FUSING SHAPE AND TEXTURE INFORMATION FOR FACIAL AGE ESTIMATION

Jiwen Lu and Yap-Peng Tan

School of Electrical and Electronic Engineering
Nanyang Technological University, Singapore

ABSTRACT

This paper presents a new human age estimation method by using multiple feature fusion via facial image analysis. Motivated by the fact that both shape and texture information of facial images can provide complementary information in characterizing human age, we propose fusing these two sources of information at the feature level by using canonical correlation analysis (CCA), a powerful and well-known tool that is well suitable for relating two sets of measurements, for enhanced facial age estimation. Then, we learn a multiple linear regression function to uncover the relation of the fused features and the ground-truth age values for age prediction. Experimental results are presented to demonstrate the efficacy of the proposed method.

Index Terms—Facial age estimation, information fusion, soft biometrics.

1. INTRODUCTION

A facial image conveys many significant human characteristics, such as identity, gender, expression, and age. While face and gender recognition [1–4] have been extensively studied in the computer vision and pattern recognition community, facial age estimation has not been well explored, and there have only been a few attempts in the literature [5–13] in recent years.

Existing methods for facial age estimation can be mainly classified into four categories: anthropometric model [5], aging pattern subspace [6, 7], aging manifold analysis [8], and age regression [9–13]. (i) The anthropometric model method applies the cranio-facial development theory and facial skin wrinkle analysis to create the aging model, whereby the changes of shape and texture patterns of facial images are measured to categorize a face into several age groups. Hence, this method is suitable for coarse age estimation rather than refined cases. (ii) The aging pattern subspace method models a sequence of facial images of each subject collected at different years old and learns a subspace to estimate the age of a test image by reconstructing the image using the learned subspace. This method needs the identity information of each facial image in the training set, which is usually not available in many practical age estimation systems. Hence, it is also not convenient for real-world applications. (iii) The aging manifold analysis method models an aging manifold by using a manifold learning algorithm to seek a lower-dimensional feature submanifold to characterize the aging structure of the training samples. This method can effectively uncover the aging manifold structure of facial images, however, it still needs to learn a regressor to establish the relation between the lower-dimensional features and the age information. (iv) The age regression method considers facial age estimation as a multiple linear regression problem and aims to seek an appropriate regression function to describe the relation between the facial image and age information. Due to the simplicity and effectiveness, the regression method has been widely used in many facial age estimation systems [9–13].

Most existing facial age estimation methods, however, usually utilize only the appearance features (texture information) of facial images for age estimation. Previous work in face modeling [14] has demonstrated that the shape information of facial images also plays an important role in human age estimation, especially for the youth persons. Motivated by this reason, we propose utilizing canonical correlation analysis (CCA) [15] to fuse both the shape and texture information from facial images to characterize human ages. To uncover the relation of the fused features and the ground-truth age values, we learn a multiple linear regression function with a quadratic model for age estimation. Experimental results on the widely used FG-NET face database [16] are presented to demonstrate the efficacy of the proposed method.

2. PROPOSED APPROACH

The flow chart of our proposed facial age estimation approach is shown in Fig. 1. We first extract the shape and texture features of each face sample, respectively, and then fuse them by the canonical correlation analysis (CCA) method to further exploit their correlation. Lastly, we perform age estimation by using a multiple linear regression model.

2.1. Feature Extraction

In order to model facial shape and texture, a statistical approach is applied to learn the way how these two types of information vary across ages.
For the shape feature, each face image was manually labeled with 68 landmark points, and the positions of which define the face shape were used for face modeling. Fig. 2(a) shows some face samples of one subject across different ages and the manually labeled landmarks, and Fig. 2(b) shows the corresponding shapes represented by the 68 points. Having acquired the 68 points, each face shape was described by a vector: 
\[
s = (x_1, x_2, \cdots, x_{68}, y_1, y_2, \cdots, y_{68}) \in \mathbb{R}^{136 \times 1}.
\]
Assume there are \(m\) face samples for modeling, we can generate \(m\) such vectors \(s_i\) \((i = 1, 2, \cdots, m)\) and group them into a \(136 \times m\) matrix for statistical shape analysis. Then, we applied Procrustes analysis to minimize the variations of face shape from factors such as angle and distance during photographing. Individual shape \(s_i\), was transformed into a common frame so that the sum of squared distances of the shape to the mean \(\sum_i |s_i - \bar{s}|^2\) is minimized, where \(\bar{s}\) is the mean shape the all the modeling samples. Fig. 2(c) shows the shapes after Procrustes analysis.

