Eigen-Aging Reference Coding for Cross-Age Face Verification and Retrieval

Kaihua Tang[1,2](✉), Sei-ichiro Kamata[1], Xiaonan Hou[2], Shouhong Ding[2], and Lizhuang Ma[2]

[1] Graduate School of Information, Production and Systems,
Waseda University, Shinjuku, Japan
kam@waseda.jp
[2] Shanghai Jiao Tong University, Shanghai, China
tkhchipaomian@gmail.com, xnhou1989@gmail.com, dingsh1987@gmail.com,
ma-lz@cs.sjtu.edu.cn

Abstract. Recent works have achieved near or over human performance in traditional face recognition under PIE (pose, illumination and expression) variation. However, few works focus on the cross-age face recognition task, which means identifying the faces from same person at different ages. Taking human-aging into consideration broadens the application area of face recognition. It comes at the cost of making existing algorithms hard to maintain effectiveness. This paper presents a new reference based approach to address cross-age problem, called Eigen-Aging Reference Coding (EARC). Different from other existing reference based methods, our reference traces eigen faces instead of specific individuals. The proposed reference has smaller size and contains more useful information. To the best of our knowledge, we achieve state-of-the-art performance and speed on CACD dataset, the largest public face dataset containing significant aging information.

1 Introduction

Growing number of corporations and organizations use face recognition algorithms to realize interaction and verification applications in recent years. In spite of the high recognition accuracy for well-captured images, it's still a tough task to maintain the effectiveness under various real-world deformation factors. Among all the factors that may cause the reduction of accuracy, PIE (pose, illumination, expression) and facial aging overwhelmed others. Compared with PIE, the facial aging is more complicated. Even for human beings, to identify the same person under different ages is not an easy job (see Fig. 1). Meanwhile, cross-age face recognition is so crucial for a real-world face recognition system, for example, it can be used to find escaped prisoners or missing people. Without cross-age face recognition ability, face information has to be updated frequently to keep the recognition system effective. Most of the previous researches achieve near or over human performance only under PIE variation [2,3,12,13]. Facial aging researches still have a lot to improve.
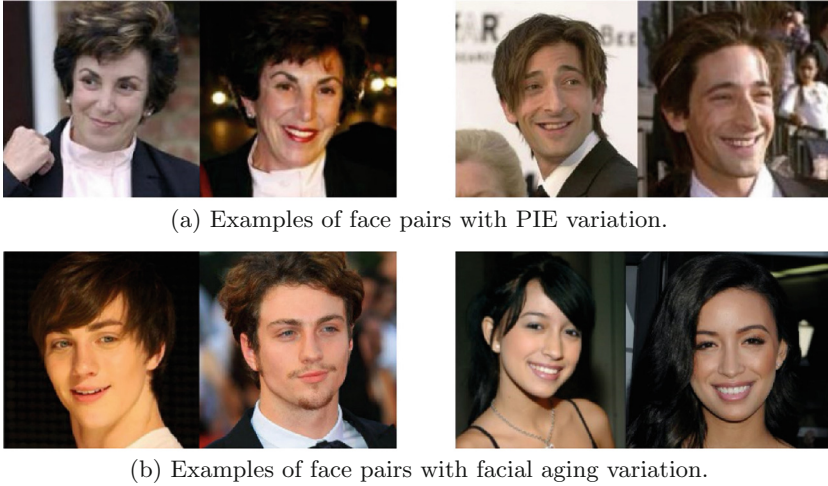
(a) Examples of face pairs with PIE variation.

(b) Examples of face pairs with facial aging variation.

Fig. 1. These image pairs show the difference between traditional PIE variation and facial aging variation. Facial aging variation is much more complicated. (a) The above image pairs come from LFW [21] dataset, a commonly used human face dataset, which doesn't contain aging information. (b) The bottom image pairs come from the largest public facial aging dataset CACD [1].

