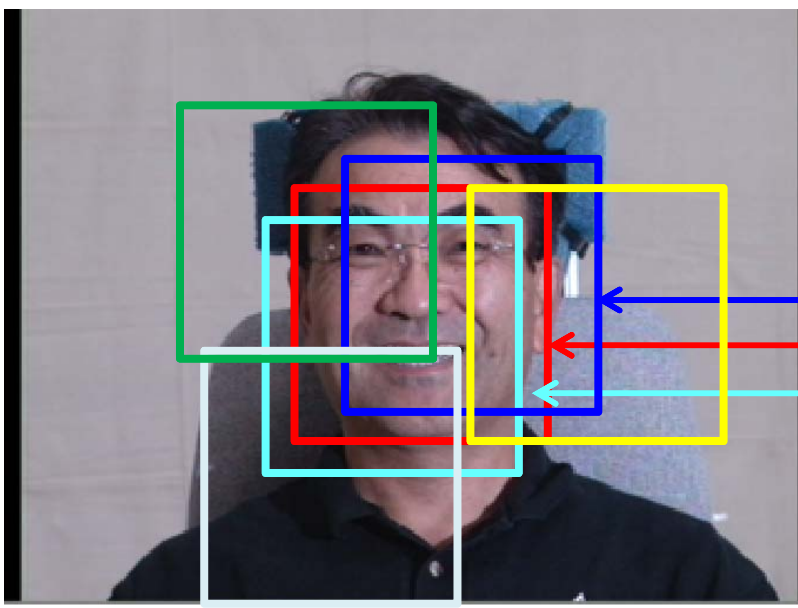
Motion parameters

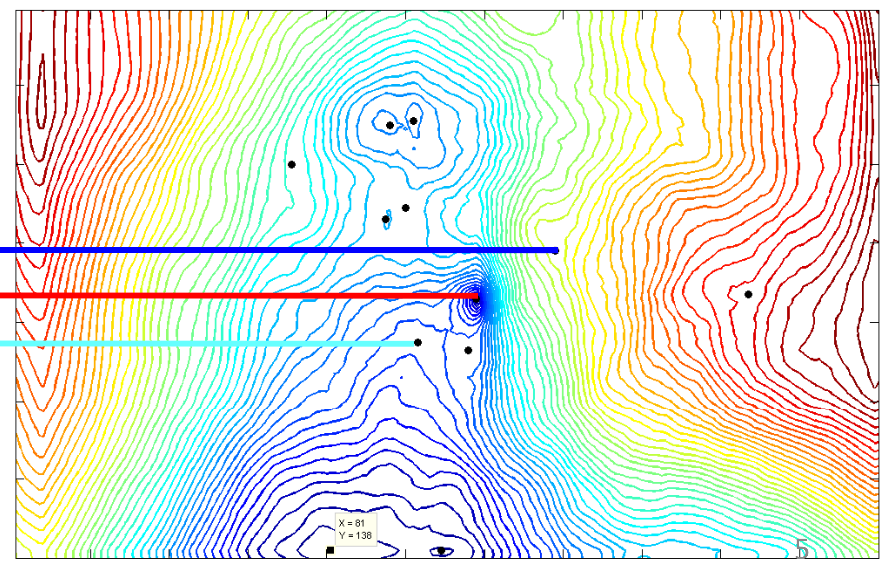
1. Search space ($\mathbf{p}$)
2. Energy function (e.g. L_2)
3. Search strategy (e.g. exhaustive search, branch & bound, gradient-based)

(Lucas-Kanade IUW81, Mathews et al PAMI04)

