Supervised Descent Method and its Applications to Face Alignment

Xuehan Xiong and Fernando De la Torre, The Robotics Institute, Carnegie Mellon University

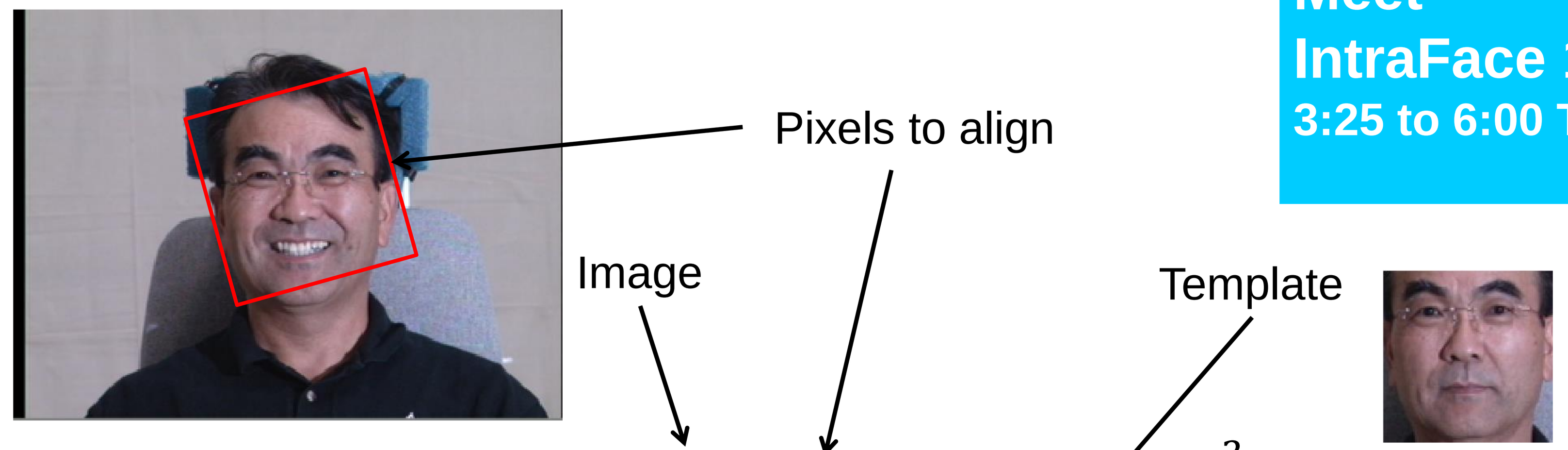
CVPR 2013 Portland, Oregon June 23-28

MOTIVATION

Meet IntraFace 1.0
3:25 to 6:00 Today

➢ Generative Methods

- Template Matching (e.g., Lucas-Kanade 81)



Pixels to align

Image

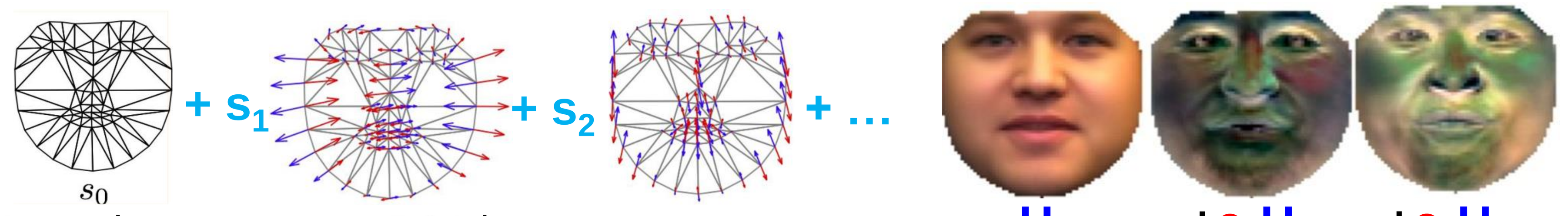
Template

$$f_{TM}(\mathbf{p}) = \|\mathbf{d}(\mathbf{f}(\mathbf{x},\mathbf{p})) - \mathbf{d}_{template}\|^2$$

Geometric transformation

Motion parameters

- Active Appearance Models (e.g., Cootes et al 98, Mathews & Baker 04, De la Torre et al. 02-08)



