Contourlet Appearance Model (CAM) for Facial Age Estimation

Khoa Luu (kh_lu@encs.concordia.ca)
Keshav Seshadri (kseshadr@andrew.cmu.edu)
Prof. Marios Savvides (msavvid@ri.cmu.edu)
Prof. Tien D. Bui (bui@encs.concordia.ca)
Prof. Ching Y. Suen (suen@encs.concordia.ca)
Overview

- Introduction
- Motivation
- Prior Work
- Nonsubsampled Contourlet Transform (NSCT)
- Our Contourlet Appearance Model (CAM)
  - Shape Modeling
  - Texture Modeling
- Experimental Results
- Conclusions
Introduction

- Face based age estimation applications
  - Law enforcement
  - Facial recognition
- Human face age progression
  - Growth and development (childhood to pre-adulthood)
  - Adult aging
- Active Appearance Models (AAMs)\(^\text{[1]}\) – tool in age estimation

Motivation

- **Active Appearance Models (AAMs)**
  - Don’t generalize well to unseen images [2]
  - Build global texture models - poor tolerance to illumination variation

- **Active Shape Models (ASMs)** [3]
  - Accurate method for automatic facial landmarking
  - Generalize better to fit unseen images [2]
  - Build local texture models – more robust to illumination variation

- **Contourlet Appearance Model (CAM)**
  - Modified Active Shape Model (MASM) [4] for facial shape determination
  - Nonsubsampled Contourlet Transform (NSCT) [5] for texture modeling
  - Results show improved facial age estimation using CAM

---

Working of AAMs

**Step 1:** Manual annotation and shape representation

Training set (n images)  Collection of n shapes  Shape representation

**Step 2:** Shape Alignment

**Step 3:** Shape Modeling using PCA

\[ \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \]

\[ \Sigma = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})^T \]

The generated shape model:

\[ x = \bar{x} + P_s b_s \]

**Step 4:** Gray-level Appearance Modeling using PCA

The gray-level appearance is generated as follows:
1. Apply appearance alignment by warping the control points to match the mean shape, by using the triangulation algorithm.
2. Correct the lighting of gray-level appearance.
3. Apply PCA to obtain the linear model for the gray-level appearance:

\[ g = \bar{g} + P_g b_g \]

**Step 5:** Principal Component Analysis

The shape model \[ x = \bar{x} + P_s b_s \]

The texture model \[ g = \bar{g} + P_g b_g \]

Combine them together \[ b = \begin{pmatrix} W_s b_s \\ b_g \end{pmatrix} = \begin{pmatrix} W_s P_s^T (x - \bar{x}) \\ P_g (g - \bar{g}) \end{pmatrix} \]

\[ b = P_c c \]
Work on Improving AAMs

For more details, see [Xinbo Gao et al. in: IEEE SMC, 2010]

**Efficiency**: Aims to reduce the time and space costs

- Iain Matthews et al., ”Active Appearance Models Revisited,” in: *IJCV*, 2004

**Discrimination**: Aims to preserve discriminative information


**Robustness**: Aims to be robust against inconstant factors (pose etc.)

- Changbo Hu et al., "Active Wavelet Networks for Face Alignment", in: *BMVC*, 2003
## Work on Age Estimation

For more details, see [Fu et al. in: *TPAMI*, 2010]