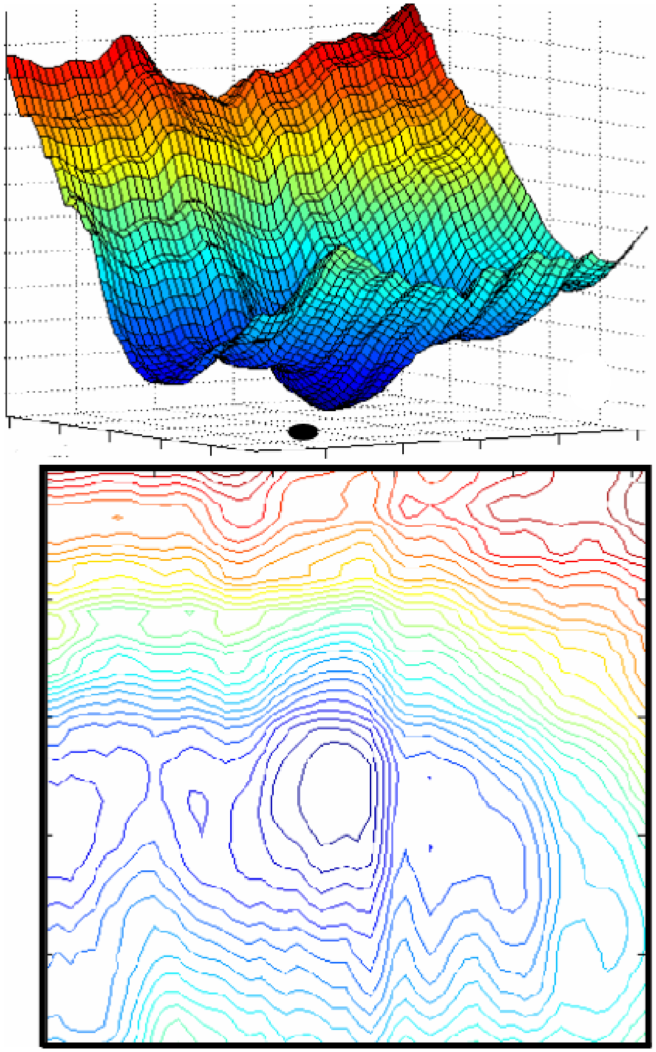
Issues of gradient-based optimization



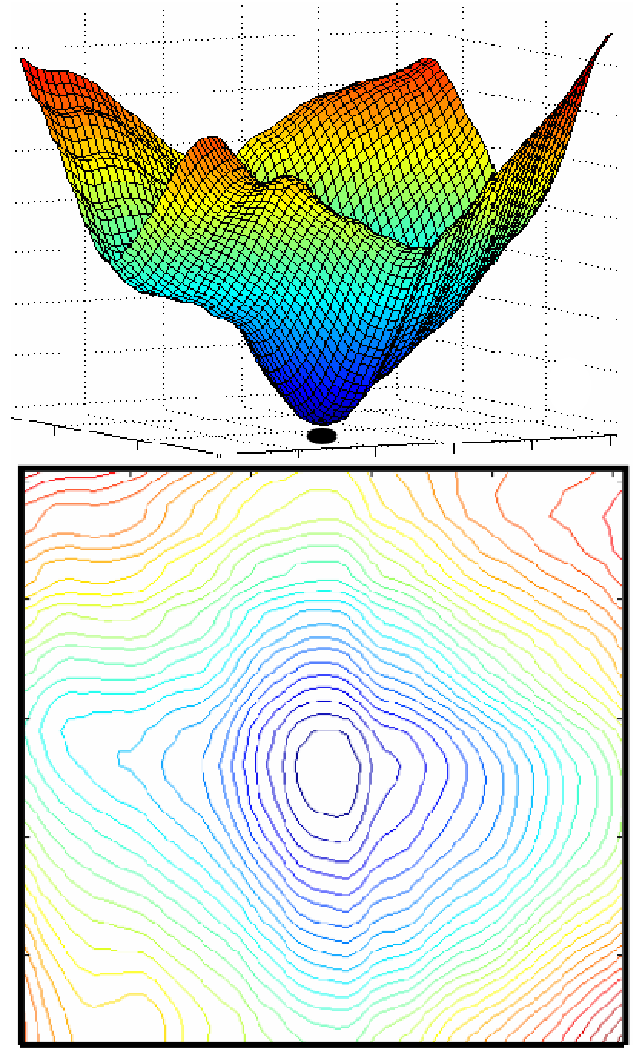




Local minima problems in image alignment



Learn the energy function



(Black & Jepson98, Cootes & Taylor 01, Baker et al 04, de la Torre et al 07)

Local minima at only at right places

(Wimmer et al 06, Wu et al 08)

Quadratic cost function

pixels to align

Template

Image

$$\underset{\mathbf{p}}{\text{minimize}} \, ||\mathbf{d}(\mathbf{f}(\mathbf{x}, \mathbf{p})) - \mathbf{d}_{ref}||_2^2$$

Geometric transformation

Motion parameters

$$E(\mathbf{d}, \mathbf{p}) = \mathbf{d}(\mathbf{f}(\mathbf{x}, \mathbf{p}))^T \mathbf{A} \mathbf{d}(\mathbf{f}(\mathbf{x}, \mathbf{p})) + 2\mathbf{b}^T \mathbf{d}(\mathbf{f}(\mathbf{x}, \mathbf{p}))$$

$$\mathbf{A} = \mathbf{I}_m \text{ and } \mathbf{b} = -\mathbf{d}_{ref}$$

Identity matrix

Quadratic cost function

Learn **A** and **b**

$$E(\mathbf{d}, \mathbf{p}) = \mathbf{d}(\mathbf{f}(\mathbf{x}, \mathbf{p}))^T \mathbf{A} \mathbf{d}(\mathbf{f}(\mathbf{x}, \mathbf{p})) + 2\mathbf{b}^T \mathbf{d}(\mathbf{f}(\mathbf{x}, \mathbf{p}))$$

