Quality-based face representation for matching in uncontrolled environments

Nelson Méndez-Llanes, Katy Castillo-Rosado, Heydi Méndez-Vázquez and Massimo Tistarelli

January 14, 2021

Advanced Technologies Application Center (CENATAV)
University of Sassari
Face Recognition in Uncontrolled Environments

- Although the high levels of accuracy achieved by deep face models, face recognition in unconstrained environments still remains an open problem.
Face Recognition in Uncontrolled Environments

- Although the high levels of accuracy achieved by deep face models, face recognition in unconstrained environments still remains an open problem.
- The majority of the approaches abruptly decrease their performance in the presence of low quality face images.
Deep learning approaches deal with this problem by using massive training sets containing face samples representative of all these factors.
Deep learning approaches deal with this problem by using massive training sets containing face samples representative of all these factors.

In most of the cases, the proposals tend to deal with only one of the most difficult problems (E.g. pose, illumination, resolution, occlusion).
Deep learning approaches deal with this problem by using massive training sets containing face samples representative of all these factors.

In most of the cases, the proposals tend to deal with only one of the most difficult problems (E.g. pose, illumination, resolution, occlusion).

There are just a few approaches trying to deal jointly with these challenges.
Face Recognition from Low Quality Images

- Deep learning approaches deal with this problem by using massive training sets containing face samples representative of all these factors.
- In most of the cases, the proposals tend to deal with only one of the most difficult problems (E.g. pose, illumination, resolution, occlusion).
- There are just a few approaches trying to deal jointly with these challenges.
- Another option is to design an adaptive face recognition system able to handling different variations based on some quality factors.
Proposal

In this work, we propose to deal with the low quality or missing information in face images, by incorporating the biometric sample quality into the representation.
In this work, we propose to deal with the low quality or missing information in face images, by incorporating the biometric sample quality into the representation.

The proposal is inspired on the fact that using local information has shown to be effective for deep learning face recognition.
• In this work, we propose to deal with the low quality or missing information in face images, by incorporating the biometric sample quality into the representation.

• The proposal is inspired on the fact that using local information has shown to be effective for deep learning face recognition.

• We propose to analyze the quality of different face regions to create an activation map and to extract deep local features from the activated areas.
In this work, we propose to deal with the low quality or missing information in face images, by incorporating the biometric sample quality into the representation. The proposal is inspired on the fact that using local information has shown to be effective for deep learning face recognition. We propose to analyze the quality of different face regions to create an activation map and to extract deep local features from the activated areas. The face parts which are not reliable are then discarded.
The face image is divided into 16 regions to make a local analysis. These regions are automatically obtained by making use of the facial landmarks.
The defined regions will be activated or not, depending on the quality information obtained from the face image.

For this aim, different quality measures are defined, taking into account the factors that have the most impact on the performance of face recognition systems: pose variations, facial expressions, occlusion, blurriness, and illumination.

Occlusion ($\Theta$), blurriness ($\beta$), and illumination ($\Phi$) are evaluated on every region, while Pose ($\Upsilon$) and facial expression ($\Xi$) are computed globally, and depending on the obtained values, the defined regions are activated or not.
Activation Vector

- \( \Lambda = \{\lambda_i | i = 1..16\} \), is the ordered activation vector for the entire face, where \( \lambda_i \) takes values of 0 and 1, indicating if the \( i^{th} \) region should be used for extracting features or not.

- This vector results from the “and” operation between all individual binary vectors that represent the defined quality measures. Thus, \( \Lambda = \beta \times \Theta \times \Phi \times \Upsilon \times \Xi \).

Some samples of different activated regions
Distance between two face samples:

\[ D(f_1, f_2) = \frac{\sum_{j=1}^{\alpha} d(\text{feat}_{1j} \ast \Lambda_{Cj}, \text{feat}_{2j} \ast \Lambda_{Cj})}{\alpha} \]  

where \( \alpha = |\Lambda_C| \) indicates the number of active regions and \( d(\text{feat}_{1j} \ast \Lambda_{Cj}, \text{feat}_{2j} \ast \Lambda_{Cj}) \) is the Euclidean distance between the two features vectors corresponding to the \( j \) region.