<table>
<thead>
<tr>
<th>Authors</th>
<th>Approach</th>
<th>Age ranges</th>
<th>Results</th>
<th>Database</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kwon et al., 1999</td>
<td>Facial geometry &amp; wrinkle analysis</td>
<td>Three groups: babies, youths, adults</td>
<td>100%</td>
<td>47 high resolution images</td>
</tr>
<tr>
<td>Horg et al., 2001</td>
<td>Facial geometry &amp; wrinkle analysis</td>
<td>Four groups: babies, youths, middle adults, old adults</td>
<td>90.5% (on training photos), 81.6% (on testing photos)</td>
<td>116 training photos, 114 testing photos</td>
</tr>
<tr>
<td>Lanitis et al., 2002</td>
<td>Weighted Person Specific Aging Function (WAS)</td>
<td>0 – 30</td>
<td>MAEs (years): 3.91</td>
<td>FG-NET (350 training, 150 testing images)</td>
</tr>
<tr>
<td>Hayashi et al., 2002</td>
<td>Wrinkle analysis and DTHT</td>
<td></td>
<td>27%</td>
<td></td>
</tr>
<tr>
<td>Geng et al., 2007</td>
<td>AGing pattErn Subspace (AGES)</td>
<td>0 – 69</td>
<td>MAEs (years): 6.77</td>
<td>FG-NET (1002 images)</td>
</tr>
<tr>
<td>Fu et al., 2007</td>
<td>Manifold</td>
<td>0 – 93</td>
<td>MAEs: 5 - 6 (F), 5-6 (M)</td>
<td>UIUC-IFP (4000 F, 4000 M)</td>
</tr>
<tr>
<td>Yan et al., 2008</td>
<td>Spatially Flexible Patches (SFPs)+ Gaussian Mixture Models (GMMs)</td>
<td>0 – 93</td>
<td>MAEs: 8.53 (F), 7.82 (M)</td>
<td>YAMAHA (8000 images)</td>
</tr>
<tr>
<td>Zhuang et al., 2008</td>
<td>HMM Supervector</td>
<td>0 – 93</td>
<td>MAEs: 6.3 (F), 5.4 (M)</td>
<td>YAMAHA (8000 images)</td>
</tr>
<tr>
<td>Yan et al., 2008</td>
<td>Regression from Patch-Kernel (RPK)</td>
<td>0 – 69 (FG-NET)</td>
<td>MAEs: 4.95</td>
<td>FG-NET, 1002 images</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0 – 93 (YAMAHA)</td>
<td>4.94 (M), 4.38 (F)</td>
<td>YAMAHA, 8000 images</td>
</tr>
<tr>
<td>Luu et al., 2009</td>
<td>AAMs + SVR</td>
<td>0 – 69</td>
<td>MAEs: 4.37</td>
<td>FG-NET (802 training, 200 testing images)</td>
</tr>
<tr>
<td>Guo et al., 2009</td>
<td>Biologically Inspired Features (BIFs)</td>
<td>0 – 93</td>
<td>MAEs: 3.47 (M), 3.91 (F)</td>
<td>YAMAHA</td>
</tr>
<tr>
<td>Guo et al., 2009</td>
<td>BIFs and Manifold</td>
<td>0 – 93</td>
<td>MAEs: 2.58 (M) and 2.61 (F)</td>
<td>YAMAHA</td>
</tr>
<tr>
<td>Montillo et al., 2009</td>
<td>Anthropometric measurements</td>
<td>0 – 20</td>
<td>MAEs: 3.43</td>
<td>FG-NET</td>
</tr>
<tr>
<td>Dehshibi et al., 2010</td>
<td>Seven geometric ratios and three wrinkle areas + Neural network</td>
<td></td>
<td>86.64%</td>
<td>Iranian face database (IFDB)</td>
</tr>
<tr>
<td>Luu et al., 2011</td>
<td>Hybrid Facial Features</td>
<td>0 – 69</td>
<td>MAEs: 4.22</td>
<td>FG-NET</td>
</tr>
<tr>
<td>Chen et al., 2011</td>
<td>Feature Fusion and Model Selection</td>
<td>18 – 93</td>
<td>MAEs: 5.65</td>
<td>PAL (540 images)</td>
</tr>
</tbody>
</table>
Nonsubsampled Contourlet Transform (NSCT)

Problems with traditional feature extraction methods:

- Face contains weak wrinkles which must be distinguished from noise
- Traditional approaches (Canny edge detection, wavelet, ridgelet, etc.) are not suitable for enhancing facial wrinkles

NSCT observations:

- Strong edges are pixels with large magnitude coefficients in all subbands.
- Weak edges embody pixels having both large and small magnitude coefficients in directional subbands within same scale
- Noise corresponds to pixels having small magnitude coefficients in all subbands
- Apply NSCT in the logarithmic domain to enhance range of values
1. Given an image $I$, calculate its logarithm: $L = \log(I)$
2. Apply NSCT on $L$ for $N$ levels
3. Estimate the noise standard deviation
4. For each level of the pyramid:
   (a) Estimate the noise variance
   (b) At each pixel, compute $m$ and $m'$ of the corresponding coefficients in all directional sub-bands at this level
5. Apply the inverse NSCT to get the processed image

$$v(x) = \begin{cases} 
  x & \text{if } (c \sigma \leq \overline{m}) \\
  \max\left(\left(\frac{c \sigma}{|x|}\right)^p, 1\right) & \text{if } (c \sigma > \overline{m}, c \sigma \leq m) \\
  0 & \text{if } (c \sigma > \overline{m}, c \sigma > m)
\end{cases}$$