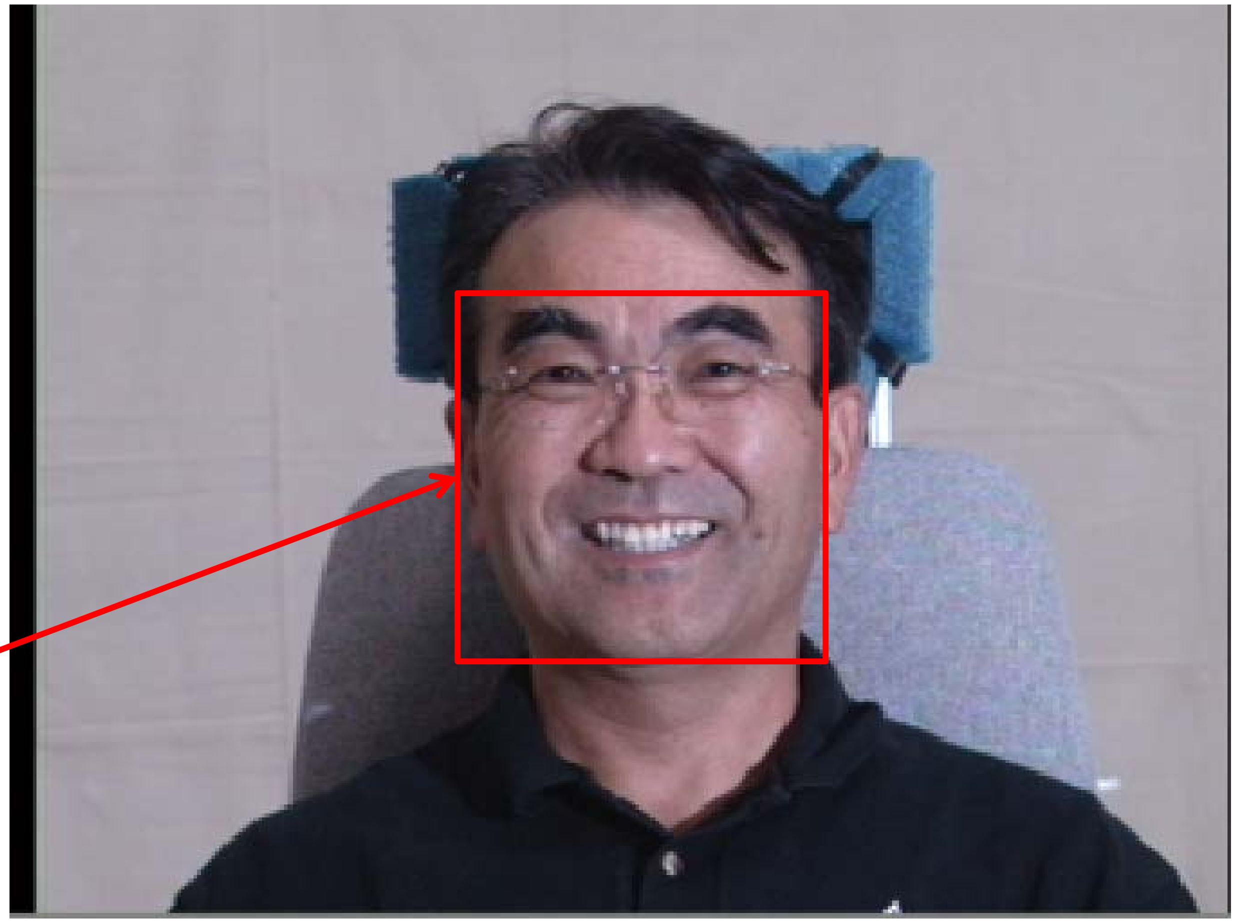
$\mathbf{A} = \mathbf{I}_m$ and $\mathbf{b} = -\mathbf{d}_{ref}$

Identity matrix

Training data

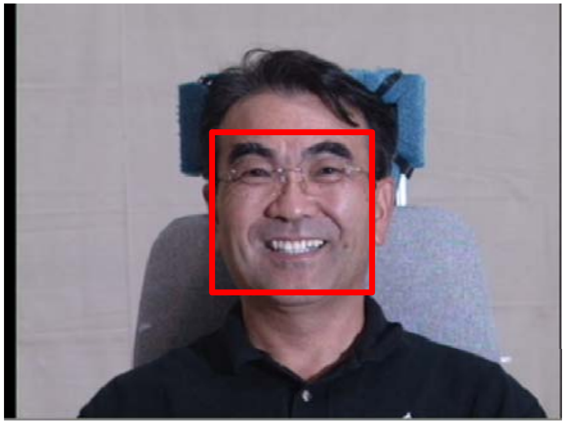
Template

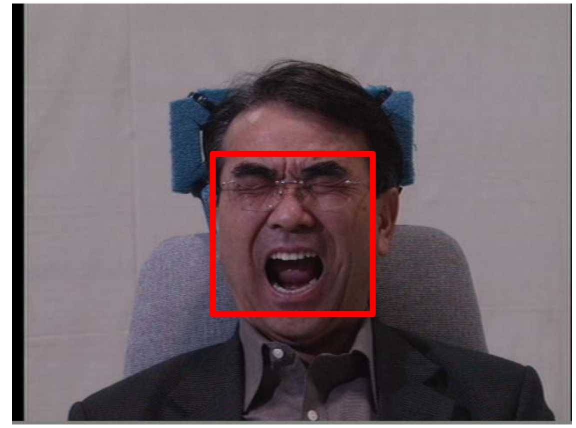
Ground truth

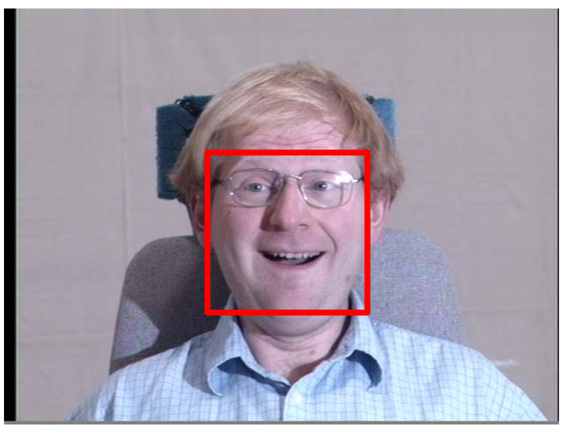


Image

Multiple training images







...

