A Unified Framework for Age Invariant Face Recognition and Age Estimation

Changhong Wu and Jianbo Su

1 Introduction

Human facial appearances change substantially over time, resulting in great challenges for the face recognition task. On the other hand, the age estimation task expects to catch the age related information. To some extent, the two tasks can be considered as competing with each other. Most existing researches investigate the two tasks independently, while in this paper we explore the coordination between them.

Existing researches on age invariant face recognition (AIFR) fall into two categories: the generative methods and the discriminative methods. The generative methods [1–3] focus on generating an aging pattern which relies on accurate age estimation and the model is computationally expensive. The discriminative methods focus on extracting age invariant features for face recognition [4–7]. However, they ignore the aging effects on facial appearance.

The age estimate task usually performs by two steps: feature extraction and classification (or regression). Most existing methods extract the geometry and texture features and combine with the AAM to simultaneously model the shape and texture of face images [1, 8]. Nevertheless, it is difficult to design effective features for age estimation due to the unclear mechanism of the ageing effect on face images [9].

The method proposed in [10] considers the two tasks with conflicting goals can help inhibit irrelevant features for each task and hence improve the performance. Since the age invariant face recognition expects to extract age insensitive features while the age estimation expects the age sensitive features, it is desired to improve one of them by incorporating information from the other. Gong et al. [11, 12] propose a probabilistic model with the two latent factors and learn an identity space

C. Wu · J. Su (✉)
Department of Automation, Research Center of Intelligent Robotics,
Shanghai Jiao Tong University, Shanghai 200240, China
e-mail: jbsu@sjtu.edu.cn

and an age space at the same time. But the model is based on the assumption that the latent factors and random noise follow the Gaussian distribution. The method proposed in [13] consider that the face image can be decomposed into a class specific part and a common part and a dictionary can be split into two corresponding dictionaries. However, they ignore the discrimination ability of the coding coefficients.

In this paper, a novel unified framework is proposed for face recognition and age estimation. The two tasks are modeled together in a joint framework. Identity and age dictionaries are introduced to encode the identity and age parts of a face image onto two separated subspaces. The learned identity subspace only catches the identity information of the face images, which can be used for face recognition. The age subspace catches the age sensitive features which can be used for age estimation.

2 The Joint Model for Face Recognition and Age Estimation

2.1 Face Images Decomposition

Firstly, Suppose the training set is composed of c classes, then the training samples of the i-th class $F_i = [f_1, \ldots, f_{n_i}] \in R^{d \times n_i}$, where $f_j \in R^d, j = 1, \ldots, n_i$. Let $F = [F_1, \ldots, F_c] \in R^{d \times n}$, $n = n_1 + \cdots + n_c$ be the training samples matrix. Each training sample $f_j \in R^{d \times 1}, j = 1, \ldots, n$ can be decomposed into four parts: the person specific part $I_j \in R^{d \times 1}$, the age related part $A_j \in R^{d \times 1}$, the mean face feature $m_j \in R^{d \times 1}$ and the random noise part $\varepsilon_j \in R^{d \times 1}$. The face image decomposition can be formulated as:

$$f_j = m_j + I_j + A_j + \varepsilon_j, \quad j = 1, \ldots n \tag{1}$$

2.2 Collaboration Coding with Identity and Age Dictionaries

In this paper, two dictionaries are introduced to encode the identity part and age part of face features: the identity dictionary $U = [U_1, U_2 \ldots U_i, \ldots U_c] \in R^{d \times p}$, $U_i \in R^{d \times p_i}$ and the age dictionary $V = [V_1, \ldots V_s, \ldots V_t] \in R^{d \times q}$, $V_s \in R^{d \times q_s}$. The coding coefficients $X_i = [x_i^1, \ldots, x_i^j, \ldots x_i^{n_i}] \in R^{p \times n_i}$ denotes the identity component of the i-th class training samples $I_i \in R^{d \times n_i}$ coded over the identity dictionary $U \in R^{d \times p}$, where $x_i^j \in R^{p \times 1}$ denotes the identity coding efficient of f_i^j. Similarly, $Y_i = [y_i^1, \ldots, y_i^j, \ldots y_i^{n_i}] \in R^{q \times n_i}$ denotes the age component of the i-th class training

samples $A_i \in R^{d \times n_i}$ coded over the age dictionary $V \in R^{d \times q}$, where $y_i^j \in R^{q \times 1}$ denotes the age coding efficient of f_i^j.

