Heterogeneous Face Recognition: Matching NIR to Visible Light Images

Brendan Klare and Anil K Jain
What is Heterogeneous Face Recognition?

- Matching two face images acquired in alternate modalities
  - Gallery images are standard face image in visible spectrum (e.g. mugshot or passport photo)
  - Probe images are in some alternate modality (only reliable information available)

Examples:

- Forensic sketch
- Near infrared
- Medium infrared
- Video
- 3D
- Face at a distance
Many scenarios exist in which we cannot control the modality of the query image.

In these cases our target databases are still regular face photographs.

Consider Near Infrared (NIR) face recognition:
- Suppose we are acquiring face images at night time.
- This is a very realistic scenario in military and law enforcement.
- NIR face image may appear different from a photograph, but it still has salient information about subject’s identity.
NIR and Illumination Variation

Appearance of NIR images does not change with change in illumination

Images from Li et al, PAMI 2007
Approaches to Heterogeneous Face Recognition

- **Image Synthesis**
  - **Pro**: Compatible with commercial face recognition systems
  - **Con**: Generative method, sensitive to parameters, requires large amounts of training data

- **Feature-based methods**
  - Extract domain invariant image feature descriptors from both the face images [Liao et al 2009, Klare and Jain 2010]
  - **Pro**: Discriminative method, fast, supports improvement via training (e.g. feature extraction)
  - **Con**: Image feature descriptor may not capture all the salient information, large feature vectors -> SSS problem
Proposed Heterogeneous FR Method

- **Representation:**
  - Feature-based
    - SIFT descriptors
    - LBP descriptors
      - [Liao et al., 2009]

- **Feature Extraction:**
  - Random Subspaces
    - Discriminant analysis performed on a series of random subspaces
      - [Wang and Tang, 2004]

- **Matching**
  - Nearest Neighbor
  - Sparse Representation
Matching Algorithm

Feature Representation

Applied to both NIR and VIS Images

Discriminative Ensemble Model

Random Image Feature-based Subspaces

Perform LDA on each random subspace

Concatenate K LDA feature projections of the face image into a single vector

Matching

Nearest Neighbor

Identity($I^N$) = argmin$_i$ ||$Y^V_i$ - $Y^N$||$_2$

Sparse Representation

\[ x = \arg\min_x \|Ax - Y^N\|_2^2 + \lambda\|x\|_1 \]

where, $A$ is column-wise matrix of $Y^V_i$ and Identity($I^N$) is subject with most non-zero coefficients in $x$
Discriminant analysis learns linear combinations of the most salient features

*Feature-based representation vs. pixel representation:*
  - Both have very high dimensionality
  - Feature-based representation is less redundant
    - Limits the use of standard approach of PCA followed by LDA

Approaches to avoid overfitting:
  - R-LDA, D-LDA, **RS-LDA**

Method used similar to Random Subspace LDA (RS-LDA):
  - Difference: Randomly sample *patches* instead of *pixels*
Matching

- Two approaches reported:
  - **Nearest Neighbor Matching**
    - Each of the k subspace vectors concatenated into a single vector
    \[
    \text{Identity}(I^N) = \arg\min_i \|Y_i^V - Y^N\|_2
    \]
  - **Sparse Representation Matching:**
    - If A is a matrix whose columns are the feature representation of each gallery, and y is the probe feature vector
    \[
    x = \arg\min_x \|Ax - y\|_2^2 + \lambda\|x\|_1
    \]
    - Then x (ideally) will contain non-zero coefficients for only those gallery images that correspond to probe subject
Experimental Results

Dataset:

- From the CASIA HFB Database
- Same data set used previously in literature [1][2]
- Details:
  - 202 Subjects
  - 3,002 NIR images
  - 2,095 VIS images
- Training / Testing Split
  - Training - 102 subjects, Testing - 100 subjects (disjoint)
  - Five random splits used to generate the results

Comparative Baseline:

- State-of-the art matcher: FaceVACS (by Cognitec)
  - FaceVACS performs extremely well on this problem (outperforms previously published methods)

---

[1] S. Liao et al., Heterogeneous face recognition from local structures of normalized appearance. ICB, 2009
Verification Scenario

<table>
<thead>
<tr>
<th>FP Rate</th>
<th>0.1 %</th>
<th>1.0 %</th>
<th>10.0%</th>
</tr>
</thead>
<tbody>
<tr>
<td>FE-NN</td>
<td>48.78 ± 3.87</td>
<td>73.49 ± 3.72</td>
<td>95.94 ± 0.56</td>
</tr>
<tr>
<td>FE-SR</td>
<td>77.56 ± 2.96</td>
<td>94.04 ± 1.49</td>
<td>99.63 ± 0.21</td>
</tr>
<tr>
<td>NNSR</td>
<td>79.05 ± 4.48</td>
<td>91.37 ± 1.99</td>
<td>98.15 ± 0.46</td>
</tr>
<tr>
<td>FV</td>
<td>85.62 ± 2.17</td>
<td>93.80 ± 0.65</td>
<td>98.24 ± 0.07</td>
</tr>
<tr>
<td>NNSR + FV</td>
<td>93.45 ± 0.96</td>
<td>97.06 ± 0.39</td>
<td>99.41 ± 0.17</td>
</tr>
</tbody>
</table>

- **Previous best results:**
  - Liao et al. [2009] 87.5% at FP rate of 1.0%
    - Used 150 subjects for training and 52 for testing

- **Interpretation:**
  - With only *falsely accepting 1% of subjects, we can truly accept ~97% of the subjects*
Fusing FaceVACS with our method:
  - ~98% Rank-1 accuracy achieved
Summary

- Proposed method performed well in matching NIR face images to visible face images: \textit{TAR}=94.0\% at \textit{FAR}=1.0\%.
- Commercial matcher (FaceVACS) performed well.
- Fusion of our method with FaceVACS improves recognition results.
- These results offer promising solution to face recognition in varying illumination.
- Future directions:
  - Incorporate NIR to VIS matching on face recognition at a distance (FRAD) scenario.
  - Matching Thermal IR to VIS face images.