$$(\mathbf{d}_i, \mathbf{d}_i^{ref}, \mathbf{p}_i)$$

Image

Template

Ground truth

1st desired property of the cost function

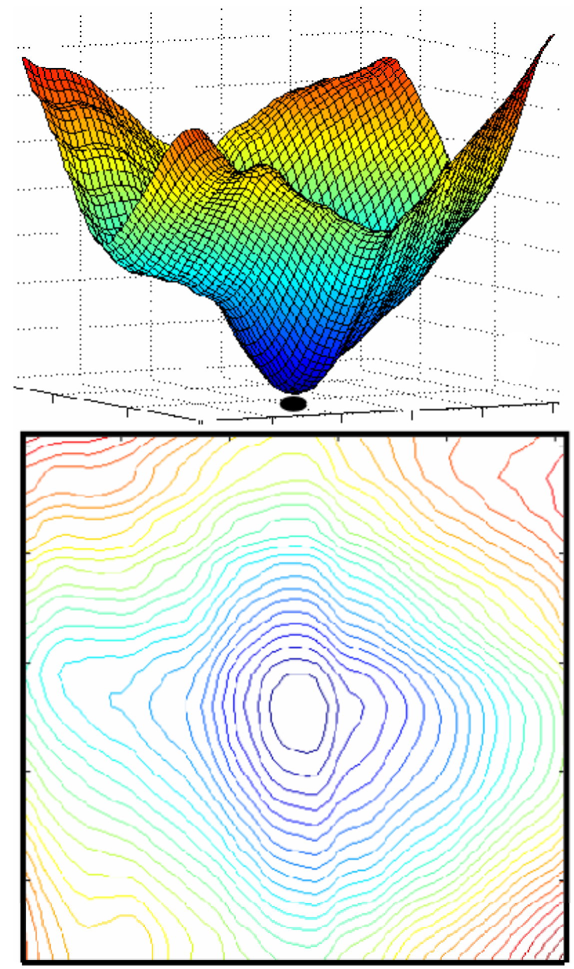
There is a local minimum at the right place!

Training image i^{th}

$$\left.\frac{\partial E(\mathbf{d}_i, \mathbf{p})}{\partial \mathbf{p}}\right|_{\mathbf{p}_i} = \vec{\mathbf{0}}$$

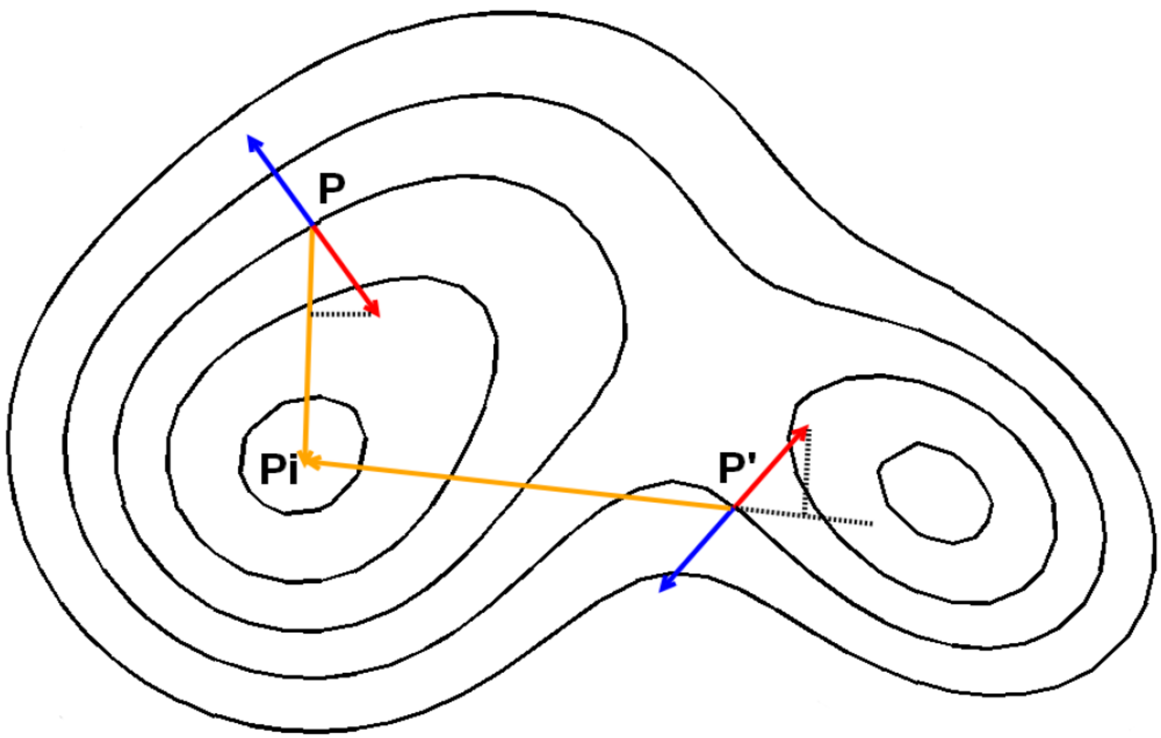
Correct alignment parameter

$$\left\| \left.\frac{\partial E(\mathbf{d}_i, \mathbf{p})}{\partial \mathbf{p}}\right|_{\mathbf{p}_i} \right\|_2^2 = 0$$



2^{nd} desired property of the cost function

Gradient descent agrees with the optimal walking direction



Pi: desired location

At P, gradient descent moves closer to Pi
At P', gradient descent moves away from Pi

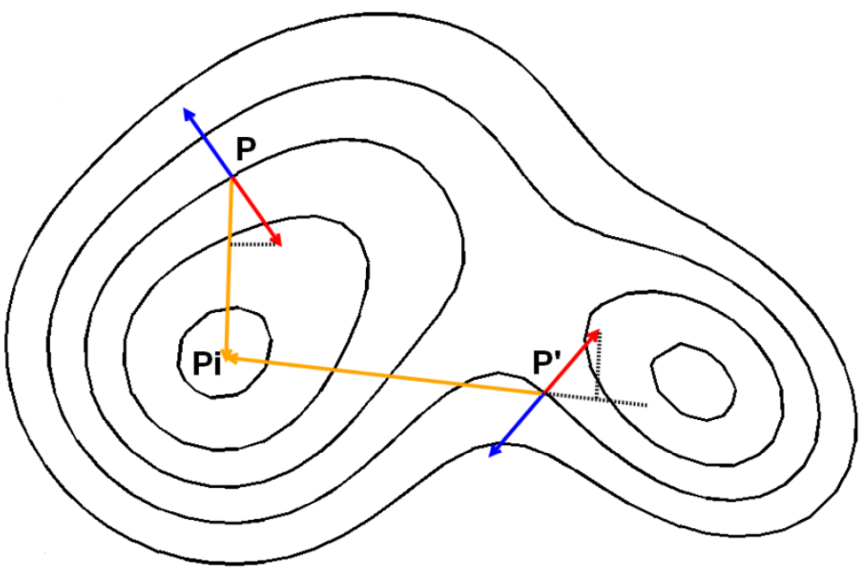
Enforcing the 2nd desired property:

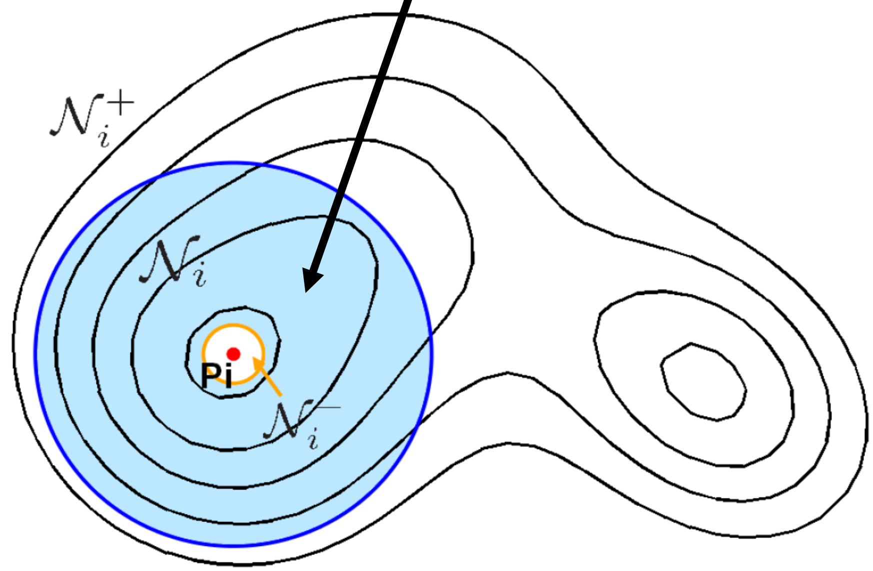
$$\left\langle -\left(\frac{\partial E(\mathbf{d}_i, \mathbf{p})}{\partial \mathbf{p}}\right)^T, \frac{\mathbf{p}_i - \mathbf{p}}{||\mathbf{p}_i - \mathbf{p}||_2} \right\rangle > 0 \;\; \forall \mathbf{p} \in \mathcal{N}_i$$

Gradient descent direction

Optimal direction

Appropriate neighbourhood





The learning problem

$$\left\| \frac{\partial E(\mathbf{d}_i, \mathbf{p})}{\partial \mathbf{p}} \Big|_{\mathbf{p}_i} \right\|_2^2 = 0 \quad \forall i$$

1st desired property

$$\left\langle -\left(\frac{\partial E(\mathbf{d}_i, \mathbf{p})}{\partial \mathbf{p}} \right)^T, \frac{\mathbf{p}_i - \mathbf{p}}{||\mathbf{p}_i - \mathbf{p}||_2} \right\rangle > 0 \ \forall \mathbf{p} \in \mathcal{N}_i \ \ \forall i$$

2nd desired property

$$\underset{\mathbf{A},\mathbf{b},\xi}{\text{minimize}} \sum_i \frac{1}{2} \left\| \frac{\partial E(\mathbf{d}_i, \mathbf{p})}{\partial \mathbf{p}} \Big|_{\mathbf{p}_i} \right\|_2^2 + C \sum_i \xi_i$$

$$\text{s.t.} \quad \left\langle -\left(\frac{\partial E(\mathbf{d}_i, \mathbf{p})}{\partial \mathbf{p}} \right)^T, \frac{\mathbf{p}_i - \mathbf{p}}{||\mathbf{p}_i - \mathbf{p}||_2} \right\rangle \geq M - \xi_i \ \forall \mathbf{p} \in \mathcal{N}_i, \forall i$$

$$\xi_i \geq 0 \quad \forall i$$

The learning problem

- **Convex optimization problem**
- **Infinite number of constraints:**
  - Subset of most violated constraints
  - Iteratively update

$$\underset{\mathbf{A},\mathbf{b},\xi}{\text{minimize}} \sum_i \frac{1}{2} \left\| \left. \frac{\partial E(\mathbf{d}_i, \mathbf{p})}{\partial \mathbf{p}} \right|_{\mathbf{p}_i} \right\|_2^2 + C \sum_i \xi_i$$

$$\text{s.t.} \quad \left\langle -\left(\frac{\partial E(\mathbf{d}_i, \mathbf{p})}{\partial \mathbf{p}} \right)^T, \frac{\mathbf{p}_i - \mathbf{p}}{\|\mathbf{p}_i - \mathbf{p}\|_2} \right\rangle \geq M - \xi_i \; \forall \mathbf{p} \in \mathcal{N}_i, \forall i$$

$$\xi_i \geq 0 \quad \forall i$$

Weighted template alignment

$$\underset{\mathbf{p}}{\text{minimize}} \, ||\mathbf{d}(\mathbf{f}(\mathbf{x}, \mathbf{p})) - \mathbf{d}_{ref}||_2^2$$

$$(\mathbf{d}(\mathbf{f}(\mathbf{x}, \mathbf{p})) - \mathbf{d}_{ref})^T diag(\mathbf{w}) (\mathbf{d}(\mathbf{f}(\mathbf{x}, \mathbf{p})) - \mathbf{d}_{ref})$$