In our model, over-complete dictionary is difficult to acquire since the number of training images per class is limited. Here, we introduce collaboration coding [15] and combine it with the reconstruction error to form a unified objective function:

$$\min \|F - M - UX - VY\|_F^2 + \lambda_1 \|X\|_F^2 + \lambda_2 \|Y\|_F^2, \tag{2}$$

where UX and VY denote the person specific component I and age related component A. The first term $\|F - m - UX - VY\|_F^2$ is based on the idea of data reconstruction. The mean face can be calculated by: $M = [m, m, \ldots, m] \in R^{d \times n}$, $m = \frac{1}{n} \sum_{i=1}^{c} f_i \in R^{d \times 1}$, where n is the size of the total training set. λ_1, λ_2 are introduced to balance the reconstruction error and the collaboration constraints.

2.3 Discriminative Coefficients

The collaboration coefficient of one face image can be regarded as a feature vector representing the data in a new space, which should be discriminant for each class. Here, two label supervised constraints are introduced [14]. Identity label supervised term $\|H_1 - AX\|_F^2$ ensures that the samples from the same identity get the similar identity coding coefficient x. $\|H_2 - BY\|_F^2$ ensures that the samples from the same age group get the similar age coding coefficient y. The unified objective function is given as follows:

$$\min \|F - M - UX - VY\|_F^2 + \|H_1 - AX\|_F^2 + \beta \|H_2 - BY\|_F^2$$
$$+ \lambda_1 \|X\|_F^2 + \lambda_2 \|Y\|_F^2, \tag{3}$$

where $H_1 \in R^{c \times p}$ and $H_2 \in R^{a \times p}$ denote the identity and age label matrixes. If the l-th training sample belongs to the k-th identity class, $H_1(k, l)$ equals 1, otherwise $H_1(k, l)$ equals 0, which is similar to the definition of $H_2(m, n)$.

2.4 Optimization of the Joint Model

The objection function in formula (3) is not convex for U, V, X, Y, A, B jointly. While it is convex with respect to each of them when the other ones are fixed.

- *Firstly, fix V, B, X, Y and update U, A. Formula (3) can be rewritten as:*

$$J_{U^*, A^*} = \arg \min_{U, A} \left\| \begin{bmatrix} F - M - VY \\ \sqrt{\alpha} H_1 \end{bmatrix} - \begin{bmatrix} U \\ \sqrt{\alpha} A \end{bmatrix} X \right\|_F^2, \qquad (4)$$

Let $D = \begin{bmatrix} U \\ \sqrt{\alpha} A \end{bmatrix}$, D can be updated class by class according to [16].

- *Secondly, fix U, A, X, Y and update V, B.*

  The update process of V, B is similar to the optimization of U, A;

- *Thirdly, fix U, V, A, B, Y and update X. Formula (3) can be rewritten:*

$$J_{X^*} = \arg \min_X \left\| \begin{bmatrix} F - M - VY \\ \sqrt{\alpha} H_1 \end{bmatrix} - \begin{bmatrix} U \\ \sqrt{\alpha} A \end{bmatrix} X \right\|_F^2 + \lambda_1 \|X\|_2, \qquad (5)$$

X can be calculated directly in formula (6).

$$X = \left(\begin{bmatrix} U^T & \sqrt{\alpha} A^T \end{bmatrix} \begin{bmatrix} U \\ \sqrt{\alpha} A \end{bmatrix} + \lambda_1 I \right)^{-1} \begin{bmatrix} U^T & \sqrt{\alpha} A^T \end{bmatrix} \begin{bmatrix} F - M - VY \\ \sqrt{\alpha} H_1 \end{bmatrix}. \qquad (6)$$

- *Fourthly, fix U, V, A, B, X and updated Y.*

  The update process of Y is similar to the optimization of X.

2.5 Face Matching

In the testing, with the learned two dictionaries U, V, any unseen face $f \in R^d$ can be encoded as follow (assume $\lambda_1 = \lambda_2$):

$$J_{x^*, y^*} = \arg \min_{x, y} \left\| f - m - \begin{bmatrix} U & V \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \right\|_F^2 + \lambda_1 \left\| \begin{bmatrix} x \\ y \end{bmatrix} \right\|_2. \qquad (7)$$

The coding coefficient x can be directly used as an age invariant feature descriptor for face recognition. The coding coefficient y can be used for age estimation. Cosine distance and the Nearest Neighbor (NN) algorithm are used for both tasks.

