CROSS-AGE FACE RECOGNITION USING REFERENCE CODING WITH KERNEL DIRECT DISCRIMINANT ANALYSIS

Haoshan Zou, Haifeng Hu

School of Electronics and Information Technology, Sun Yat-Sen University

ABSTRACT

While face recognition methods have received wide application for decades, the aging process on human face could disable the original method. In this paper, we present a modification on a cross-age face recognition model which utilizes a reference set arranged in time order to eliminate the age difference of input images. In the proposed method, we utilize the identity information of the gallery set to perform discriminative analysis, which can further discriminate between persons in a subspace after the reference coding. Compared with classical cross-age reference coding method, our experiments on Cross-age Celebrity Dataset (CACD) acquire a 35% improvement in recognition rate.

1. INTRODUCTION

In the recent decades, more and more researchers working on computer vision have set their eyes on cross-age face recognition. Since current works involving faces as well as the aging process have mostly focused on facial age estimation or the description of facial aging process, they can be the basis for further cross-age recognition work. Hadchum et al. [1] present age estimation by observing wrinkles and skin color using a hybrid of Support Vector Machines and Fuzzy Logic. Gong et al. [2] propose a maximum entropy feature descriptor (MEFD) to encode facial images by embodying maximum entropy, and then by densely sampling the encoded face image, the desirable discriminatory information is obtained. Liu et al. [3] exert deep learning for cross-age face verification under large age gaps. They trained an autoencoder to synthesize faces in four age stages for the two input faces of different age, and then a trained CNN could tell whether the synthesized faces of each same stage belongs to the same person. Bouchaffra [4] introduces nonlinear topological component analysis which nonlinearly maps the data manifold to a low-dimensional latent variable space, extracts topological features through these latent variables, and performs shape classification. Li et al. [5] propose a two-level learning hierarchical model. At the first level they extract feature on their proposed local pattern selection feature descriptor that minimizes intra-user dissimilarity, and then at the second level, higher level visual information is further refined.

Corresponding Author: Haifeng Hu

Recently, Chen et al. propose a data-driven method called Cross-age Reference Coding (CARC) [6], which intends to eliminate the difference between young and old face images by encoding them into a reference set space that embodies the face difference between ages. The basic assumption for the method is that if two people look alike when they are young, they might also look similar when they both grow older. By taking advantage of the similarity, we identify images that have similar relations to the reference set to be the same person. Because CARC is intended for face retrieval, only the reference set images make use of the age and identity information, but the gallery and probe sets for testing are randomly provided without any information. In real application scenario, face verification devices such as face locks and attendance machines acquire the gallery face image sets with identity labels. Therefore, we present a new reference coding model with the extra information about identity. In our method, the age-invariant features of gallery images are used as training data to obtain an optimal subspace. In this way, by projecting both the gallery and the probe into the subspace, we should be able to maximize the difference between persons in the subspace and thereby enhance performance in the subsequent cosine similarity scoring. Considering the nonlinear and complex distribution of the large number of aging face images under a perceivable variation in age, we adopt the classical Kernel Direct Discriminant Analysis (KDDA) [7, 8, 9, 10] for exploiting the labeling information and maximizing the difference between persons in the subspace.

2. THE PROPOSED APPROACH

Our combined method as shown in Fig.1 is to perform CARC in the first stage, then in the second stage we utilize the identity information of the gallery set to implement KDDA. By implementing KDDA, we derive a projection matrix to further distinguish between test persons when applied on the CARC coded gallery and probe.

2.1. Cross-Age Reference Coding

As shown in Fig.2, the CARC method mainly involves three steps:

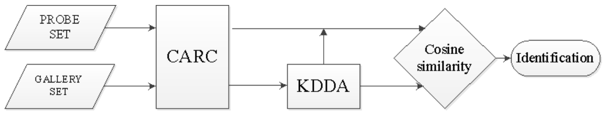

Fig. 1. The overall process of our proposed model.