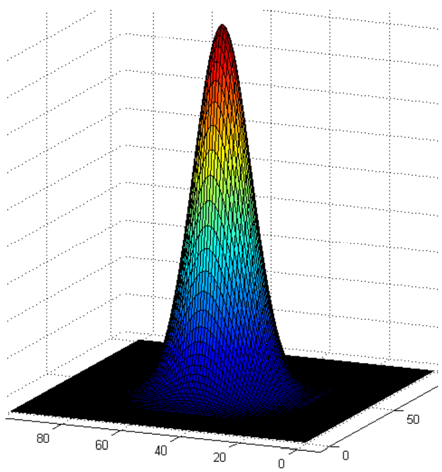
$$E(\mathrm{d}, \mathrm{p}) = \mathrm{d}(\mathbf{f}(\mathrm{x}, \mathrm{p}))^T \mathbf{A} \mathrm{d}(\mathbf{f}(\mathrm{x}, \mathrm{p})) + 2\mathrm{b}^T \mathbf{d}(\mathbf{f}(\mathbf{x}, \mathbf{p}))$$

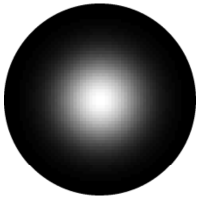
$$\mathbf{A} = diag(\mathbf{w})$$

$$\mathbf{b} = -diag(\mathbf{w})\mathbf{d}_{ref}$$

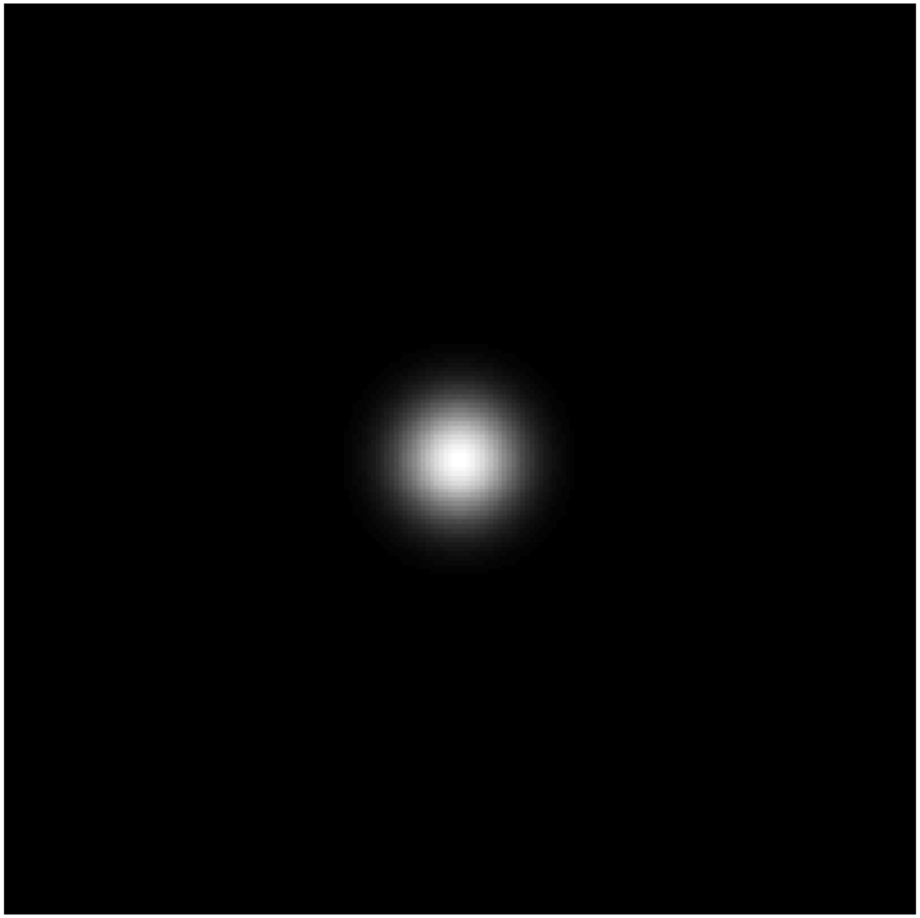
$$0 \leq w_i \leq 1$$

Experiment with the Gaussian



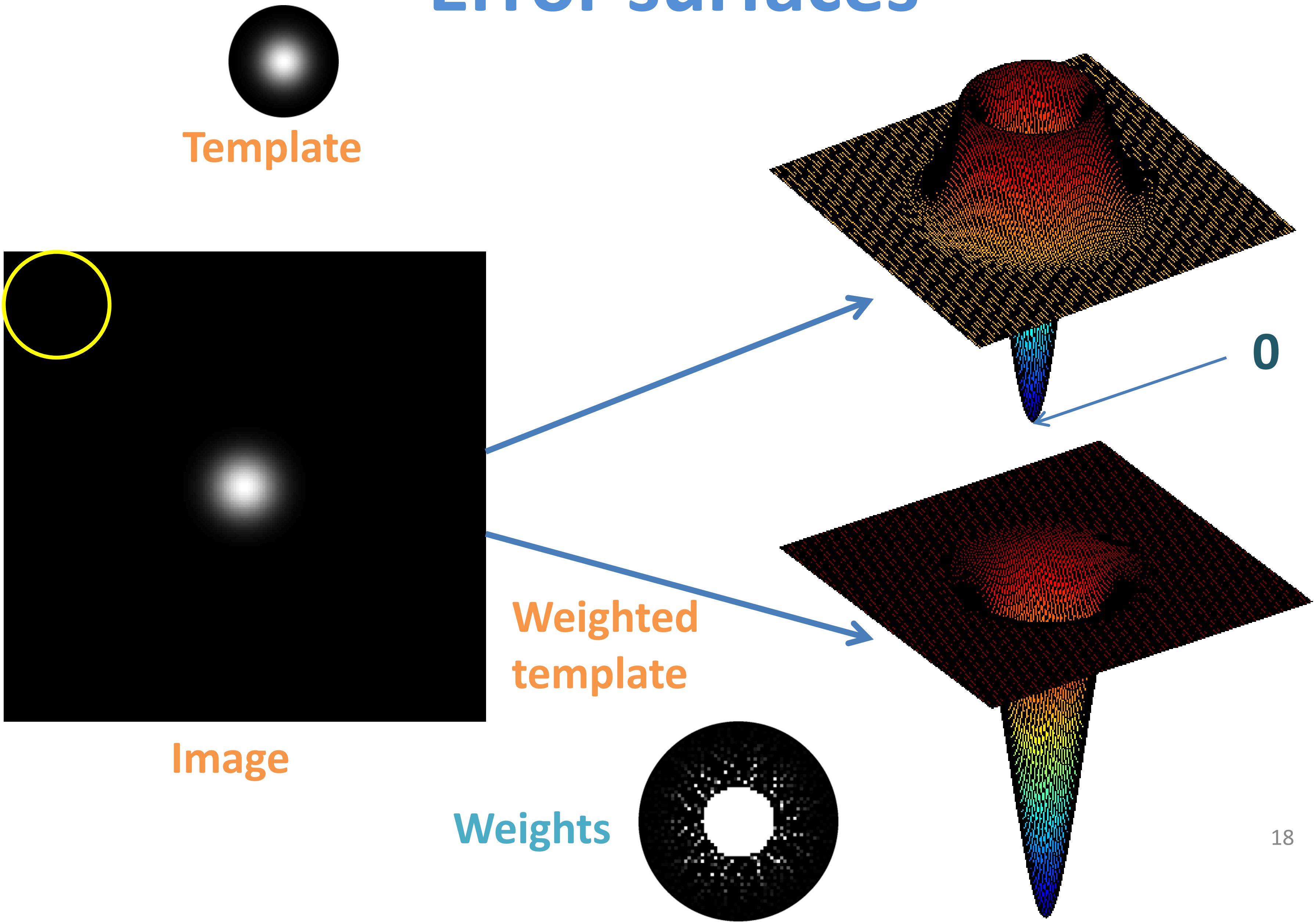


Template

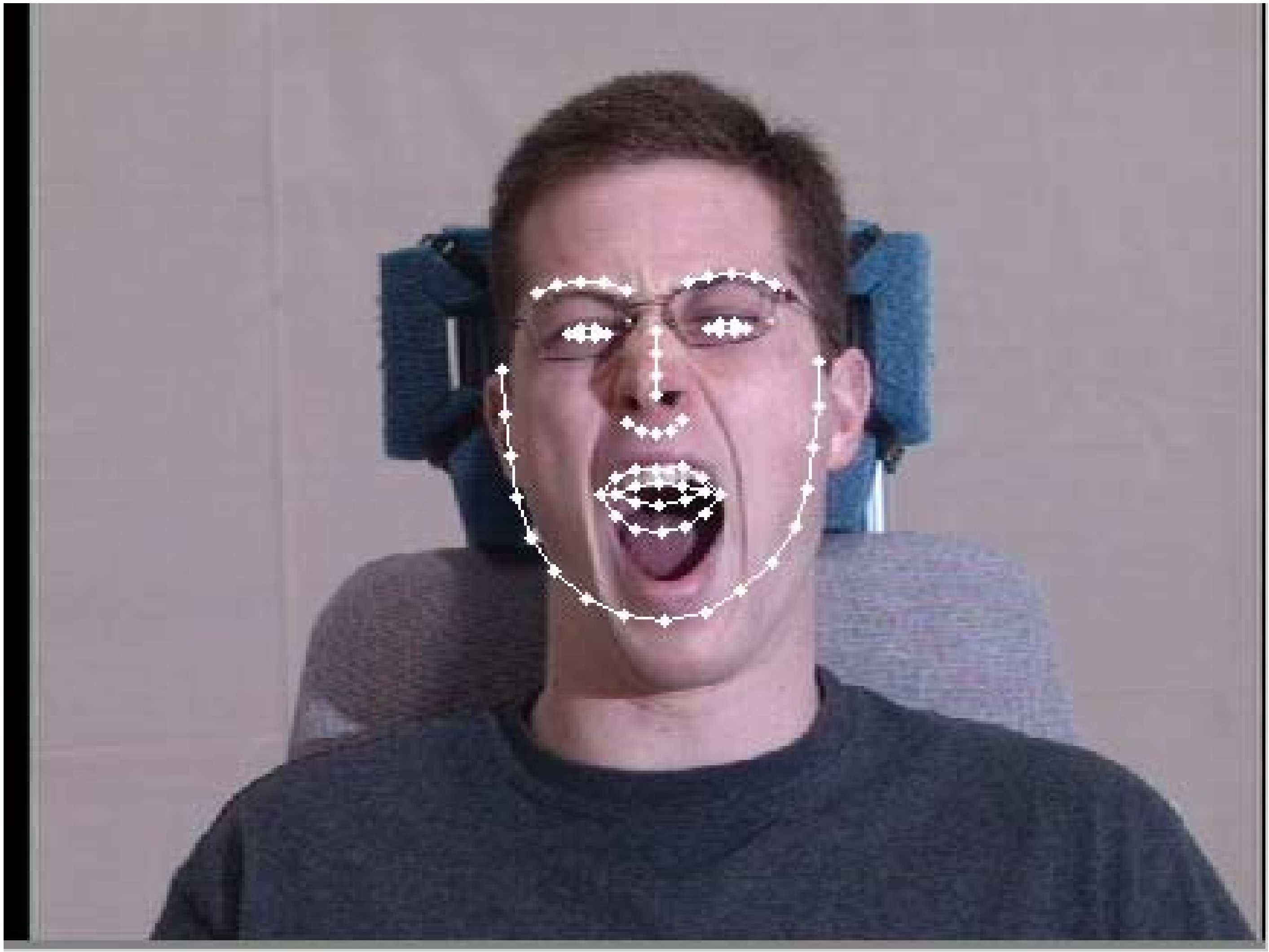


Image

Error surfaces



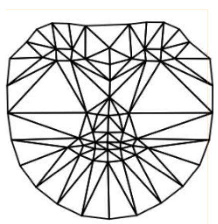
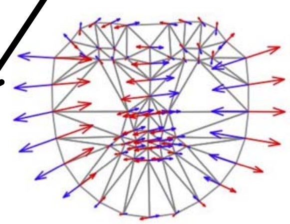
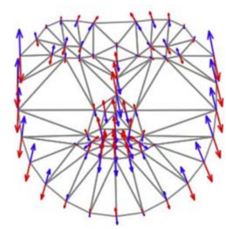
Active Appearance Models (AAMs)



Energy function in AAMs

$$\underset{\mathbf{p},\mathbf{c}}{\text{minimize}} \, ||\mathbf{d}(\mathbf{f}(\mathbf{x},\mathbf{p})) - \mathbf{U}\mathbf{c}||_2^2$$

Affine parameters + Shape coefficients

