Improving RGB-D face recognition via transferring pretrained 2D networks

Xingwang Xiong, Xu Wen, and Cheng Huang

https://github.com/XingwXiong/Face3D-Pytorch

3D Face Recognition Algorithm Challenge (3DFRAC)

ICT, Chinese Academy of Sciences
Face Representations

RGB image

Point cloud

3D Mesh

Depth image

3D Face Images
3D Face Recognition Algorithm Challenge

- RGB-D Face Recognition
- The value in the depth image reflects the distance of scene object surface from the viewpoint.
Why do we need RGB-D FR?

- 2D FR is sensitive to external variations
  - Poses
  - Facial expressions
  - Illuminations

- Extra **low-level** patterns on depth images
  - Smooth variations
  - Contracts
  - Borders & global layouts

- **Face Anti-spoofing** (ICPR 2018)
Open-set vs. Closed-set FR

- Open-set FR
  - Classification
- Closed-set FR
  - Face embedding
  - Similarity comparison
    - SVM
    - KNN
- 3DFRAC
  - a closed-set problem
RGB Images vs. Depth Images

RGB images
- High frequency patterns
  - Textures & Details
- Easy to obtain
- Massive scale
  - \( \sim 3.3 \text{ million faces}^1 \)
  - \( \sim 9K \text{ identities}^1 \)

Depth images
- Low frequency patterns
  - Smooth variations
  - Contracts
  - Borders & Global layouts
- Not enough to learn a deep CNN
  - \( \sim 403K \text{ million faces}^2 \)
  - \( \sim 1.2K \text{ identities}^2 \)

1. VGGFace2
2. Intellifusion RGB-D face dataset
Goal

- To leverage both conventional RGB-based works and depth features
Inter-modal Transfer Learning

2D pretrained network

RGB-D network

Copy weights
Inter-modal Transfer Learning

- Use **ResNet-50** as the backbone network
- Copy pretrained weights of middle layers
- Fine-tune the whole model with 224x224 RGB-D images
Preprocessing

- Face Detection & Alignment
  - MTCNN (SPL 2016)
- Randomly horizontally flipping
- Normalizing
  - 0-1 range
Results

- **94.64% accuracy** on the Intellifusion RGB-D dataset
- **Won 1st** on 3DFRAC

<table>
<thead>
<tr>
<th>Method</th>
<th>Input data</th>
<th>CNN Models</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretrained on ImageNet</td>
<td>RGB images</td>
<td>ResNet-50</td>
<td>94.47</td>
</tr>
<tr>
<td>Train from scratch</td>
<td>RGB-D images</td>
<td>RGB-D ResNet50</td>
<td>88.36</td>
</tr>
<tr>
<td>Pretrained on ImageNet</td>
<td>RGB-D images</td>
<td>RGB-D ResNet50</td>
<td><strong>94.64</strong></td>
</tr>
</tbody>
</table>
Conclusions

- Inter-modal transfer learning from pretrained 2D networks to RGB-D networks improves recognition accuracy

- Code is open-sourced
  - https://github.com/XingwXiong/Face3D-Pytorch

- Email
  - xiongxingwang@ict.ac.cn
QUESTIONS AND ANSWERS