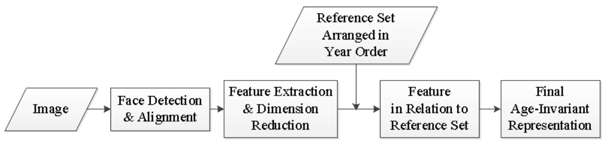

Fig. 2. The general flowchart of the reference coding.

1) Compute from a training set of age-varying face images of n people and m years to construct a reference set.

2) Exploit the reference set to encode the test image feature.

3) Aggregating the feature in step 2 to obtain an age-invariant representation.

Before implementing CARC steps, our pre-processing stage has detected $p = 16$ landmarks locating around eyebrows, eyes, nose tips and upper lip, as shown in Fig. 3. At each landmark position we extract a 4720-dimension HD-LBP[13] feature and cascade all 16 features, obtaining a 75520-dimension vector to describe an input face x. For the ease of calculation, we perform PCA for the feature vectors. For example, the dimension is reduced to $d = 500$ at every landmark position k, so we have a 500*16=8000-dimension pca feature at the beginning of our CARC steps. We calculate CARC features for each landmark feature individually, thus the input part $x^{(k)}$ for each $C^{(k)}$ round would be in 500-dimension and we are to have 16 rounds of calculation.

In step 1, we build the reference set by arranging images of the n reference persons into m years, where at person i, year j, landmark k the reference set $C_i^{(j,k)}$ is a $d \times n$ matrix. If there exist several images for one single year, the mean value of the image features is calculated to be the representation vector,

$$C_i^{(j,k)} = \frac{1}{N_{ij}} \sum_{\substack{identity(x^{(k)})=i \\ year(x^{(k)})=j}} x^{(k)}, \qquad (1)$$
$$\forall i = 1, ..., n; j = 1, ..., m; k = 1, ..., p.$$

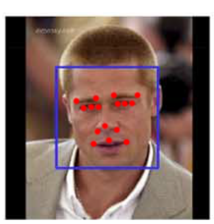

Fig. 3. The detection of 16 facial landmarks.

In step 2, the process of encoding feature into the reference space is by solving a least-square problem with Tikhonov regularization

$$\min_{\alpha^{(j,k)}} \|x^{(k)} - C^{(j,k)}\alpha^{(j,k)}\|^2 + \lambda\|\alpha^{(j,k)}\|^2, \forall j, k, \qquad (2)$$

where the $\alpha^{(j,k)} \in R^{n\times 1}$ we desire to obtain represents the relation of the input image to the n reference people in year j at landmark k. With this regulation, the relation α should be large if the input x is close to some $i - th$ reference people, and small if not.

Moreover, we would imply a temporal constraint on the regulation based on the intuition that if a input person is similar to the $i - th$ reference person at year j, he should still be more or less similar to the same reference person at year $j-1$ or $j+1$. Therefore, we have

$$L = \begin{bmatrix} 1 & -2 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & -2 & 1 \end{bmatrix} \in R^{(m-2)\times m}, \qquad (3)$$

the L smoothness operator for the temporal consistency that minimizes the difference in the adjacent years.

In all, the regularization becomes

$$\min_{A^{(k)}} \sum_{j=1}^{m} (\|x^{(k)} - C^{(j,k)}\alpha^{(j,k)}\|^2 + \lambda_1\|\alpha^{(j,k)}\|^2) + \lambda_2\|LA^{(k)\mathrm{T}}\|^2, \forall k, \qquad (4)$$

where $A^{(k)} = [\alpha^{(1,k)}, \alpha^{(2,k)}, ..., \alpha^{(m,k)}], \forall k$.