Among all the age related researches, age estimation [4,14,15] and aging simulation [5,16,17] researches have narrow application area, so we mainly investigate cross-age face recognition. As far as we know, the existing works can be roughly divided into three categories: the modeling approaches [6,7], the discriminative approaches [8,9,18,19] and the reference-based approaches [1]. Modeling approaches use aging simulation model to change query faces into the same age as gallery ones. Discriminative approaches try to separate or eliminate the age-sensitive features to increase the recognition ability. Reference-based approaches achieve age-invariance by comparing the face features with reference individuals in different ages. Figure 2 provides an example of how reference-based approaches achieve age-invariance.

The reference-based cross-age face recognition is first proposed by Bor-Chun Chen et al. [1], which is called Cross-Age Reference Coding (CARC). It achieves remarkable improvement compared to previous researches. But it still has some drawbacks. First, the reference set is extremely large, since it has to cover the diversity among race, gender and so on. Second, due to the expensive computational cost, it can't make full use of all training individuals. At last, in sparse coding, CARC only adds locality constraint to ensure the smoothness, but the global distribution may be changed.

To tackle these drawbacks, this paper introduces a new reference-based method called Eigen-Aging Reference Coding (EARC). Although eigen face isn't the first choice of recent face recognition researches, we find that it's effective in building eigen face reference in our task. Actually, it's the first time that eigen
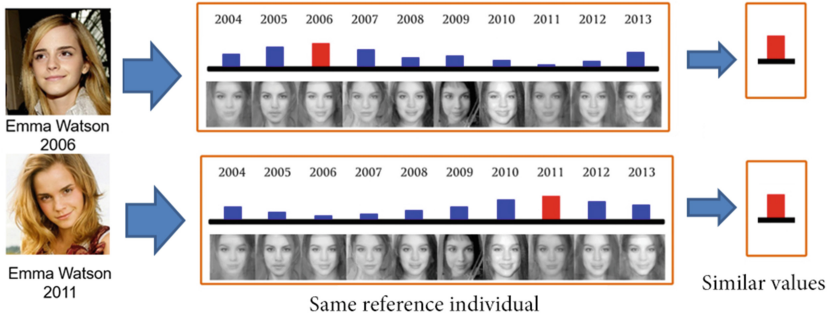
Fig. 2. After a pair of faces are encoded by a reference individual, they pool out their absolute maximums. If two faces come from the same person, the activated values are supposed to be similar.

face is used to build reference set. We use PCA to select some eigen components of face features just like how we calculate eigen faces [22]. Instead of tracing the aging process for specific individuals, the proposed reference set traces the aging process of these eigen face components as eigen-aging reference. We also add distribution constraint in our sparse coding to guarantee the encoded results follow good distributions. Contributions of this paper can be concluded into following four parts:

– Using PCA to remove the redundancy and cover the diversity among training individuals, which makes each of our eigen-aging reference contain more useful information.
– The number of reference and result dimension of our method are dramatically reduced compared with previous reference-based methods. EARC can make full use of all training individuals without increasing the computational cost.
– The proposed distribution constraint ensures that our encoded result is of good distributions. It improves the performance of sparse coding.
– The proposed method achieves state-of-the-art performance in both retrieval and verification, which is better than human average performance and very close to human voting result.

The rest of this paper is organized as follows. In Sect. 2, we introduce some related works. In Sects. 3 and 4, we separate our Eigen-Aging Reference Coding into reference construction part and encoding part. In Sect. 5, we provide some experiment results running on CACD. This paper will be concluded in Sect. 6.

2 Related Work

2.1 Cross-Age Face Recognition

Most of the highly qualified researches in cross-age face recognition start after MORPH [20] dataset is published. It keeps being the largest facial aging dataset

