Large-Pose Face Alignment via CNN-Based Dense 3D Model Fitting

Amin Jourabloo, Xiaoming Liu
Computer Vision Lab, Department of Computer Science and Engineering, Michigan State University

Introduction

The main contributions of proposed method:
- Large-pose face alignment by fitting a dense 3DMM.
- Cascaded CNN-based 3D face model fitting with integrated landmark marching.
- Dense 3D face-enabled pose-invariant local features.

Prior Work

<table>
<thead>
<tr>
<th>Method</th>
<th>Dense 3D model fitting</th>
<th>Visibility</th>
<th>Pose-related database</th>
<th>Pose range</th>
<th>Landmark #</th>
<th>Estimation Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>TSPM (CVPR2013)</td>
<td>NO</td>
<td>Yes</td>
<td>COFW</td>
<td>frontal w. occlu</td>
<td>19</td>
<td>8.5</td>
</tr>
<tr>
<td>RCPR (ECCV2015)</td>
<td>NO</td>
<td>Yes</td>
<td>AFLW, AFW</td>
<td>all poses</td>
<td>6</td>
<td>11.1</td>
</tr>
<tr>
<td>CDMM (ECCV2013)</td>
<td>NO</td>
<td>No</td>
<td>AFW</td>
<td>all poses</td>
<td>6</td>
<td>9.1</td>
</tr>
<tr>
<td>TCDCN (ECCV2014)</td>
<td>NO</td>
<td>Yes</td>
<td>AFLW, AFW</td>
<td>all poses</td>
<td>5</td>
<td>8.0, 8.2</td>
</tr>
<tr>
<td>PIFA (ECCV2015)</td>
<td>Yes</td>
<td>Yes</td>
<td>AFLW, AFW</td>
<td>all poses</td>
<td>21, 6</td>
<td>6.5, 8.6</td>
</tr>
</tbody>
</table>

Proposed Method

3D Surface-Enable Visibility

- We use the average of normal around a 3D landmark as the person-specific 3D surface normal.
- The sign of z coordinate indicates 2D landmark visibility.

Experimental Results

- AFLW dataset experiments
  - 5200 images selected evenly within \([0°, 30°, 30°, 60°, 60°, 90°]\) yaw angles.
  - Randomly partitioned into 3901 training and 1299 testing images.

3D Face Modeling

- Proposed method is represented by a cascaded CNN regressor.

Cascaded Couple-Regressor

- Projection Matrix Parameter Regressor
  \[ \theta_0 = \arg \min_{\theta} \sum_{i} |\text{CNN}(U, x, t, \theta) - y_i| \]

- 3D Shape Parameter Regressor
  \[ \theta_1 = \arg \min_{\theta} \sum_{i} |\text{CNN}(A, x, t, \theta) - y_i| \]

- The average NME of each landmark
  \[ \text{NME} = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i| \]

- Result across stages
  - Extracted features
  - Alignment results

- Binary code: http://cvlab.cse.msu.edu/project-pifa.html

3D scans w. labels

2D image w. landmarks

2D Landmarks \(U\) are represented as pair of projection matrix parameter \(m\) and 3D shape parameter \(p\).

Piecewise Affine-Wrapped Feature (PAWF)

Direct 3D Projected Feature (D3PF)