It has an analytic solution

$$\hat{\alpha}^{(k)} = (\hat{C}^{(k)\mathrm{T}}\hat{C}^{(k)} + \hat{\lambda}_1 I + \hat{\lambda}_2 \hat{L}^{\mathrm{T}}\hat{L})^{-1}\hat{C}^{(k)\mathrm{T}}\hat{x}^{(k)}, \forall k. \qquad (5)$$

In stage 3, the relation α is aggregated through m years to obtain one single representation for the young and old faces of one input person. Such representations would desirably be age-invariant, i.e., they do not vary with time. The aggregation is accomplished with max-pooling, so if any input person has a high response to certain reference persons in any year, they would be aggregated into the representation that indicates the same identity, and thus the age-invariant representation is obtained.

2.2. KDDA

To begin the KDDA stage, we have a set of N gallery representations $\{\alpha_m\}_{m=1}^{L}$ available. Each representation is a vector of length $L = 500 \times 16 = 8000$. Our object is to obtain an optimal projection $y_m = \phi(\alpha_m)$ based on some certain separability criteria, which produces a mapping that results in an enhanced separability of different face subjects.

While the original LDA criterion is to minimize within-class distance as well as maximize between-class distance, the modified criterion that solves for singular within-class conditions for KDDA[10] can be expressed as

$$\Psi = arg\,\max_{\Psi} \frac{|(\Psi^{\mathrm{T}} S_{BTW} \Psi)|}{|(\Psi^{\mathrm{T}} S_{BTW} \Psi)| + |(\Psi^{\mathrm{T}} S_{WTH} \Psi)|} \ . \quad (6)$$

Meanwhile, in the representation of S_{BTW} and S_{WTH} for nonlinear conditions, we adopt a RBF kernel function

$$k(\alpha_m, \alpha_n) = \phi(\alpha_m) \cdot \phi(\alpha_n) = exp((-\|\alpha_m - \alpha_n\|^2)/\sigma^2) \ . \quad (7)$$

For more details about KDDA implementation, refer to the original paper.

After obtaining the final representation, cosine similarity is adopted for similarity score calculation.

3. EXPERIMENTS

FG-Net and MORPH[14] are by far the most popular aging face datasets, but FG-NET only contains 1,002 images of 82 people which is not enough for our data-driven method. MORPH contains 52099 images of 12936 people. The author of the CARC coding framework collected from the Internet the photos of 2,000 celebrities within 10 years, obtaining a dataset of 163,446 images with age ranging from 14 to 63, called Cross-age Celebrity Dataset (CACD).

3.1. Experiments on CACD

We followed the experimental settings in the original paper[6]. The CACD dataset is divided into two parts. Part 1 containing images of 200 celebrities was manually annotated their name and age. The other 1800 celebrities in part 2 was automatically annotated, allowing the possible existence of noise images that does not match its name or age. We select 120 celebrities from part 1 for testing, among which images taken in 2013 serve as query images and images taken in other years (2004-2012) serve as database images. Images in part 2 are either used to calculating the PCA subspace or as the reference set. In addition, we have the PCA feature dimensions d=500, the regularization parameters $\lambda_1 = 10, \lambda_2 = 10^4$, and the number of reference celebrities n=600.

Evaluation metrics include mean average precision (MAP) and precision at first place. The MAP metric is widely applied in retrieval tasks. For the retrieval results of each query image, AP is the average of precision at every recall level. MAP is then computed by averaging the APs of all queries. More specifically, let $q_j \in Q$ be the query images, and the positive images (images of the same person as query) in the database for q_j are $\{I_1, I_2, \ldots, I_{m_j}\}$. We rank the retrieval results of q_j in a descending order, and let R_{jk} be the ranked retrieval results from the top to the image I_k. The MAP is calculated as

$$MAP(Q) = \frac{1}{|Q|}\sum_{j=1}^{|Q|}\frac{1}{m_j}\sum_{k=1}^{m_j} Precision(R_{jk}) \quad (8)$$

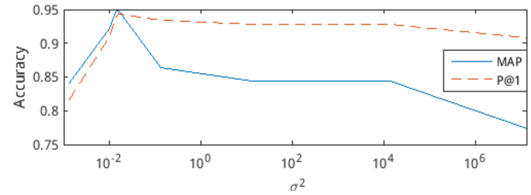

Fig. 4. Selection of RBF kernel variance σ^2.