3 Experiment

3.1 Implementation Details

Follows the settings in [12], HOG feature is used as feature descriptor. Feature slicing and PCA are used for dimension reduction. The PCA feature dimension d, the number of slices p, label parameters α, β and regularization parameters λ_1, λ_2 are set as follows: $d = 900$, $p = 6$, $(\alpha, \beta, \lambda_1, \lambda_2) = (2.3, 1.2, 0.005, 0.005)$ for MORPH database, $(\alpha, \beta, \lambda_1, \lambda_2) = (2.3, 0.01, 0.002, 0.002)$ for FGNET database.

3.2 Experiment on the FGNET Database

The FGNET database contains1002 face images with 82 different individuals. Each one has 13 images on average collected at ages in the range of 0 to 69. To train the joint model, the face images in FGNET database is divided into 9 age groups every 5 age gap.

Following the training and testing split scheme used in [5, 11], the leave-one-person-out (LOPO) scheme is used for validation. Table 1 records the rank-1 recognition rate of these methods. It shows that our method can achieve an accuracy higher than the best published result (76.2%). Moreover, the proposed model outperforms the HOG feature baseline method by a clear margin.

## 3.3 Experiment on the MORPH Database

The MORPH Album2 database contains 55,134 facial images of 13,618 subjects. Following the experimental settings in [17], we randomly select 10,000 subjects from the MORPH Album2. For each subject, the images with the youngest age is

Table 1 Comparison of age invariant face recognition methods on FGNET database

	Database (subjects, images) in probe and gallery	Face recognition accuracies (%)	Age estimation accuracies (%)
Geng et al. [1]	FGNET(10,10)	38.1	–
Park et al. [2]	FGNET(82,1002)	37.4	–
Li et al. [5]	FGNET(82,82)	47.5	–
Gong et al. [11]	FGNET(82,82)	69.0	–
Gong et al. [15]	FGNET(82,82)	76.2	–
HOG only	FGNET(82,82)	53.7	–
The joint model	FGNET(82,82)	79.8	74.4

Table 2 Performance of different methods on the MORPH database

Algorithms	Rank-1 face identification (%)	Rank-1 age estimation (%)
Park et al. [2]	79.80	–
Li et al. [5]	83.90	–
Du et al. [3]	79.24	–
Klare et al. [4]	79.08	–
Otto et al. [7]	81.27	–
HOG + HFA (2013)	91.14	–
HLBP + CARC(2014)	92.80	–
HOG + IFA (2015)	92.26	–
Method (2015) in [18]	87.13	–
HOG only	43.65	–
The proposed method	89.57	83.19

True	41	46	24	30	45	29
Predict	35-40	40-45	25-30	35-40	40-45	30-35
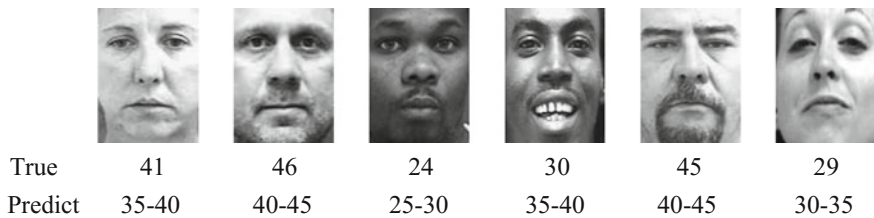
Fig. 1 Some failed age estimation results in MORPH Album2

selected as gallery set and the images with the oldest age is used as probe set. Then another 1000 subjects are randomly selected from the dataset as the training set.

The experiment results are shown in Table 2. One thing to note is that the age span of one subject on the MORPH dataset is relative small and face images contain more expression and pose variations. Nevertheless, our approach still achieves a good result on both two tasks. Some failed examples of age estimation are shown in Fig. 1. Some failed results of face matching are shown in Fig. 2.
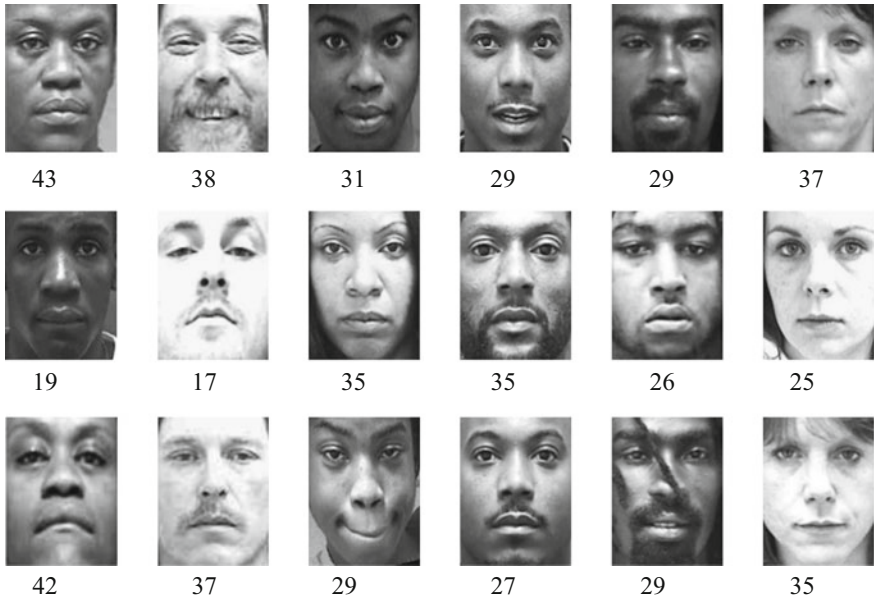
| 43 | 38 | 31 | 29 | 29 | 37 |