until CACD [1] occurs. Except these two, the other facial aging datasets either contain small data size or have low quality. Limited by the rareness of datasets, there are still few researchers focusing on this field. To the best of our knowledge, we divide existing cross-age face recognition methods into three categories: the modeling approaches, the discriminative approaches and the reference-based approaches. The modeling approaches [6,7] change the query faces into the same age as gallery one. Although it removes some variations caused by facial aging, the main problem is that the diversity of aging process between different race or gender can't be covered, which makes most of the modeling approaches hard to be general. The discriminative approaches have become popular in recent years. Most of them seem to be very effective in improving the recognition ability by eliminating age-sensitive features. Zhifeng Li et al. [8] build Local Patterns Selection feature descriptor to achieve age-invariance. It applies clustering encoding tree on feature space and removes facial aging variation by minimizing intra-user dissimilarity among different ages. Dihong Gong et al. [9] use hidden factor analysis to separate the features into age-sensitive factors and age-invariant factors. However, aging process is not just a process in chaos. Age-sensitive features can also be used for cross-age face recognition, if we take advantage of their inner regularity: similar faces are supposed to have similar aging process. That's why two twins look alike all through their life. Reference-based cross-age approaches make use of this regularity. They trace aging processes of some standard reference individuals.

Before reference set is used in cross-age face recognition, it has already been applied by lots of methods to improve traditional face recognition systems [10,11] and achieves quite good results. Kumar et al. [11] present attribute classifier and simile classifier, the simile classifier uses reference people to do classification. Qin Yin et al. [10] propose an associate-predict model based on a 200 identities reference set. However, their reference sets don't contain aging variation. Bor-Chun Chen et al. [1] further find the value of reference in solving facial aging problem and use it to achieve age-invariance. They build a reference based on 600 individuals to deal with cross-age face recognition, which is called cross-age reference coding (CARC). CARC assumes that similar faces should still look alike when they both get older. It's an obvious phenomenon if we think about the above twins example. Its reference traces aging processes of reference individuals, then it encodes the faces by pooling out the maximum similarity between input face and reference individuals at different ages. Both younger faces or elder faces are supposed to activate their own corresponding age at reference and have similar activated values (see Fig. 2), so the age-invariant encoded feature can be obtained. Compared with other approaches, CARC shows much better performance and robustness.

However, all the existing reference methods mentioned above [1,10,11] use specific individuals as reference. Inspired by eigen face [22], the proposed eigen-aging reference first uses eigen faces as reference. It achieves much better performance and less computational consumption.

2.2 High-Dimensional LBP

Among all the feature descriptors applied in face recognition applications, we select high-dimensional LBP (HD-LBP) [2] to extract face features. It has been proved to have near human performance in traditional face recognition.

High-dimensional LBP features will be extracted on multi-scale patches around each landmark. Landmarks are some fixed points on human face like centers of eyes, corners of the mouth, tip of the nose and so on. Each landmark will have an independent high-dimensional feature vector. Because the multi-scale sampling extracts too much redundant information, we further use PCA to reduce its dimension. It will maintain the performance and reduce the computational consumption for further processing.

3 Eigen-Aging Reference

3.1 Form Training Individual Sets

Before calculating aging processes of eigen components, we have to obtain the training individual representations using Eq. (1). Average HD-LBP features of those images from same person at same age are used by:

$$R_{i,j}^k = \frac{1}{N_{ij}} \sum_{individual(x^k)=i, year(x^k)=j} x^k,$$
$$\forall i = 1, 2, ..., n; \forall j = 1, 2, ..., m; \forall k = 1, 2, ..., q; \tag{1}$$

where $x^k \in R^d$ is the HD-LBP feature at landmark k, $R_{i,j}^k$ means the average face feature of individual i in year j at landmark k, n is the number of training individuals, m is range of ages, and q is the number of landmarks. In our experiments, $n = 1200$, $m = 10$, $q = 16$. N_{ij} is the number of all images from individual i in year j. In CACD dataset, N_{ij} is nonzero for any i and j.
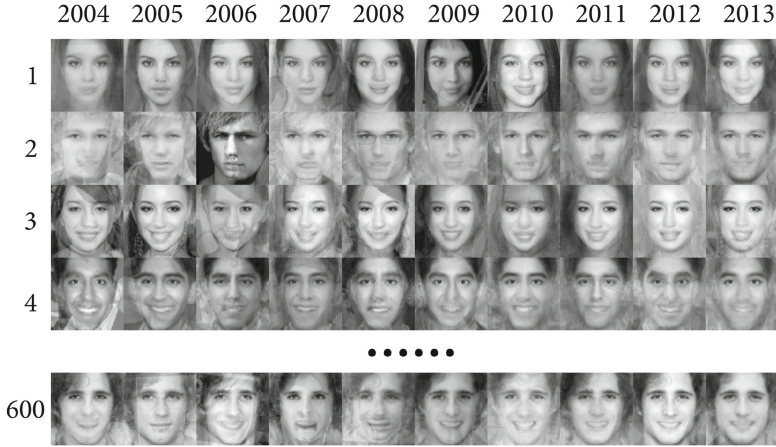
3.2 Train Eigen-Aging Reference