Shape =  + s_1  + s_2  + ...

Mean shape

PCA basis
for shape variation

U_0 U_1 U_2

Appearance Variation

(Cootes et al 98, Mathews & Baker 04, Gong et al 00)

Learning AAM energy function

$$\underset{\mathbf{p},\mathbf{c}}{\text{minimize}} \, ||\mathbf{d}(\mathbf{f}(\mathbf{x},\mathbf{p})) - \mathbf{U}\mathbf{c}||_2^2$$

$$E(\mathbf{d},\mathbf{p}) = \mathbf{d}(\mathbf{f}(\mathbf{x},\mathbf{p}))^T \mathbf{A}\mathbf{d}(\mathbf{f}(\mathbf{x},\mathbf{p})) + 2\mathbf{b}^T \mathbf{d}(\mathbf{f}(\mathbf{x},\mathbf{p}))$$

$$\mathbf{A} = \mathbf{I}_m - \mathbf{U}\mathbf{U}^T = \mathbf{I}_m - \sum_{i=1}^{k} \mathbf{u}_i \mathbf{u}_i^T$$

$$\mathbf{b} = \mathbf{0}_m$$

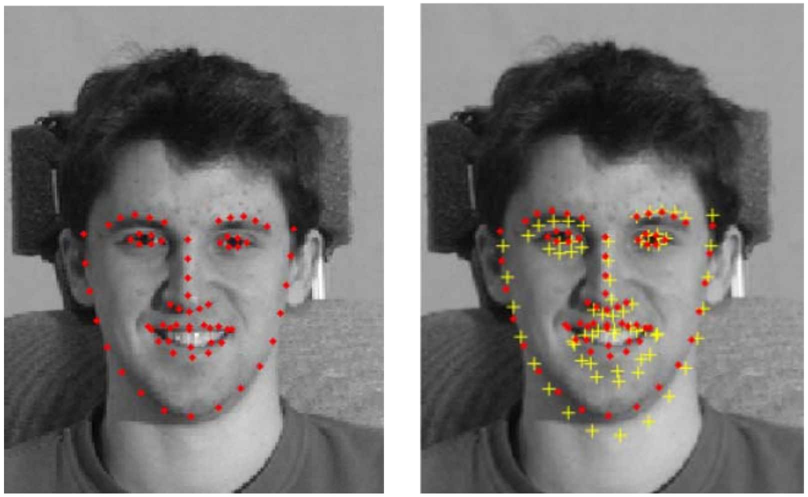
Weighted Basis AAM alignment:

$$\mathbf{A} = \mathbf{I}_m - \sum_{i=1}^{K} \lambda_i \mathbf{u}_i \mathbf{u}_i^T \qquad 0 \leq \lambda_i \leq 1$$

$\mathbf{b}$ **No constraint**

Experiments with Multi-PIE database

- Data:
  - 337 subjects
