IEEE Conference in Computer Vision and Pattern Recognition 2014

· U C ·
UNIVERSIDADE DE COIMBRA
UNIVERSITY OF COIMBRA

Non-Parametric Bayesian Constrained Local Models

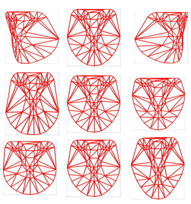
Pedro Martins, Rui Caseiro, Jorge Batista
Institute of Systems and Robotics - University of Coimbra - Portugal

Overview:

- **Goal:** Face alignment in unseen images.
- Constrained Local Models (CLM): combine an ensemble of local detectors with a global optimization strategy that constrains the feature points to lie in the subspace spanned by a linear shape model (Point Distribution Model - PDM).
- CLM two step fitting approach:
  - (1) Local search using the detectors (likelihood map for each landmark).
  - (2) Global optimization strategy that estimates the PDM parameters that jointly maximize all the detections.
- Non-Parametric Bayesian global optimization strategy that models the posterior distribution by a Kernel Density Estimator (KDE).

CLM: Shape Model (PDM) and Local Detectors

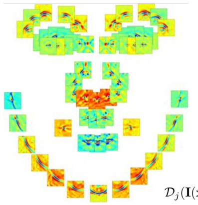
$\mathbf{s} = (x_1, y_1, \ldots, x_v, y_v)^T$



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

$\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$

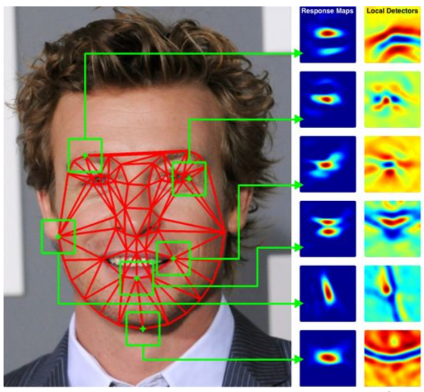
Local Landmark Detectors



$\mathcal{D}_j(\mathbf{I}(\mathbf{x}_j)) = \mathbf{h}_j^T \mathbf{I}(\mathbf{x}_j)$

$\mathbf{s} = \mathcal{S}(\mathbf{s}_0 + \Phi\mathbf{b}, \mathbf{q})$ → (**b**) shape and (**q**) pose parameters

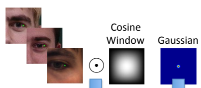
Local Detectors (MOSSE Filters)



$p(a_j = 1|\mathcal{D}_j, \mathbf{I}(\mathbf{z}_j))$ $\mathcal{F}^{-1}\{\mathbf{H}_j^*\}$

Correlation in Fourier Domain

$$\mathbf{G} = \mathcal{F}\{\mathbf{I}\} \odot \mathbf{H}^*$$



$$\min_{\mathbf{H}^*} \sum_{i=1}^{N} (\mathcal{F}\{\mathbf{I}_i\} \odot \mathbf{H}^* - \mathbf{G}_i)^2$$

MOSSE Filter

$$\mathbf{H}^* = \frac{\sum_{i=1}^{N} \mathbf{G}_i \odot \mathcal{F}\{\mathbf{I}_i\}^*}{\sum_{i=1}^{N} \mathcal{F}\{\mathbf{I}_i\} \odot \mathcal{F}\{\mathbf{I}_i\}^* + \epsilon}$$

Prob. Landmark (j) being aligned

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

The Alignment Goal

Given a shape observation (**y**), find the optimal set of shape (**b**) and pose parameters that maximize the posterior probability

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

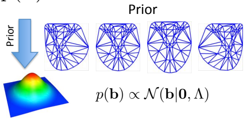
Posterior Expectation

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

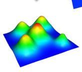
Posterior

Likelihood

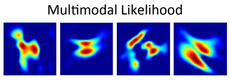
Prior



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



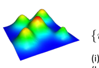
Multimodal Likelihood



Posterior Distribution (KDE)