Mean shape PCA bases

Affine parameters + Shape coefficients

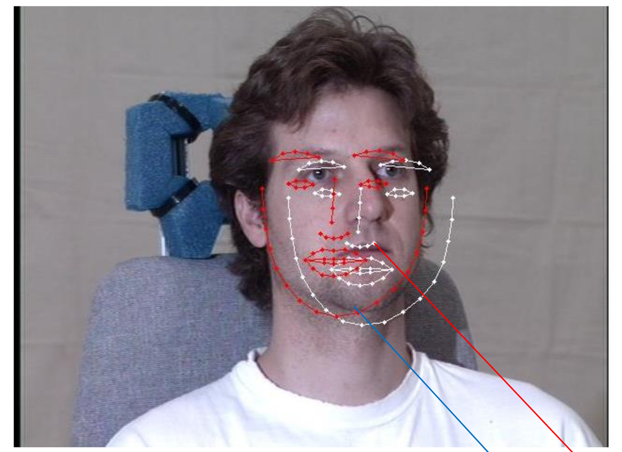
Appearance variation

$$f_{AAM}(\mathbf{p},\mathbf{c}) = \|\mathbf{d}(\mathbf{f}(\mathbf{x},\mathbf{p})) - \mathbf{U}\mathbf{c}\|^2$$

- Problems: prone to local minima

➢ Discriminative Methods

(e.g., Saragih & Goecke '09, Liu '09, Sauer et al. '11, Saragih '11, Cao et al. '12, Ribera & Martinez '12))



1. What are discriminative methods minimizing?
2. Is there any relation between discriminative and traditional image alignment methods?
3. How to learn discriminative methods effectively?

$\Delta\mathbf{p}_1 \longleftarrow \text{features}(\mathbf{d}(\mathbf{f}(\mathbf{x},\mathbf{p}_0 + \Delta\mathbf{p}_1)))$

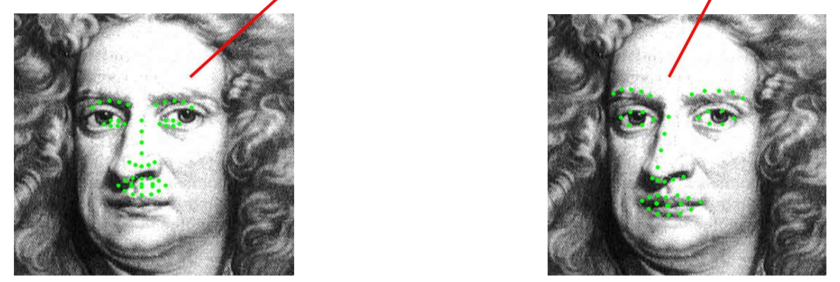
$\Delta\mathbf{p}_2 \longleftarrow \text{features}(\mathbf{d}(\mathbf{f}(\mathbf{x},\mathbf{p}_0 + \Delta\mathbf{p}_2)))$

Perturbation of motion parameters

SUPERVISED DESCENT METHOD

➢ Objective Function

$$f_{SDM}(\Delta\mathbf{x}) = \|\mathbf{h}(\mathbf{d}(\mathbf{x}_0 + \Delta\mathbf{x})) - \mathbf{h}(\mathbf{d}(\mathbf{x}_*))\|^2$$



- SDM vs. TM vs. AAM

$$f_{SDM}(\Delta\mathbf{x}) = \|\mathbf{h}(\mathbf{d}(\mathbf{x}_0 + \Delta\mathbf{x})) - \boldsymbol{\phi}_*\|^2$$
$$f_{TM}(\mathbf{p}) = \|\mathbf{d}(\mathbf{f}(\mathbf{x},\mathbf{p})) - \mathbf{d}_{template}\|^2$$
$$f_{AAM}(\mathbf{p},\mathbf{c}) = \|\mathbf{d}(\mathbf{f}(\mathbf{x},\mathbf{p})) - \mathbf{U}\mathbf{c}\|^2$$

- No model of shape or appearance
  1. asymmetric facial gestures
  2. no coupling between landmarks (e.g., brows and mouth)
- Richer image descriptors (e.g., SIFT)

➢ How to Optimize

Newton update

$$\mathbf{x}_k = \mathbf{x}_{k-1} - \mathbf{H}^{-1}\mathbf{J}_f$$
$$= \mathbf{x}_{k-1} + 2\mathbf{H}^{-1}\mathbf{J}_h^T(\boldsymbol{\phi}_* - \boldsymbol{\phi}_{k-1})$$

$\boldsymbol{\phi}_* = \mathbf{h}(\mathbf{d}(\mathbf{x}_*))$ $\quad \boldsymbol{\phi}_{k-1} = \mathbf{h}(\mathbf{d}(\mathbf{x}_{k-1}))$

Problems:

- Hessian is expensive to invert
- Hessian is not positive definite in all domain
- $\mathbf{h}$ is a non-differentiable image operator
- In testing we do not know $\boldsymbol{\phi}_*$