Table 1. Comparison of accuracy between three discriminant analysis methods.

Method	MAP	P@1
HD-LBP	43.96%	85.46%
CARC	61.11%	92.15%
CARC+DCC	65.07%	92.69%
CARC+LDA	89.68%	94.92%
CARC+KDDA	96.70%	94.46%

Parameter Selection: In the RBF kernel function $k(\alpha_1, \alpha_2) = exp((-\|\alpha_1 - \alpha_2\|^2)/\sigma^2)$, we can set different variance value σ^2. After trying multiple values as shown in Fig. 4, we determine a best-performing $\sigma^2 = 1.6e - 2$ and fix it for the subsequent experiments.

Results and Discussion: We compare the results of our kdda-CARC to the original CARC and HD-LBP features, as well as other discriminant analysis methods such as LDA and DCC[15]. For the ease of comparison, the following results are all based on the 2010-2012 gallery set.

From Table 1, CARC+KDDA have achieved a remarkably high MAP reaching 96.70%, much higher than 61.11% when the label information in gallery set is not given. The rank-1 precision has also improved slightly from 92.15% to 94.46%.

However, applying KDDA demands extra training time, in this case, 729 seconds. Besides, to project the gallery and the probe into the subspace together consumes 493 seconds. The final evaluation stage has accelerated because of lower dimension features, but not significantly.

Fig. 5 shows a retrieval example of multiple methods. Our method that incorporates label information and KDDA

Fig. 5. The top-10 retrieval result of the query image. The red box indicates false result, and numbers in the parenthesis are APs.

outperforms the previous methods.

3.2. STASM

Before going on to implement CARC on MORPH, we adjust the pre-processing stage of Chen et al.'s cross-age face recognition system. Their pre-processing stage includes face detection, landmark detection and face alignment. They employ Viola-Jones face detector for face detection, IntraFace by CMU Human Sensing Laboratory for landmark detection and the line segment connecting two eye balls for face alignment. While IntraFace is a robust landmark detector that works well with facial expressions and angle variation, their source code is under patent protection and not publicly accessible. Therefore we select another landmark detector with similar performance but easier accessibility.

STASM is a c++ packet that performs face detection as well as landmark detection. They incorporate SIFT descriptors into classical Active Shape Model (ASM) for face detection and use Multivariate Adaptive Regression Splines (MARS) to match with the descriptor. Though it is intended for frontal face with neural facial expression, the left-out images would not affect our data-driven CARC.

For CACD dataset, in the 163446 face images that IntraFace manages to recognize the landmarks, STASM only failed to recognize 746 of them, reaching a success rate of 99.54%.

For Morph dataset, in 52099 face images STASM recognize 51861, reaching a success rate of 99.54%.

3.3. Experiments on MORPH

We follow the experimental setup in the original paper[6]. They perform PCA and LDA to reduce the feature dimension to 1001 and 1000 before CARC. 10,000 subjects are randomly selected from the MORPH dataset, among which the youngest are used as the gallery set and the oldest are used as the probe set. Because not many people have multiple images in one year in MORPH, we directly use 600 persons with images in any 6 years as the reference set. The other remaining subjects are used for building PCA and LDA subspaces. In this part of the experiment, the original paper has exploited the label information of the set of images used for building subspaces. At this stage, we can achieve the MAP of 92.40% and P@1 of 90.36%.

After CARC coding, we perform KDDA using the label information of the gallery set. The final performance has decreased to the MAP of 88.07% and the P@1 of 85.75%. Since some label information of the MORPH dataset is already used in previous steps, it is reasonable that it does not make improvement. Moreover, the gallery set contains only one image for the 10,000 subjects and thereby KDDA is unable to utilize within class information. Meanwhile, it takes a large amount of time to find the optimal projection matrix for the gallery set with 10,000 classes.