To build a statistical model for the facial texture features, all face images were wrapped usually to the mean shape such that the 68 feature points were exactly matched. We then performed wrapping through a piece-wise affine transformation. Fig. 3 shows some face texture images of one subject across ages from the FG-NET database. (a) Sample face images obtained by (a) Procrustes analysis and (b) wrapping, respectively.

\[
C_{xx} = E[xx^T], \quad C_{yy} = E[yy^T], \quad C_{xy} = E[xy^T].
\]
The projections of CCA can be easily obtained by solving the following generalized eigenvalue equation:

\[
\begin{bmatrix}
XX^T & 0 \\
0 & YY^T
\end{bmatrix}
\begin{bmatrix}
w_x \\
w_y
\end{bmatrix} = \lambda \begin{bmatrix}
XX^T \\
0
\end{bmatrix}
\begin{bmatrix}
w_x \\
w_y
\end{bmatrix}
\]

whence \(\lambda\) is the exact correlation between \(x\) and \(y\).

### 2.3. Age Regression

Let \(T = [t_1, t_2, \cdots, t_N]\) and \(L = [l_1, l_2, \cdots, l_N]\) be the training feature set and the corresponding age labels, where \(t_i \in \mathbb{R}^d, \ l_i \in \mathbb{R}^1, \ i = 1, 2, \cdots, N, \ d\) is the feature dimension of each facial image, and \(N\) is number of samples in the training set. The age regression method minimizes an objective function as [9, 11]:

\[
\min \sum_i ||l_i - \hat{l}_i||_2^2 \iff \min \sum_i ||l_i - f(t_i)||_2^2
\]
where \( \hat{l}_i = f(x_i) \) is the estimated age of the image \( x_i \), and \( f \) is the regression function to be sought. Previous work \([9, 11]\) has demonstrated that better age estimation performance can be usually obtained if \( f \) was selected to be a quadratic model (QM) rather than a linear model (LM) and a cubic model (CM). In such cases, \( f \) was defined as

\[
f(t_i) = \hat{a}_0 + \hat{a}_1 t_i + \hat{a}_2 t_i^2
\]

(4)

where \( \hat{a}_0 \) denotes the estimated intercept term, \( \hat{a}_1 \) and \( \hat{a}_2 \) the estimated parameter vectors, \( \hat{a}_1, \hat{a}_2 \in \mathbb{R}^d \), and \( t_i^2 \) the elementwise square of \( t_i \).

Let \( \hat{A} = [\hat{a}_0 \hat{a}_1^{(1)} \cdots \hat{a}_d^{(1)} \hat{a}_2^{(1)} \cdots \hat{a}_d^{(d)}]^T \) and \( \hat{T} = [\hat{t}_1 \hat{t}_2 \cdots \hat{t}_N]^T \). By using the well-known ordinary least square solution \([8, 9]\), the parameters \( \hat{a}_0, \hat{a}_1 \) and \( \hat{a}_2 \) in \( \hat{A} \) can be obtained as

\[
\hat{A} = (\hat{T}^T \hat{T})^{-1} \hat{T}^T \hat{L}
\]

(5)

Having obtained \( \hat{A} \), we can easily estimate the age value of the new test sample \( t \) by using Eq. (4).

3. EXPERIMENTAL RESULTS

In this section, we evaluated the effectiveness of our proposed method for facial age estimation. We have used the FG-NET \([16]\) face databases to evaluate the efficacy of the proposed approach for facial age estimation. The FG-NET database consists of 1002 face images with large variation of lighting, pose, and expression. There are 82 subjects in total with ages ranging from 0 to 69. Some sample images of two subjects, each from one database, are shown in Fig. 4.

Fig. 4. Sample face images with different age values from the FG-NET database, and the number below each image is the corresponding age value.

