A Robust Feature-processing Method for Age-invariant Face Recognition

Xiaonan Hou¹,²*, Xuetao Qiu², Sishuang Wan² and Tao Tang¹,²

¹Postdoctoral Science Research workstation of China UnionPay, Shanghai, China
²China Unionpay, Shanghai, China

*houxiaonan@unionpay.com

Abstract. In this paper, we present a novel feature-processing method for age-invariant face recognition so that it is robust to aging process. The main purpose of feature-processing is to remove aging effects while keep personalized properties stable simultaneously. In order to achieve this, we try to learn a space map and then encode the mapped feature to an age-invariant representation. In the encoding step, we introduce two kinds of constraints: the temporal constraint (local constraint) and boundary constraint (global constraint). We applied our feature-processing method to Cross-Age Celebrity Dataset (CACD). In order to verify the versatility of our method, we apply it to both high-dimensional LBP feature and deep feature. Results show that our feature-processing method works well on CACD and the face verification subset of CACD (CACD-VS).

1. Introduction

In face recognition, pose, illumination, expression and aging are four key factors affecting the accuracy. In recent days, due to the significant improvement of face detection, alignment, and using of deep feature, face verification accuracies have outperformed human performance on face verification benchmark such as Labelled Faces in the Wild dataset (LFW). However, LFW contains little age variation. When dealing with faces under aging changes, most existing methods still face great challenges. The challenge of this problem, to a great extent, arises from the fact that the feature of a person face is obviously changed by the aging process. Examples of face aging is shown in figure 1.

The rest of this paper is organized as follows. Section 2 gives the detail of our proposed method. We specifically presented the experimental results in section 3. Section 4 gives the conclusion.
2. The Proposed Method

Figure 1. Examples of CACD dataset (Chen et al., 2014). Each column corresponds to one single person’s images of year. 2004, 2008 and 2013.

There are mainly two steps in our feature-processing method: mapping and encoding. We first learn a robust feature map in order to remove the aging effects in features; And in the encoding step, we induced a temporal constraint and a boundary constraint, which is proved helpful to improve the recognition accuracy. The overview of our method is shown in figure 2.

Figure 2. Overview of the proposed method.

2.1. Robust Feature Mapping

In this part, we will focus on learning a feature map which is to remove effects of face aging. To achieve this, we consider intra-personal similarity of the whole age range. Next is the detail.

Figure 3. Take the equal interval \( l \) as 3, which means dividing the age range into 3 intervals.
Denote the $ith$ reference person’s feature set as $X^{(i)}$ and the age range as $T$. We divide $T$ into $l$ equal periods: $[T_1, T_2, ..., T_l]$. We assume that all features in $X^{(i)}$ are ordered by time and are divided into $[X_1^{(i)}, ..., X_l^{(i)}]$ (see figure 3) according to time period $T_t$, $t = 1, ..., l$. Assuming $\mu^{(i)}_t$ as the average feature of $X_t^{(i)}$, we can get the average features of the $ith$ reference in all age periods: $\mu^{(i)} = [\mu^{(i)}_1, ..., \mu^{(i)}_l]$. Averaging the feature vectors can save computing time and remove some noise. Considering the fact that prominent face changes usually come, after all other confounding factors have been eliminated, from large age variation, we need to learn a space map $M \in \mathbb{R}^{d \times d}$ in order to make average features in $\mu^{(i)}$ as close as possible. We denote $d_w$ as the space dimension we learn and $d$ as the dimension of the input feature. Denoting $K_i = [\mu^{(i)}_1, ..., \mu^{(i)}_n]$, $n$ is the size of the reference set, we have the objective function below to solve $M$:

$$
\min_M \sum_{i=1}^{l} \sum_{t=1}^{T} \| M (K_i - K_{t,j}) \|_F^2 + \sigma \| M - I_0 \|_F^2
$$

(1)

$\| M (K_i - K_{t,j}) \|_F^2$ is used to make the processed features of the same subject in all age periods as close as possible. We can see that (1) is easily to be solved and have a closed-form solution of $M$:

$$
M = \sum_{i=1}^{l} \sum_{t=1}^{T} ((K_i - K_{t,j})(K_i - K_{t,j})^T + \sigma I)^{-1}(I)
$$

(2)

Thus, we only need to do a linear computation to remove the aging effects of the original feature and therefore it is highly scalable. After solving $M$ for each landmark, we can map all the original feature vectors to a subspace that is robust to face aging.

2.2. Modified Feature Encoding

In order to encode the mapped feature, we combine the temporal constraint which is local and boundary constraint which is global. There are two main steps in this procedure: 1) computing reference set representations; 2) encoding local feature vector into reference space.

2.2.1. Reference Set Representation. Denote $n$ as the size of reference set, $m$ as the number of years. $Z_j^{(i)}$ is the average of the feature vectors of the same reference person from the same year. The reference set in year $j$ is $Z^{(i)} = [Z_1^{(i)}, ..., Z_j^{(i)}]$ and the whole reference set in a span of $m$ years is $Z = [Z^{(1)}, ..., Z^{(m)}]$.