4. CONCLUSIONS AND FUTURE WORK

In this paper, we address the cross-age face recognition problem by combining the CARC method with a kernel discriminant analysis model. Our work has focus on a specific application scenario where the gallery set has labeled identity. By taking advantage of this extra information, our method can improve the recognition rate compared to the CARC approach, which is verified by the experimental results on the CACD dataset.

5. ACKNOWLEDGEMENT

This work was supported in part by the NSFC, under Grant no. 61673402, the NSF of Guangdong (grant nos. 2014A030313173 and 2016B010109002), the Science and Technology Program of Guangzhou, China (grant no. 201704020180) and the Fundamental Research Funds for the Central Universities of China.

6. REFERENCES

[1] Phittaya Hadchum and Sartra Wongthanavasu, “Facial age estimation using a hybrid of svm and fuzzy logic,” in *Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2015 12th International Conference on*. IEEE, 2015, pp. 1–6.

[2] Dihong Gong, Zhifeng Li, Dacheng Tao, Jianzhuang Liu, and Xuelong Li, “A maximum entropy feature descriptor for age invariant face recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 5289–5297.

[3] Luoqi Liu, Chao Xiong, Hanwang Zhang, Zhiheng Niu, Meng Wang, and Shuicheng Yan, “Deep aging face verification with large gaps,” *IEEE Transactions on Multimedia*, vol. 18, no. 1, pp. 64–75, 2016.

[4] D. Bouchaffra, “Nonlinear topological component analysis: Application to age-invariant face recognition,”

IEEE Transactions on Neural Networks and Learning Systems, vol. 26, no. 7, pp. 1375–1387, July 2015.

[5] Zhifeng Li, Dihong Gong, Xuelong Li, and Dacheng Tao, "Aging face recognition: A hierarchical learning model based on local patterns selection," *IEEE Transactions on Image Processing*, vol. 25, no. 5, pp. 2146–2154, 2016.

[6] Bor-Chun Chen, Chu-Song Chen, and Winston H Hsu, "Face recognition and retrieval using cross-age reference coding with cross-age celebrity dataset," *Multimedia, IEEE Transactions on*, vol. 17, no. 6, pp. 804–815, 2015.

[7] G Baudat and Fatiha Anouar, "Generalized discriminant analysis using a kernel approach," *Neural Computation*, vol. 12, no. 10, pp. 2385–2404, 2000.

[8] Lifen Chen, Hongyuan Mark Liao, Mingtat Ko, Jachen Lin, and Gwojong Yu, "A new lda-based face recognition system which can solve the small sample size problem," *Pattern Recognition*, vol. 33, no. 10, pp. 1713–1726, 2000.

[9] Hua Yu and Jie Yang, "A direct lda algorithm for high-dimensional data ł with application to face recognition," *Pattern Recognition*, 2001.

[10] Juwei Lu, Konstantinos N Plataniotis, and A N Venetsanopoulos, "Face recognition using kernel direct discriminant analysis algorithms," *IEEE Transactions on Neural Networks*, vol. 14, no. 1, pp. 117–126, 2003.

[11] S. Milborrow and F. Nicolls, "Active Shape Models with SIFT Descriptors and MARS," *VISAPP*, 2014.

[12] X. Xiong and F. De la Torre, "Supervised descent method and its applications to face alignment," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, June 2013, pp. 532–539.

[13] Bor-Chun Chen, Chu-Song Chen, and Winston H. Hsu, "Review and implementation of high-dimensional local binary patterns and its application to face recognition," Tech. Rep. TR-IIS-14-003, Institute of Information Science, Academia Sinica, 2014.

[14] K. Ricanek and T. Tesafaye, "Morph: a longitudinal image database of normal adult age-progression," in *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, April 2006, pp. 341–345.

[15] Taekyun Kim, Josef Kittler, and Roberto Cipolla, "Discriminative learning and recognition of image set classes using canonical correlations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 1005–1018, 2007.