In our experiments, two measures were used to evaluate the performance of the proposed approach. The first one is the mean absolute error (MAE) criterion \([6–13]\), which is defined as the average of the absolute errors between the estimated and the ground truth age values:

\[
MAE = \frac{1}{N_t} \sum_{i=1}^{N_t} |\hat{A}_i - A_i|
\]

(6)

where \( N_t \) is the number of the test sample, and \( \hat{A}_i \) and \( A_i \) are the estimated and ground truth age values for the \( i \)th test sample, and \( \hat{T} = \frac{\hat{T}}{100} \times N_t \)

Fig. 5. Cumulative scores of the age estimation at error levels form 0 to 10 years old on the FG-NET database. Results obtained when (a) PCA and (b) LDA were applied, respectively.

\[
\text{CumScore}(\theta) = \frac{N_{e\leq\theta}}{N_t} \times 100\%
\]

(7)

where \( N_{e\leq\theta} \) is the number of test face samples on which the estimation makes an absolute error less than \( \theta \) (years old).

We compared our feature fusion method with three other methods: shape only (SO), texture only (TO), direct combination of shape and texture (DC). Since it is usually not stable to perform regression for high-dimensional data, we performed principal component analysis (PCA) and linear discriminant analysis (LDA) on the extracted features before age estimation. Table 1 tabulates the MAEs of the methods, and Fig. 4 shows the cumulative scores. We also compared our approach with some state-of-the-art facial age estimation methods and tabulated the results obtained on the FE-NET database in Table 2.

We can make the following three observations from the above experimental results:

1) Exploiting both shape and texture information consistently outperform the methods which use only shape or texture feature for facial age estimation, which implies that multiple feature fusion is better than single feature-based methods for facial age estimation.

2) Our proposed approach performs better that the method which direct combines the shape and texture features for facial age estimation. The reason is that CCA can better exploit the correlation of shape and texture information in the fusion procedure.

3) Our proposed approach is comparative to state-of-the-art facial age estimation algorithms in terms of the age estimation performance and it is much easier to be implemented than these existing methods.
Table 1. Comparison on MAE (years old) using different feature representations.

<table>
<thead>
<tr>
<th>Method</th>
<th>PCA</th>
<th>LDA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Range</td>
<td>SO</td>
<td>TO</td>
</tr>
<tr>
<td>0-9</td>
<td>5.26</td>
<td>5.11</td>
</tr>
<tr>
<td>10-19</td>
<td>5.24</td>
<td>5.08</td>
</tr>
<tr>
<td>20-29</td>
<td>5.20</td>
<td>6.13</td>
</tr>
<tr>
<td>30-39</td>
<td>10.29</td>
<td>10.59</td>
</tr>
<tr>
<td>40-49</td>
<td>16.79</td>
<td>18.32</td>
</tr>
<tr>
<td>50-59</td>
<td>27.87</td>
<td>30.18</td>
</tr>
<tr>
<td>60-69</td>
<td>33.81</td>
<td>41.18</td>
</tr>
<tr>
<td>Average</td>
<td>6.73</td>
<td>6.95</td>
</tr>
</tbody>
</table>

Table 2. MAE comparison with state-of-the-art facial age estimation methods on the FG-NET database

<table>
<thead>
<tr>
<th>Method</th>
<th>MAE year</th>
</tr>
</thead>
<tbody>
<tr>
<td>RUN [10]</td>
<td>5.78 2008</td>
</tr>
<tr>
<td>MISS [17]</td>
<td>5.36 2009</td>
</tr>
<tr>
<td>AO graph [18]</td>
<td>5.97 2010</td>
</tr>
<tr>
<td>IIS-LLD [19]</td>
<td>5.77 2010</td>
</tr>
<tr>
<td>Ours (this paper)</td>
<td>5.75 /</td>
</tr>
</tbody>
</table>

4. CONCLUSION AND FUTURE WORK

We have proposed in this paper a new human age estimation method by using multiple feature fusion via facial image analysis. Since the shape and texture information from a same facial image can provide complementary information in characterizing human ages, we have fused them by the canonical correlation analysis (CCA) method to better characterize their correlation. To uncover the relation of the fused features and the ground-truth age values, we have used a multiple linear regression function with a quadratic model for age estimation. Experimental results have demonstrated the efficacy of the proposed method. Some interesting future directions of this work include: 1) to apply an automatic shape localization method to extract the shape information of facial images; 2) to design more elaborate multiple feature fusion algorithms to better improve the age estimation performance.

5. ACKNOWLEDGEMENT

The authors would like to thank Mr. Xue Qi for help us conducting partial experiments in this paper.

6. REFERENCES