Let the Shape Speak - Discriminative Face Alignment using Conjugate Priors

Pedro Martins, Rui Caseiro, João F. Henriques, Jorge Batista
http://www.isr.uc.pt/~pedromartins

Institute of Systems and Robotics
University of Coimbra
Portugal

British Machine Vision Conference 2012
Introduction

• **Goal:** Face alignment in unseen images.

• Closely related to Constrained Local Models (CLM) and Active Shape Models (ASM), where a set of local detectors is constrained to lie in the subspace spanned by a Point Distribution Model (PDM).

  • Two step fitting approach:
    • (1) Local search using the local detectors (response maps for each landmark)
    • (2) Global optimization strategy that finds the PDM parameters that jointly maximize all the detections.

• **Proposed Work:** New Bayesian global optimization strategy where the prior distribution encodes the transition of the PDM parameters.

  • The prior distribution is modeled using recursive Bayesian estimation. The mean and covariance are assumed to be unknown and treated as random variables.
Related Work - Parametric Image Alignment

- **Generative / Holistic methods**
  - Active Appearance Models (AAM)
    T.F.Cootes, G.J.Edwards, C.J.Taylor - ECCV 98
  - 3D Morphable Models (3DMM)
    V.Blanz, T.Vetter - SIGGRAPH 99
  - Real Time Combined 2D+3D Active Appearance Models
    J.Xing, S.Baker, I.Matthews, T.Kanade - CVPR 2004

- **Discriminative / Patch-Based**
  - Active Shape Models (ASM)
    T.F.Cootes, G.J.Edwards, C.J.Taylor - CVIU 95
  - Constrained Local Model (CLM)
    D.Cristinance, T.F.Cootes - BMVC 2006
  - Convex Quadratic Fitting (CQF)
    Y.Wang, S.Lucey, J.Cohn - CVPR 2008
  - Bayesian Constrained Local Model (BCLM)
    U.Paquet - CVPR 2009
  - Subspace Constrained Mean-Shifts (SCMS)
    J.Saragih, S.Lucey, J.Cohn - ICCV 2009

\[ s = S (s_0 + \Phi b; q) \]

- **Point Distribution Model**
- **Shape Parameters**
- **Pose Parameters**
The Alignment Goal

• Given a shape observation vector \( \mathbf{y} \), find the optimal set of shape (and pose) parameters \( \mathbf{b} \) that maximize the posterior probability

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

• Assuming:
  • Conditional independence between landmarks
  • Close to a solution

\[
p(\mathbf{b}|\mathbf{y}) \propto \left( \prod_{i=1}^{v} p(\mathbf{y}_i|\mathbf{b}) \right) p(\mathbf{b}|\mathbf{b}_{k-1}^*)
\]

Likelihood from the local detectors

Prior on how parameters change
The Likelihood Term

• Convex energy function:

\[
p(y|b) \propto \exp \left( -\frac{1}{2} \left( y - (s_0 + \Phi b) \right)^T \Sigma_y^{-1} \left( y - (s_0 + \Phi b) \right) \right)
\]

Observed shape \( (y) \)

Difference between the observed and the mean shape

Uncertainty covariance

The likelihood follow a Gaussian distribution

\[
p(y|b) \propto \mathcal{N}(\Delta y|\Phi b, \Sigma_y)
\]
Local Landmark Detectors

Linear SVM

(+) Aligned Examples

(-) Misaligned Examples

\[ D_i^{\text{linear}}(I(y_i)) = w_i^T I(y_i) + b_i \]

\[ i = 1, \ldots, v \text{ landmarks} \]
Local Landmark Detectors - MOSSE Filters

- Correlation in Fourier Domain
  
  \[ G = \mathcal{F}\{I\} \odot H^* \]

- MOSSE Filter
  
  \[
  H^* = \frac{\sum_{j=1}^{N} G_j \odot \mathcal{F}\{I_j\}^*}{\sum_{j=1}^{N} \mathcal{F}\{I_j\} \odot \mathcal{F}\{I_j\}^*}
  \]

- Visual object tracking using adaptive correlation filters
  D.Bolme, J.Beveridge, B.Draper, Y.Lui, CVPR 2010
Local Optimization Strategies