The Eigen-Aging Reference will be trained according to eigen face algorithm [22]. We make use of the above $R_{i,j}^k$ as training representations and calculate eigen components of them. For the convenience of formulization, we concatenate training representations into vectors by year as:
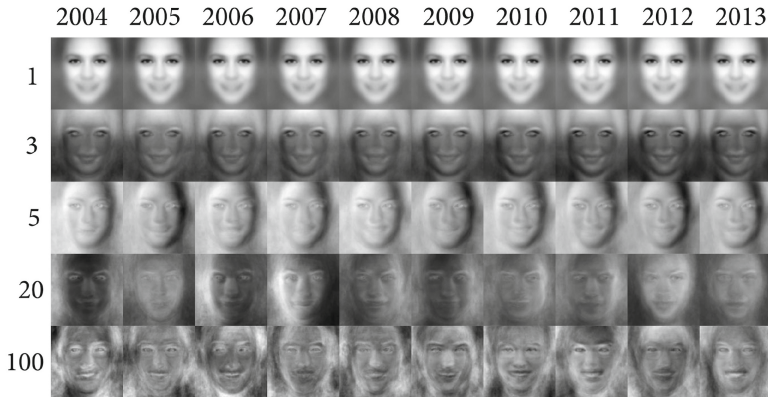
$$\hat{R}_i^k = [(R_{i,1}^k)^T, (R_{i,2}^k)^T, ..., (R_{i,m}^k)^T]^T \in R^{md}. \tag{2}$$

After that, PCA is applied to obtain eigen components of n training feature vectors $\hat{R}_i^k$, then we can get the average vector M^k and difference vectors Φ_i^k in the following:

$$M^k = \frac{1}{n} \sum_{i=1}^{n} \hat{R}_i^k, \qquad \Phi_i^k = \hat{R}_i^k - M^k. \tag{3}$$

2004 2005 2006 2007 2008 2009 2010 2011 2012 2013

1

2

3

4

• • • • • •

600

(a) The examples of Cross-Age Reference.

2004 2005 2006 2007 2008 2009 2010 2011 2012 2013

1

3

5

20

100

(b) The examples of proposed Eigen-Aging Reference.

Fig. 3. The above illustrations use raw images to make reference set easier to understand. (a) Cross-Age Reference is based on specific individuals. (b) Eigen-Aging Reference is based on eigen faces. The number on the left of each row is the rank of the corresponding eigen face.

Next we construct matrixes $\Phi^k = [\Phi_1^k, \Phi_2^k, ..., \Phi_n^k] \in R^{md \times n}$, and calculate eigenvalues λ_l^k and eigenvectors u_l^k of $(\Phi^k)^T \Phi^k$. Top-p λ_l^k and u_l^k are used, where $l = 1, 2, ..., p$. According to Sect. 5, the best performance can be achieved when $p = 50$ and $n = 1200$, so we come up with eigen-aging vector $\hat{E}_l^k$ for each eigen face component l by:

$$\hat{E}_l^k = \frac{1}{sqrt(\lambda_l^k)} \Phi^k u_l^k, \qquad l = 1, 2, ..., p. \qquad (4)$$

We further separate $\hat{E}_l^k \in R^{md}$ into sub-vectors $E_{l,j}^k \in R^d$ by year as follows:

$$\hat{E}_l^k = [(E_{l,1}^k)^T, (E_{l,2}^k)^T, ..., (E_{l,m}^k)^T]^T. \qquad (5)$$

$E_{l,j}^k$ is our proposed reference set representation, which is smaller and contain more useful information.