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)

Whitening

$\mathbf{b}_k^{(i)} \rightarrow \mathbf{A}^{-1}\mathbf{b}_k^{(i)}$

$\mathbf{S} = \mathbf{A}\mathbf{A}^T$

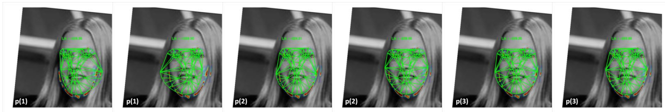
→ Rescaled Regularization

$\frac{\det(\mathbf{A})^{-1}}{h^n} K\left(\frac{\mathbf{A}^{-1}\mathbf{b}_k}{h}\right)$

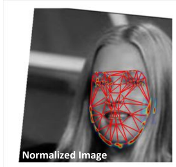
→ Bandwidth Gaussian Kernel

$h_{\mathrm{opt}} = \left(\frac{4}{2N(n+2)}\right)^{\frac{1}{n+4}}$

Non-Parametric Global Optimization



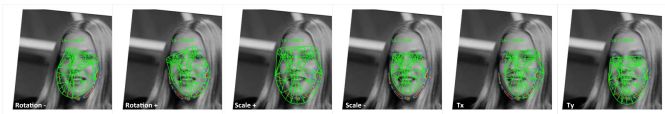
Shape Parameters



Alignment Quality: 'Robust' Product of All Landmarks

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

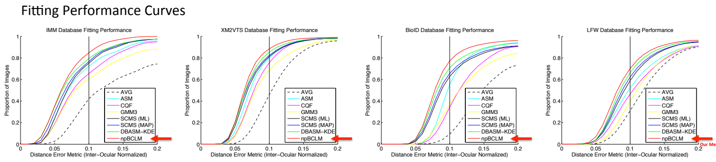
$$\mathbf{b}_k^{(i)} \sim p(\mathbf{b}_k|\mathbf{b}_{k-1}^{(i)}) \propto \mathcal{N}(\mathbf{b}_k|\mathbf{b}_{k-1}, \Sigma_{\mathbf{b}})$$



Pose Parameters

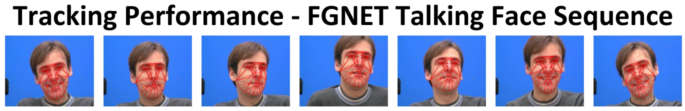
Fitting Performance - Labeled Faces in the Wild (LFW)

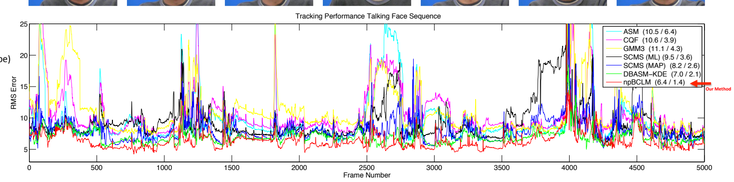
Fitting Performance Curves



$$e_m(\mathbf{s}) = \frac{1}{v\, d_{eyes}} \sum_{i=1}^{v} \|\mathbf{s}_i - \mathbf{s}_i^{gt}\|$$

Tracking Performance - FGNET Talking Face Sequence





INSTITUTE OF SYSTEMS AND ROBOTICS
UNIVERSITY OF COIMBRA
· U C ·
FCTUC FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE DE COIMBRA

This work was supported by the Portuguese Science Foundation (Fundação para a Ciência e Tecnologia - FCT), under the research project "Differential Geometry for Computer Vision and Pattern Recognition – DG2CVPR"(grant PTDC/EIA-CRO/122812/2010). Pedro Martins and Rui Caseiro also acknowledge the FCT through the grants SFRH/BDP/90200/2012 and SFRH/ BD74152/2010, respectively.