Weighted Peak Response (WPR)

\[
p_i(z_i) \quad \text{Prob. } z_i \text{ is aligned}
\]

\[
y_i^{WPR} = \max_{z_i \in \Omega_{y_i}} (p_i(z_i))
\]

\[
\Sigma_{y_i} = \text{diag}(p_i(y_i^{WPR})^{-1})
\]

Gaussian Response (GR)

\[
y_i^{GR} = \frac{1}{d} \sum_{z_i \in \Omega_{y_i}} p_i(z_i)z_i
\]

\[
d = \sum_{z_i \in \Omega_{y_i}} p_i(z_i)
\]

\[
\Sigma_{y_i}^{GR} = \frac{1}{d-1} \sum_{z_i \in \Omega_{y_i}} p_i(z_i)(z_i - y_i^{GR})(z_i - y_i^{GR})^T
\]

Kernel Density Estimator (KDE)

\[
y_i^{KDE(\tau+1)} \leftarrow \frac{\sum_{z_i \in \Omega_{y_i}} p_i(z_i) \mathcal{N}(y_i^{KDE(\tau)}|z_i, \sigma_h^2 I_2)}{\sum_{z_i \in \Omega_{y_i}} p_i(z_i) \mathcal{N}(y_i^{KDE(\tau)}|z_i, \sigma_h^2 I_2)}
\]

\[
\Sigma_{y_i}^{KDE} = \frac{1}{d-1} \sum_{z_i \in \Omega_{y_i}} p_i(z_i)(z_i - y_i^{KDE})(z_i - y_i^{KDE})^T
\]
KDE Demo Video
MAP Global Alignment

**Likelihood**

\[ p(y|b) \propto \mathcal{N}(\Delta y|\Phi b, \Sigma_y) \]

Mean that is function of \( b \)
Covariance independent of \( b \)

**Prior**

\[ p(b_k|b_{k-1}) \propto \mathcal{N}(b_k|\mu_b, \Sigma_b) \]

\[ p(b_k|y_k, \ldots, y_0) \propto \mathcal{N}(b_k|\mu_k, \Sigma_k) \]

\[
\Sigma_k = \left( (\Sigma_{b_k} + \Sigma_{k-1})^{-1} + \Phi^T \sum_{m=1}^{M} \left( \Sigma_{y(m)}^{-1} \right) \Phi \right)^{-1}
\]

\[
\mu_k = \Sigma_k \left( \Phi^T \sum_{m=1}^{M} \left( \Sigma_{y(m)}^{-1} \Delta y_{(m)} \right) + (\Sigma_{b_k} + \Sigma_{k-1})^{-1} \mu_{b_k} \right)
\]

**Posterior**

Bayes’s Theorem for Gaussian variables

\[ p(b_k|y) \propto \mathcal{N}(b_k|\mu, \Sigma) \]

\[ \Sigma = (\Sigma_b^{-1} + \Phi^T \Sigma_y^{-1} \Phi)^{-1} \]

\[ \mu = \Sigma(\Phi^T \Sigma_y^{-1} y + \Sigma_b^{-1} \mu_b) \]

Model the Covariance of \( b \)
(2\text{nd Order Estimate})

+ Bayesian Fusion of Multiple Detectors

\[ \Delta y_{(m)}, \Sigma y_{(m)} \]

Multiple Likelihood Observations
The Prior Term

- Mean and Covariance \((\mu_b, \Sigma_b)\) are assumed to be unknown and modeled as random variables.

\[
p(b_k | b_{k-1}) \propto \mathcal{N}(b_k | \mu_b, \Sigma_b)
\]

## Bayes Theorem:

\[
p(\mu_b, \Sigma_b | b) \propto p(b | \mu_b, \Sigma_b) \cdot p(\mu_b, \Sigma_b)
\]