| 19 | 17 | 35 | 35 | 26 | 25 |

| 42 | 37 | 29 | 27 | 29 | 35 |

Fig. 2 Some failed retrieve results in MORPH Album2. The *first row* shows the probe images and the *second row* is the incorrect rank-1 matching results using the proposed model. The *third* shows the correct match in the gallery

4   Conclusions

In this paper, we develop a joint model to tackle age invariant face recognition and age estimation at the same time. The result of the experiments on the MORPH and FGNET databases indicate the effectiveness of the proposed method.

References

1. Geng X, Zhou ZH, Smith-Miles K (2007) Automatic age estimation based on facial aging patterns. IEEE Trans Pattern Anal Mach Intell 29(12):2234–2240
2. Park U, Tong Y, Jain AK (2010) Age-invariant face recognition. IEEE Trans Pattern Anal Mach Intell 32(5):947–954
3. Du JX, Zhai CM, Ye YQ (2013) Face aging simulation and recognition based on NMF algorithm with sparseness constraints. Neurocomputing 116:250–259
4. Klare B, Jain AK (2011) Face recognition across time lapse: on learning feature subspaces. In: International joint conference on biometrics (IJCB). IEEE, pp 1–8

5. Li Z, Park U, Jain AK (2011) A discriminative model for age invariant face recognition. IEEE Trans Inf Forensics Secur 6(3):1028–1037
6. Ling H, Soatto S, Ramanathan N, Jacobs DW (2010) Face verification across age progression using discriminative methods. IEEE Trans Inf Forensics Secur 5(1):82–91
7. Otto C, Han H, Jain A (2012) How does aging affect facial components? In: Computer Vision_ECCV 2012. Workshops and demonstrations. Springer, Berlin, pp 189–198
8. Lanitis A, Draganova C, Christodoulou C (2004) Comparing different classifiers for automatic age estimation. IEEE Trans Syst Man Cybern 34(1):621–628
9. Niu Z, Zhou M, Wang L, Gao X, Hua G (2016) Ordinal regression with multiple output cnn for age estimation. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4920–4928
10. Du L, Ling H (2015) Cross-age face verification by coordinating with cross-face age verification. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2329–2338
11. Gong D, Li Z, Lin D, Liu J, Tang X (2013) Hidden factor analysis for age invariant face recognition. In: Proceedings of the IEEE international conference on computer vision, pp 2872–2879
12. Gong D, Li Z, Tao D, Liu J, Li X (2015) A maximum entropy feature descriptor for age invariant face recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 5289–5297
13. Sun Y, Liu Q, Tang J, Tao D (2014) Learning discriminative dictionary for group sparse representation. IEEE Trans Image Process 23(9):3816–3828
14. Jiang Z, Lin Z, Davis LS (2013) Label consistent K-SVD: Learning a discriminative dictionary for recognition. IEEE Trans Pattern Anal Mach Intell 35(11):2651–2664
15. Zhang L, Yang M, Feng X (2011) Sparse representation or collaborative representation: which helps face recognition? In: IEEE international conference on computer vision (ICCV). IEEE, pp 471–478
16. Yang M, Zhang L, Yang J, Zhang D (2010) Metaface learning for sparse representation based face recognition. In: 17th IEEE international conference on image processing (ICIP). IEEE, pp 1601–1604
17. Chen BC, Chen CS, Hsu WH (2014) Cross-age reference coding for age-invariant face recognition and retrieval. In: European conference on computer vision. Springer International Publishing, Berlin, pp 768–783
18. Li Z, Gong D, Li X, Tao D (2015) Learning compact feature descriptor and adaptive matching framework for face recognition. IEEE Trans Image Process 24(9):2736–2745