Lowpass subband
Bandpass directional subbands

Low-pass image
Strong edge pixels
Weak edge pixels
Noise
NSCT Features

(a) Original facial image
(b) Noisy image ($\sigma = 0.1$)
(c) Low-pass image extracted by NCST on noisy image
(d) Strong edge image extracted by NCST on noisy image
(e) Weak edge image extracted by NCST on noisy image
(f) Noise components extracted by NCST on noisy image
(g) LBP image of noisy image
(h) Noisy image filtered using a Gabor filter
CAM Shape Modeling

- 68 point facial landmarking scheme
- Modified Active Shape Model (MASM)\(^4\)
  - 2D profiles and search space
  - PCA subspace of local texture around landmarks
  - Edge detection and weighting for improved facial boundary fitting

**CAM Texture Modeling**

**Step 1:** Facial landmarks extracted using MASM and shape parameters $b_s$ obtained.

**Step 2:** CAM texture modeling using PCA on convex hull of “shape-free” texture images and weak edge images.

\[
\begin{align*}
\text{warped gray-level texture} & \quad b_t = P_t^T(t - \bar{t}) \\
\text{warped weak edges} & \quad b_w = P_w^T(w - \bar{w})
\end{align*}
\]

**Step 3:** PCA on combined shape and texture parameters vector $b$ resulting in $b = P_c c$.

---

[Diagram showing the process of CAM texture modeling with steps illustrated through images of input images, shapes, and texture reconstructions]
Experimental Results

- CAM for Age Estimation
  - 2 aging functions for youths (0 – 20 years) and adults (21+) built using Support Vector Regression (SVR) on training set
  - Combined CAM features extracted and appropriate age function selected

- Training
  - MASM: Trained on a set of 500 images (115 subjects) from the NIST Multiple Biometric Grand Challenge - 2008 (MBGC – 2008)[6] database

- Testing
  - Leave One Person Out (LOPO) scheme is used on FG-NET[7] (1002 images, 82 subjects, age range 0-69) and PAL[8] (443 images, age range 18-91) databases

- Evaluation Methodology
  - Mean Absolute Error (MAE) \[ MAE = \frac{\sum_{i=1}^{N_t} |\hat{x}_i - x_i|}{N_t} \]
  - Cumulative Score (CS) \[ \text{CumScore}(\theta) = \frac{N_e \leq \theta}{N_t} \times 100 \]

Experimental Results

### MAE values (in years) when LOPO scheme was used on FG-NET database

<table>
<thead>
<tr>
<th>Method</th>
<th>MAE (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAM</td>
<td>4.12</td>
</tr>
<tr>
<td>BIFS (Guo. et al.)</td>
<td>4.77</td>
</tr>
<tr>
<td>AAM + LBP</td>
<td>6.17</td>
</tr>
<tr>
<td>AAM</td>
<td>6.77</td>
</tr>
<tr>
<td>AGES (Geng. Et al.)</td>
<td>6.77</td>
</tr>
</tbody>
</table>

### Cumulative scores by different methods at error levels from 0-10 years when tested on FG-NET database

<table>
<thead>
<tr>
<th>Method</th>
<th>MAE (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAM</td>
<td>6.00</td>
</tr>
<tr>
<td>AAM + LBP</td>
<td>7.17</td>
</tr>
</tbody>
</table>

MAE values (in years) when LOPO scheme was used on FG-NET database

Cumulative scores by different methods at error levels from 0-10 years when tested on FG-NET database
Conclusions

- Contourlet Appearance Model (CAM) is proposed to localize facial landmarks and extract texture based features

- Modified Active Shape Model (MASM) helps to accurately locate a dense set of facial landmarks

- Nonsubsampled Contourlet Transform helps to obtain discriminative features for age estimation

- Results of using CAM based age estimation technique on databases (such as FG-NET and PAL) highlight improved accuracy of the method
Thank You