2.2.2. Encoding Mapped Feature. In this step, we want to encode $w$ with the reference set $(Z_j^{(i)})_j$ to get the new feature vector $A = [\alpha^{(1)}, ..., \alpha^{(m)}]$, $m$ is the number of years, and $\alpha^{(i)}$ is the encoded feature. Then we can obtain $A$ by solving the optimization problem below:

$$
\min_A \sum_{j=1}^{m} \| w - Z^{(i)} \alpha^{(i)} \|_F^2 + \lambda \| \alpha^{(i)} \|^2 + \lambda \| \alpha^{(i)} \|^2 + \lambda \| LA^T \|^2 + \| BA^T \|^2
$$

(3)

There are three regularization terms in this optimization problem. The first one is Tikhonov Regularization. In order to keep the identity permanence during aging, we use the second regularization to satisfy temporal constraint that one person is similar to a reference person at year $j$, he/she is most likely similar to the same reference person in the former year $j-1$ and the latter year $j+1$. Thus, the smooth operator $L \in \mathbb{R}^{(m-2)\times m}$ is defined as:
\[
L = \begin{bmatrix}
1 & -2 & 1 & 0 & \ldots & 0 & 0 & 0 \\
0 & 1 & -2 & 1 & \ldots & 0 & 0 & 0 \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
0 & 0 & 0 & 0 & \ldots & 1 & -2 & 1
\end{bmatrix}
\]  

(4)

Besides, we assume that the changes caused by age-variance is approximately monotonous. Thus, given the encoded feature of a face, the value corresponding to the \(j\)th person in the reference set is \([\alpha_{j}^{(0)}, \ldots, \alpha_{j}^{(m)}]\). Without losing generality, we suppose the value \(\alpha_{j}^{(0)}\) corresponding to age \(i_{0}\) is the maximum value, then the value of other years will decrease progressively, which means that one \(j\) should have only one extreme point while the bad distribution may have several. Thus, we define \(B\) as:

\[B = [1, 0, \ldots, 0, -1] \in \mathbb{R}^{1 \times m}\]

(5)

Denote:

\[
W = \begin{bmatrix}
Z^{(1)} & 0 & \ldots & 0 \\
0 & Z^{(2)} & \ldots & 0 \\
0 & 0 & \ldots & Z^{(m)}
\end{bmatrix}
\]

(6)

\[
L = \begin{bmatrix}
1 & -21 & 1 & 0 & \ldots & 0 & 0 & 0 \\
0 & 1 & -21 & 1 & \ldots & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \ldots & 1 & -21 & 1
\end{bmatrix}
\]

(7)

\[\hat{B} = [1, 0, \ldots, 0, -1]\]

(8)

Here \(\hat{\alpha} = [\alpha^{(1)^T}, \ldots, \alpha^{(m)^T}]^T\) \(\hat{w} = [w^T, \ldots, w^T]^T\). Then we update (3) into:

\[
\min_{\hat{\alpha}} \sum_{j=1}^{m} \| \hat{w} - W \hat{\alpha} \|^2 + \lambda_1 \| \hat{\alpha} \| + \lambda_2 (\| \hat{L} \hat{\alpha} \|^2 - \| \hat{B} \hat{\alpha} \|^2)
\]

(9)

We can easily get the closed-form solution:

\[
\hat{\alpha} = (W^T W + \lambda_1 I + \lambda_2 (\hat{L}^T \hat{L} - \| \hat{B} \|^2))^{-1}
\]

(10)

2.2.3. Aggregating Representation Across Different Years. In this step, we use Max Pooling:

\[
\alpha_i = [\alpha_i^{(1)}, \ldots, \alpha_i^{(m)}], \forall m
\]

(11)

to aggregate representation in reference space. Thus, the final representation will have a high response to one reference person as long as it has a high response to this person in any year.

3. Experiments

![Figure 4](image-url) The curve of several age-invariant face recognition results.
3.1. Experiment on CACD Dataset
CACD contains 163,446 images of 2000 celebrities. Its age gap is 0~10. We strictly follow the settings of [1].

Our experiments are conducted on three different subsets. In all three subsets, query images are taken in 2013. These three subsets are taken in 2004∼2006, 2007∼2009, 2010∼2012 respectively. The comparison results are reported in table 1. When using HD-LBP feature, our method achieves 57.9% in the subset taken in 2004∼2006, 60.7% in 2007∼2009, and 65.7% in 2010∼2012. When using deep feature, our method achieves 68.1%, 71.5% and 76.7% respectively. As shown in figure 4, we can see that even using the same HD-LBP feature, our approach outperforms CARC for more than 4%.

Besides, as shown in table 1, Cosine Metric(Deep feature) gives better results than CARC (Deep feature), which means that not only CARC not make full use of feature’s deep information, but ends up influencing deep feature in a negative way. CARC cannot make full use of the high representativeness of deep feature. In contrast, our method (Deep feature) improves Cosine Metric (Deep feature) by around 4% in every age-range, and about 6% higher than CARC (Deep feature). This proves that our method is robust to both shallow and deep features. In our opinion, the reason for such dramatically improvement is the robust feature mapping step, which is the key step in our model. This step focuses on removing the great intra-variance caused by aging process. After this step, features of the same identity are much more alike than the original features, which makes the encoding step easier and more accurate.