The proposed approach can be applied with any pre-trained CNN: Vgg-Face model, Dlib model, MobileFaceNet-Y2.
Databases and Scenarios

**LFW database**
- Contains 13,233 face images corresponding to 5,749 subjects.

**CFPW database**
- Composed of face images of celebrities in frontal and profile views. It contains 10 frontal and 4 profile images from 500 subjects.

**AR database**
- Contains around 4000 images from 126 subjects captured on two different sessions. Each person has up to 13 images per session with different expressions, illuminations and occlusions.
### Results with LFW database

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vgg-face</td>
<td>97.27 ± 0.25</td>
</tr>
<tr>
<td>Quality Map</td>
<td>98.54 ± 0.60</td>
</tr>
<tr>
<td>Dlib</td>
<td>99.38 ± 0.27</td>
</tr>
<tr>
<td>Quality Map</td>
<td>99.85 ± 0.20</td>
</tr>
<tr>
<td>MobileFaceNet-Y2</td>
<td>99.68 ± 0.09</td>
</tr>
<tr>
<td>Quality Map</td>
<td>99.90 ± 0.04</td>
</tr>
</tbody>
</table>

Méndez-Llanes et al

Quality-Based Face Representation
Results with CFPW database

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vgg-face</td>
<td>89.91 ± 0.51</td>
</tr>
<tr>
<td>Quality Map</td>
<td>92.74 ± 0.31</td>
</tr>
<tr>
<td>Dlib</td>
<td>87.65 ± 0.40</td>
</tr>
<tr>
<td>Quality Map</td>
<td>94.69 ± 1.10</td>
</tr>
<tr>
<td>MobileFaceNet-Y2</td>
<td>94.50 ± 0.11</td>
</tr>
<tr>
<td>Quality Map</td>
<td>96.05 ± 0.80</td>
</tr>
</tbody>
</table>
Results with AR database

<table>
<thead>
<tr>
<th>Model</th>
<th>Expression</th>
<th>Sunglasses</th>
<th>Scarf</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vgg-face</td>
<td>85.36</td>
<td>88.23</td>
<td>81.04</td>
</tr>
<tr>
<td>Quality Map</td>
<td><strong>93.65</strong></td>
<td><strong>95.42</strong></td>
<td><strong>90.74</strong></td>
</tr>
<tr>
<td>Dlib</td>
<td>96.74</td>
<td>97.35</td>
<td>94.09</td>
</tr>
<tr>
<td>Quality Map</td>
<td><strong>100.0</strong></td>
<td><strong>99.59</strong></td>
<td><strong>95.92</strong></td>
</tr>
<tr>
<td>MobileFaceNet-Y2</td>
<td>30.54</td>
<td>95.11</td>
<td>63.13</td>
</tr>
<tr>
<td>Quality Map</td>
<td><strong>61.71</strong></td>
<td><strong>98.77</strong></td>
<td><strong>82.48</strong></td>
</tr>
</tbody>
</table>
Conclusions

- An activation map based on the face quality is presented, in order to select the face regions that have enough identifying information to be used in the recognition stage.
Conclusions

- An activation map based on the face quality is presented, in order to select the face regions that have enough identifying information to be used in the recognition stage.

- An extensive experimental evaluation was conducted on three challenging public databases, with three different deep face models. The achieved results show that in all cases, the proposed quality-based local approach outperforms the original models.
Conclusions

- An activation map based on the face quality is presented, in order to select the face regions that have enough identifying information to be used in the recognition stage.

- An extensive experimental evaluation was conducted on three challenging public databases, with three different deep face models. The achieved results show that in all cases, the proposed quality-based local approach outperforms the original models.

- On unconstrained scenarios, it is better to represent the faces with less information that contains the higher identifying and discriminative information, rather than use the entire face that can have spurious information due to real-life quality problems.
Thanks for your attention

Questions?