SDM update

$$\mathbf{x}_k = \mathbf{x}_{k-1} + \mathbf{R}_{k-1}\boldsymbol{\phi}_{k-1} + \mathbf{b}_{k-1}$$

Generic descent direction

Solutions:

- $2\mathbf{H}^{-1}\mathbf{J}_h^T$ is learnt from data
- $\boldsymbol{\phi}_*$ is learnt through a bias term $\mathbf{b}_{k-1}$

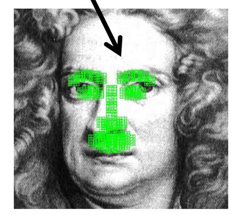
➢ Learning SDM

SDM learns a sequence of $\{\mathbf{R}_k, \mathbf{b}_k\}$ such that the succession of $\mathbf{x}_k$ will converge to $\mathbf{x}_*$ for all images in the training data

For example, learning $\mathbf{R}_0, \mathbf{b}_0$

$$\arg\min_{\mathbf{R}_0,\mathbf{b}_0} \sum_{\mathbf{d}^i}\sum_{\mathbf{x}_0^i} \|\Delta\mathbf{x}_*^i - \mathbf{R}_0\boldsymbol{\phi}_0^i - \mathbf{b}_0\|^2$$

$$\Delta\mathbf{x}_*^i = \mathbf{x}_*^i - \mathbf{x}_0^i$$



RESULTS

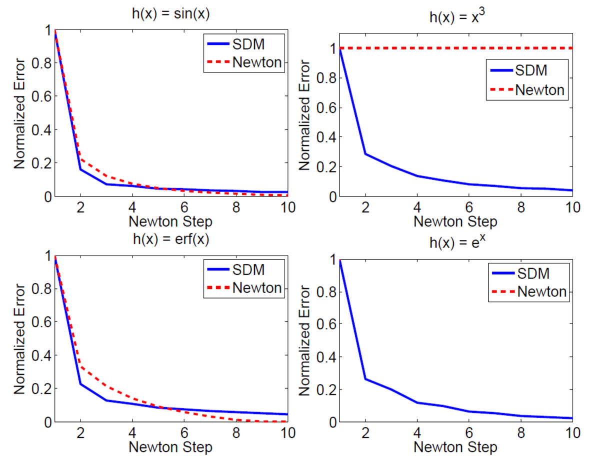
➢ Analytic Functions

Problem:

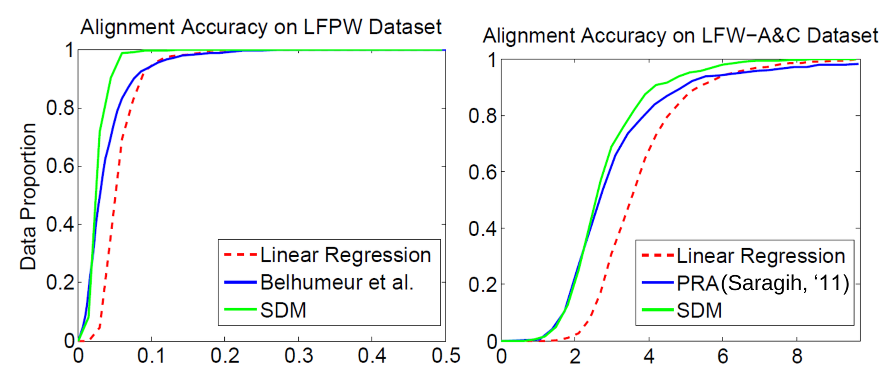
$$\min_x f(x) = (h(x) - y^*)^2$$

Experimental setup:

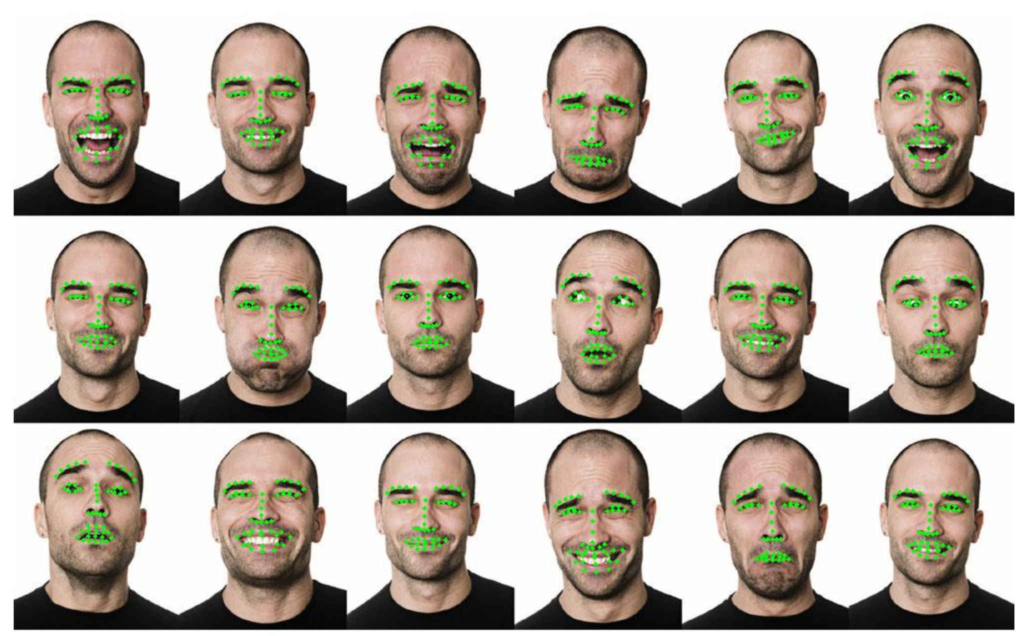
Function $h(x)$	Training Set $\mathbf{y}$	Training Set $\mathbf{x} = h^{-1}(\mathbf{y})$	Test Set $\mathbf{y}^*$
$\sin(x)$	[-1:0.2:1]	$\arcsin(\mathbf{y})$	[-1:0.05:1]
x^3	[-27:3:27]	$\mathbf{y}^{\frac{1}{3}}$	[-27:0.5:27]
$\text{erf}(x)$	[-0.99:0.11:0.99]	$\text{erf}^{-1}(\mathbf{y})$	[-0.99:0.03:0.99]
e^x	[1:3:28]	$\log(\mathbf{y})$	[1:0.5:28]



➢ Comparison with Previous Methods

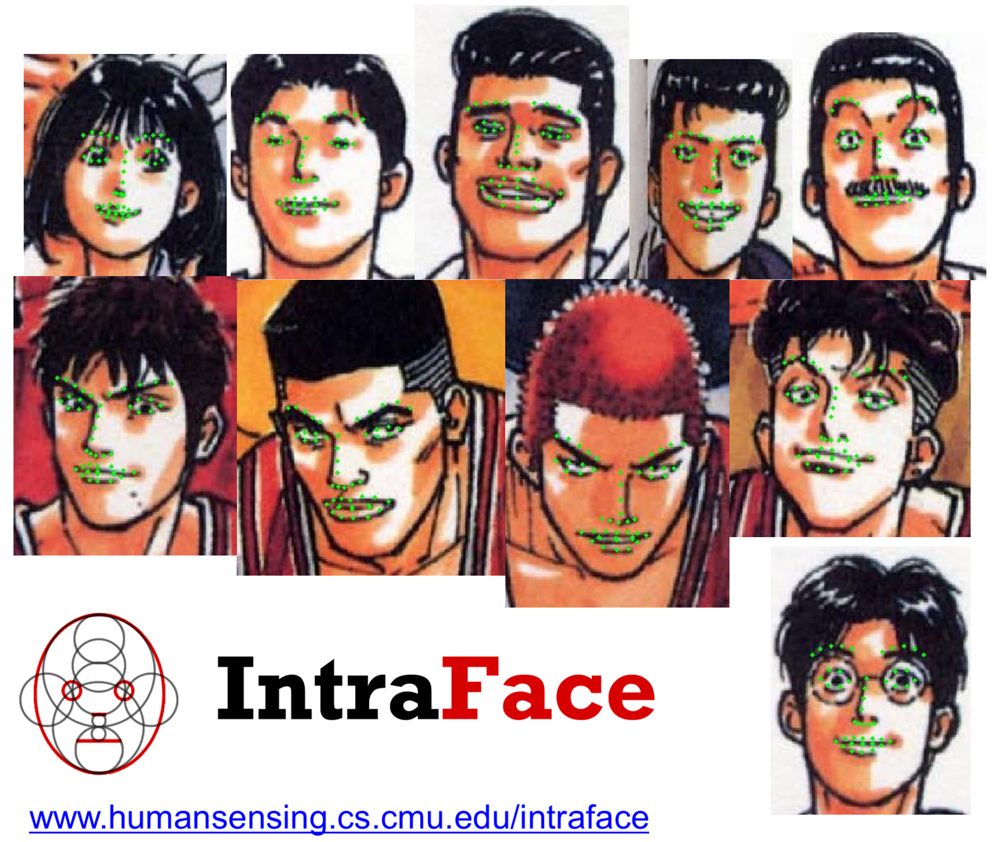


➢ Expression



➢ "faces in the wild"

➢ Cartoon and Drawings



IntraFace

www.humansensing.cs.cmu.edu/intraface

NSF