  - Frontal, directly illuminated
  - Expressions: smile, disgust, squint, surprise, scream
  - 68 hand labeled landmarks
  - Training, validation, test sets: 400/200/500
- Motion:
  - Affine + non-rigid
- Testing:
  - Random perturbation



Results of weighted PCA basis

Alignment problem becomes harder →

Perturbation amount	0.75	1.00	1.25	1.5
Initial	1.50 ± 0.50	2.16 ± 0.76	2.74 ± 1.04	3.08 ± 1.08
PCA 90%	**0.72 ± 0.40**	0.86 ± 0.66	0.94 ± 0.72	1.20 ± 1.30
PCA 80%	0.80 ± 0.46	0.86 ± 0.68	0.98 ± 0.74	1.14 ± 1.00
PCA 70%	0.82 ± 0.40	0.86 ± 0.50	0.94 ± 0.60	1.10 ± 0.92
Ours	0.74 ± 0.38	**0.80 ± 0.50**	**0.86 ± 0.58**	**0.96 ± 0.78**

Mean error + std, the smaller the better!

Summary

- Learning an optimal metric (A, b of the cost function) for image alignment
  - Template alignment
  - Active Appearance Models (AAMs)
- Local minima at and only at the right places
- Convex quadratic formulation

Less on search strategies
More on what we are searching for