3.2. Experiment on CACD-VS Dataset
CACD-VS is the verification subset of CACD. There are 2000 positive pairs and 2000 negative pairs in it. Unlike CACD which has some wrong annotations, this subset is carefully annotated by checking both of the associated image and web contents. To conduct face verification experiment on CACD-VS, we divide it into 10 folds. Each fold contains 200 positive pairs and 200 negative pairs. Subjects in each folds are exclusive.

Table 1. The comparison of our method with several state-of-the-art methods on CACD dataset.

| Method | Year Interval | 2004~2006 | 2007~2009 | 2010~2012 |
|--------|---------------|-----------|-----------|-----------|
| HD-LBP [2] | 36.6% | 38.9% | 44.0% |
| HFA [3] | 50.2% | 52.0% | 57.5% |
| CARC [1](LBP feature) | 52.9% | 55.5% | 61.1% |
| Ours(LBP feature) | 52.9% | 55.5% | 61.1% |
| CARC [1](Deep feature) | 61.7% | 65.1% | 71.6% |
| Cosine (Deep feature) | 63.6% | 66.8% | 73.1% |
| Ours(LBP feature) | 52.9% | 55.5% | 61.1% |
| Ours(Deep feature) | 68.1% | 71.5% | 76.7% |

We compare our method with high dimensional CARC [1], LBP [2], HFA [3], and DeepID [4] et al. The parameters used in this part is the same with experiments on CACD. From table 2 we can see that our method’s face verification accuracy is 91.2% when using HD-LBP feature, which outperforms other methods by a clear margin. We should note that, even using shallow feature, our method performs better than several deep learning methods among those compared ones, which use much more training data than our method does. The results of high dimensional LBP [2], HFA [3], Human Average, and Human Voted are provided by [1]. Furthermore, we apply our method to deep feature. Surprisingly, the mean accuracy is 95.9% . We also compare our method with human performance. From table 2 we know that the average human performance is 85.7% , which is much lower than ours. By combining the decisions from 9 users, the human voting accuracy achieves
94.2%. The ROC curves are shown in figure 5. We can see that our method’s performance is competitive with human voting and much higher than other compared methods.

Table 2. The face verification accuracy comparison on CACD dataset.

| Method                | Accuracy |
|-----------------------|----------|
| HD-LBP [2]            | 81.6%    |
| HFA [3]               | 84.4%    |
| Deepface [5]          | 85.4%    |
| DeepID [4]            | 87.2%    |
| GSM [6]               | 89.8%    |
| Ours(LBP feature)     | 91.2%    |
| Ours(Deep feature)    | 95.9%    |
| Human, Average [7]    | 85.7%    |
| Human, Voted [7]      | 94.2%    |

4. Conclusions
We proposed a feature-processing method which is to remove aging effects while keep personalized properties stable simultaneously. We divide our method into two steps: encoding and mapping, which makes the optimization problem much easier to solve; besides, we modify the encoding step by using two constraints: local constraint and global constraint. Our experiments were conducted on CACD and MORPH dataset. We use both shallow and deep feature to verify the versatility of our method. Our method is concise and effective and outperforms many state-of-the-art methods on CACD. Our performance on face verification on CACD-VS dataset is even higher than human voting.

Figure 5. ROC curves of state-of-the-art methods.

References
[1] Chen, B.C., Chen, C.S., Hsu, W.H. Cross-age reference coding for age-invariant face recognition and retrieval. *European 2014 Conference on Computer Vision* pp 768–783
[2] Barkan, O., Weill, J., Wolf, L., Aronowitz, H. Fast high dimensional vector multiplication face recognition 2013 *Proceedings of the IEEE International Conference on Computer Vision* pp 1960–67
[3] Gong, D., Li, Z., Lin, D., Liu, J., Tang, X.: Hidden factor analysis for age invariant face recognition
2013 Proceedings of the IEEE International Conference on Computer Vision pp 2872–79

[4] Sun, Y., Chen, Y., Wang, X., Tang, X.: Deep learning face representation by joint identification-
verification 2014 Advances in neural information processing systems pp 1988–96

[5] Taigman, Y., Yang, M., Ranzato, M., Wolf, L. Deepface: Closing the gap to human-level performance
in face verification. 2014 Proceedings of the IEEE Conference on Computer Vision and Pattern
Recognition pp 1701–08

[6] Lin, L., Wang, G., Zuo, W., Xiangchu, F., Zhang, L. Cross-domain visual matching via generalized
similarity measure and feature learning 2016 IEEE transactions on pattern analysis and machine
intelligence

[7] Chen, B.C., Chen, C.S., Hsu, W.H.: Face recognition and retrieval using cross-age reference coding
with cross-age celebrity dataset 2015 IEEE Transactions on Multimedia 17 pp 804–815