Joint Posterior

Normal Inverse-Wishart

Joint Prior

Normal Inverse-Wishart

**Conjugate Prior** for a Gaussian with unknown mean and covariance is a Normal Inverse-Wishart distribution

- **Parameters**

  - Degrees of freedom
    \[ u_k = u_{k-1} + m \]
  - Number of measurements
    \[ \kappa_k = \kappa_{k-1} + m \]
  - Mean
    \[ \theta_k = \frac{\kappa_{k-1}}{\kappa_{k-1} + m} \theta_{k-1} + \frac{m}{\kappa_{k-1} + m} \bar{b} \]
  - Scale matrix
    \[ \Lambda_k = \Lambda_{k-1} + \frac{\kappa_{k-1} m}{\kappa_{k-1} + m} (\bar{b} - \theta_{k-1})(\bar{b} - \theta_{k-1})^T \]

- **m** - Number of samples to update the model
- **\(\bar{b}\)** - Mean of all samples
The Prior Term

• Using the expectation of marginal posterior distributions as the model parameters update.

Joint Posterior
Normal Inverse-Wishart

\[ p(\mu_b, \Sigma_b | b) \]

Marginalizing w.r.t \( \Sigma_b \)

\[ p(\mu_b | b) \propto \text{Multivariate Student t} \]

\[ \mu_{b_k} = E(\mu_b | b) = \theta_k \]

Marginalizing w.r.t \( \mu_b \)

\[ p(\Sigma_b | b) \propto \text{Inv-Wishart} \]

\[ \Sigma_{b_k} = E(\Sigma_b | b) = (\nu_k - n - 1)^{-1} \Lambda_k \]

The Prior distribution is continuously kept up to date
The Algorithm

Precompute: PDM: \( s_0, \Phi, \Psi, \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n) \)

Initial estimate (\( b_0, \Sigma_0 \), (\( q_0, \Sigma_0^q \))

for \( k = 1:1:\text{MaxIterations} \)

Warp Image to the base mesh, using the current pose parameters

Generate current shape \( s = S(s_0 + \Phi b_k; q_k) \)

for \( i = 1:1:\text{Landmarks} \)

Evaluate detectors response

Find the likelihood parameters \( y_i, \Sigma y_i \)

end

Estimate the shape/pose parameters:

Update the parameters of Normal Inv-Wishart distribution \( \nu_k, \kappa_k, \theta_k, \Lambda_k \)

Expectation of the prior shape parameters \( \mu_{b_k} = \theta_k, \Sigma_{b_k} = (\nu_k - n - 1)^{-1} \Lambda_k \)

Evaluate the global shape parameters and the covariance \( \mu_k, \Sigma_k \)
Hierarchical Search (BASM-KDE-H)

- When response maps are approximated by KDE representations.

\[
y_{i}^{\text{KDE}(\tau+1)} = \frac{\sum_{\mathbf{z}_i \in \Omega_{y_i}} \mathbf{z}_i \cdot p_i(\mathbf{z}_i) \cdot \mathcal{N}(\mathbf{y}_i^{\text{KDE}(\tau)}|\mathbf{z}_i, \sigma_{h_j}^2 \mathbf{I}_2)}{\sum_{\mathbf{z}_i \in \Omega_{y_i}} p_i(\mathbf{z}_i) \cdot \mathcal{N}(\mathbf{y}_i^{\text{KDE}(\tau)}|\mathbf{z}_i, \sigma_{h_j}^2 \mathbf{I}_2)}
\]

Bandwidth schedule
\[
\sigma_{h}^2 = (15, 10, 5, 2)
\]

Standard Search

for \( k=1:1:\text{MaxIterations} \)
  for \( i=1:1:v \) (LandMarks)
    Evaluate de Detectors Response
    Mean-Shift Landmark Update \( \sigma_{h}^2 = (15, 10, 5, 2) \)
  end
end

Global Optimization

Hierarchical Search

for \( k=1:1:\text{MaxIterations} \)
  for \( \sigma_{h}^2 = (15, 10, 5, 2) \)
    for \( i=1:1:v \) (LandMarks)
      Evaluate de Detectors Response
      Mean-Shift Landmark Update \( \sigma_{h_j}^2 \)
    end
  end
end

