A Hierarchical Training and Identification Method using Gaussian Process Models for Face Recognition in Videos

Negar Hassanpour and Liang Chen
University of Northern British Columbia, Canada

1. Ensemble of Abstract Sequence Representatives (EASR)

EASR is a vector-based representation for each image set that addresses the uncertainties in data as follows:
1. Noise relaxation by transferring data to a higher level representation structure using sampling and superposition.
2. Outlier filtering to monitor the quality of ASRs generated by random sampling and counter face tracking errors.

\[ \beta = \frac{\gamma}{|\Psi|} \quad \gamma = \sum_{i=1}^{m} n_i \psi_{ij} = \left( \beta_p, \beta_q \right) \| \Delta \cos(\theta_{pq}) = \cos(\theta_{pq}) \quad \text{Reject}(\beta_0) \Psi_0 < \Psi - 2\Delta \]

Similarity Calculation and Identification: \( S_{ij} = \text{Max} \left( \Psi_{ij} \right) \quad \text{identity} = \text{ArgMax} \{ S_p \}

Although EASRs are designed to be noise resilient, their inherent linear structure does not allow for capturing complex variations in data. Thus, a second nonlinear component (i.e., Gaussian process models) is added to address this issue.

2. Gaussian Process Models

Gaussian process is a generalization of Gaussian probability distribution and a Bayesian alternative to kernel methods. Since models learned by GP are non-parametric, any hard assumption on the structure of the model is safely avoided.

GP for Regression:
\[
k(x_i, x_j) = \sigma_f^2 \exp(-\frac{1}{2\sigma_x^2}(x_i - x_j)^2)
\]
\[
p(f(x_i), X, y) = N(f(x_i|\mu, \Sigma))
\]
\[
\mu_j = \mu(x_j) + K_j^T K^{-1} y
\]
\[
\Sigma_j = K_j - K_j^T K^{-1} K_j
\]

GP for Classification:
- Use \( \mu_j \) for one vs. rest classification at the frame level.
- Aggregate the output for each GP model by sum-fusion of \( f_j \) s in all frames in the respective probe video.
- The identity associated with the highest aggregated output among all GP models is reported by the GP ensemble.

3. Learning Scheme: Specialization—Generalization

Starting with \( n \) subjects and \( m \) sequences for each subject in training data, \( SQ_{in} \) contains all training sequences.

Specialization step:
1. Calculate EASRs for all training sequences
2. Calculate the pair-wise similarity \( S_{ij} \) between each two subjects \( i \) and \( j \)
3. For each subject \( i \) find the top \( k \) nearest subjects with the highest \( S_{ij} \) and store \( j \) s in \( NS_i \)
4. Train GP, with all frames from \( SQ_{i\text{;im}} \) as (+) instances and randomly sample equal number of frames from \( SQ_{j\text{;im}} \), \( j \in NS_i \) as (-) instances

Generalization step:
5. Use GP to label each sequence in \( SQ_{j\text{;im}} \), \( j \in NS_i \), for each frame \( f \) if \( GP(f) > \theta \) (i.e., mislabelled) add it to \( GenL_i \) list to be retrained to GP
6. Update GP with all frames \( f \) in \( GenL_i \) as (-) instances


Goal: Aggregating the predictions of the EASR module and GP module, to come up with the most accurate prediction of identity for a probe video sequence.
1. Clearly, if predictions of both modules agree on the same identity, that prediction is reported.
2. Otherwise, the hierarchical approach is used as follows:
   - EASR module has the priority since it is more noise tolerant. We trust EASR-based prediction when it identifies the probe video sequence by a clear winner (a pre-defined threshold).
   - If the constraint for the cut-off is not satisfied, it indicates that the EASR module is not confident in its prediction, therefore the label generated by the GP module is reported as the final predicted identity.

5. Performance Evaluation

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Honda/UCSD</th>
<th>CMU-Mobo</th>
<th>YouTube Celebrities</th>
<th>Computation Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
<td>Year</td>
<td>50</td>
<td>All</td>
<td>50</td>
</tr>
<tr>
<td>MSM</td>
<td>1998</td>
<td>87.69 ± 6.12</td>
<td>90.26 ± 2.15</td>
<td>92.50 ± 2.71</td>
</tr>
<tr>
<td>MDA</td>
<td>2009</td>
<td>87.69 ± 2.81</td>
<td>96.41 ± 1.40</td>
<td>84.17 ± 6.56</td>
</tr>
<tr>
<td>AHSID</td>
<td>2010</td>
<td>88.21 ± 3.89</td>
<td>83.59 ± 3.89</td>
<td>92.50 ± 2.71</td>
</tr>
<tr>
<td>CHSID</td>
<td>2010</td>
<td>86.13 ± 3.92</td>
<td>91.28 ± 2.29</td>
<td>92.50 ± 2.71</td>
</tr>
<tr>
<td>SANP</td>
<td>2011</td>
<td>87.18 ± 6.01</td>
<td>96.41 ± 2.29</td>
<td>92.50 ± 2.71</td>
</tr>
<tr>
<td>RNP</td>
<td>2013</td>
<td>88.21 ± 4.66</td>
<td>93.33 ± 2.29</td>
<td>92.50 ± 2.71</td>
</tr>
<tr>
<td>MSSRC</td>
<td>2013</td>
<td>91.18 ± 1.40</td>
<td>93.85 ± 2.29</td>
<td>91.11 ± 3.49</td>
</tr>
<tr>
<td>ISRC</td>
<td>2014</td>
<td>92.31 ± 4.44</td>
<td>95.38 ± 1.11</td>
<td>94.44 ± 2.20*</td>
</tr>
<tr>
<td>EASR+GP</td>
<td>2015</td>
<td>94.87 ± 4.05</td>
<td>99.40 ± 1.5*</td>
<td>93.61 ± 2.71</td>
</tr>
</tbody>
</table>

Notes:
- * indicates statistically significant improvement of accuracy compared to the second best result at \( \alpha = 0.05 \).
- # indicates statistically significant improvement of accuracy compared to the second best result at \( \alpha = 0.1 \), and
- \( * \) indicates no significant difference between EASR+GP and the best result in the rest of the column (statistically).
- N/A indicates online-only methods