Global Optimization
Evaluation Results

Reference 7.5 RMS

<table>
<thead>
<tr>
<th>Method</th>
<th>IMM (240 images)</th>
<th>XM2VTS (2360 images)</th>
<th>BioID (1521 images)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASM</td>
<td>50.0</td>
<td>47.4 (+16.7)</td>
<td>70.0</td>
</tr>
<tr>
<td>BASM-WPR (our method)</td>
<td><strong>58.4</strong> (+8.4)</td>
<td><strong>47.4</strong> (+16.7)</td>
<td><strong>77.1</strong> (+7.1)</td>
</tr>
<tr>
<td>CQF</td>
<td>45.4</td>
<td>10.9</td>
<td>47.0</td>
</tr>
<tr>
<td>GMM3</td>
<td>40.8 (-4.6)</td>
<td>10.4 (-0.5)</td>
<td>51.7 (+4.7)</td>
</tr>
<tr>
<td>BCLM-GR</td>
<td>48.3 (+2.9)</td>
<td>15.9 (+5.0)</td>
<td>54.2 (+7.2)</td>
</tr>
<tr>
<td>BASM-GR (our method)</td>
<td><strong>51.8</strong> (+6.4)</td>
<td><strong>19.7</strong> (+8.8)</td>
<td><strong>63.5</strong> (+16.5)</td>
</tr>
<tr>
<td>SCMS-KDE</td>
<td>54.6</td>
<td>35.7</td>
<td>69.0</td>
</tr>
<tr>
<td>BCLM-KDE</td>
<td>57.1 (+2.5)</td>
<td><strong>43.4</strong> (+7.7)</td>
<td><strong>71.9</strong> (+2.9)</td>
</tr>
<tr>
<td>BASM-KDE (our method)</td>
<td><strong>65.4</strong> (+10.8)</td>
<td><strong>57.0</strong> (+21.3)</td>
<td><strong>80.3</strong> (+11.3)</td>
</tr>
<tr>
<td>BASM-KDE-H (our method)</td>
<td>64.0 (+9.4)</td>
<td>56.6 (+20.9)</td>
<td>79.9 (+10.9)</td>
</tr>
<tr>
<td>BASM-KDE Fusion of 2 Detectors</td>
<td><strong>72.5</strong> (+17.9)</td>
<td><strong>58.7</strong> (+23.0)</td>
<td><strong>88.2</strong> (+19.2)</td>
</tr>
</tbody>
</table>
Tracking Performance

- FGNET Talking Face Video Sequence

<table>
<thead>
<tr>
<th>RMS Error</th>
<th>ASM</th>
<th>CQF</th>
<th>GMM3</th>
<th>SCMS-KDE</th>
<th>BCLM-KDE</th>
<th>BASM-KDE</th>
<th>BASM-KDE-H</th>
<th>BASM-KDE Fusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>10.5</td>
<td>10.6</td>
<td>11.1</td>
<td>8.2</td>
<td>9.5</td>
<td>6.4</td>
<td>6.3</td>
<td>5.9</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>6.4</td>
<td>3.9</td>
<td>4.3</td>
<td>2.6</td>
<td>3.6</td>
<td>1.7</td>
<td>1.5</td>
<td>1.5</td>
</tr>
</tbody>
</table>
Tracking Performance
Labeled Faces in the Wild (LFW)
Conclusions

• Bayesian formulation for aligning faces in unseen images.
• New global optimization strategy infers both shape and pose parameters, in MAP sense, by explicitly modeling the prior distribution.
• The prior distribution is continuously kept up to date.
• Recursive Bayesian estimation is used to treat the mean and covariance as random variables.
• Extensive evaluations were performed on several standard datasets (IMM, XM2VTS, BioID, LFW and FGNET Talking Face) against state-of-the-art methods while using the same local detectors.

Acknowledgements

• Work supported by the Portuguese Science Foundation (Fundação para a Ciência e Tecnologia - FCT) under the project grant PTDC/EIA-CCO/108791/2008.
• Pedro Martins, Rui Caseiro and João F. Henriques also acknowledge the FCT through the grants SFRH/BD/45178/2008, SFRH/BD74152/2010 and SFRH/BD/75459/2010, respectively.
Qualitative Evaluation - Labeled Faces in the Wild