Figure 3 shows the difference between EARC and CARC. We use raw images to make reference set easier to understand. As we can see, specific individual based reference contains too much noise and has redundancy among different individuals while the proposed eigen component based reference is more representative and clear. With the above advantage, EARC only use tens of eigen face references but achieve better performance than CARC, the method based on hundreds of specific individuals. In the example of eigen aging reference, we display 1st, 3rd, 5th, 20th and 100th eigen face references. They are ranked according to their eigenvalues. We also find that the higher ranked components contain more structure information while the lower ranked ones may have more noise, so if we use too much eigen faces, the result could become worse. In Sect. 5, we will discuss how many eigen faces we should use.

4 Coding and Pooling

4.1 Sparse Coding

Before we apply sparse coding, We need to define relationship values $\alpha_{l,j}^k$ between input feature and l^{th} eigen face reference in year j at landmark k (we denote matrix $\check{E}_j^k = [E_{1,j}^k, E_{2,j}^k, ..., E_{p,j}^k] \in R^{d \times p}$ and vector $\check{\alpha}_j^k = [\alpha_{1,j}^k, \alpha_{2,j}^k, ..., \alpha_{p,j}^k] \in R^p$). Our reference set is used as dictionary, so it can be considered as solving a Tikhonov regularization problem:

$$\underset{\alpha_{i,j}^k}{\text{minimize}} ||x^k - \check{E}_j^k \check{\alpha}_j^k||^2 + \lambda ||\check{\alpha}_j^k||^2, \forall j, k. \qquad (6)$$

An additional locality constraint [1] will also be used to improve performance. It will guarantee the smoothness of encoded results, which means that similar high dimensional points in original feature space still have similar value in encoded results. In other word, the relationship values between input face and eigen face reference l in year j should be similar to the values between input face and the same reference face in year $j + 1$ and $j - 1$, so for any k and l, $\alpha_{l,j}^k$ will always be similar to $\alpha_{l,j+1}^k$ and $\alpha_{l,j-1}^k$.

This locality smoothness constraint is defined as $\lambda ||LA^k||^2$. We let $A^k = [(\check{\alpha}_1^k)^T, (\check{\alpha}_2^k)^T, ..., (\check{\alpha}_m^k)^T]^T \in R^{mp}$ and matrix L:

$$L = \begin{bmatrix} I & -2I & I & 0 & \cdots & 0 & 0 & 0 \\ 0 & I & -2I & I & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & I & -2I & I \end{bmatrix} \in R^{(m-2)p \times (mp)}. \qquad (7)$$

4.2   Distribution Constraint

In spite of that locality constraint $\lambda||LA^k||^2$ ensures the smoothness of encoded results and improves the coding performance, it can't keep the distribution still being similar to original features. Mathematically, the statement that near points maintain closer is not equal to that far points keep far away. Only when both of these two statements are satisfied, the global distribution can be maintained after changing feature spaces. So we propose a new constraint to guarantee the encoded features follow the distributions we need.

With smoothness constraint, we may have four possible distributions of encoded $\alpha_{l,j}^k$ sequences (see Fig. 4). Based on common sense, if we compare a face with a series of faces from identical individual in different ages, there should be only one most similar face. It can locate either at the boundary of age sequence or at one certain age inside the sequence, so there is supposed to be one and only one extreme point. The locality smoothness constraint may lead to distribution 4 in Fig. 4, which is not a good distribution in our work, So we propose two kinds of constraint terms that may force the encoded features follow the first three distributions.

Maximize boundary difference: We assume that the extreme point will occur at the boundary in most of the cases. It's quite rare for the extreme point located at exact center of the reference age. So we can simply try to maximize the difference of two boundary ages. The new constraint will be $-\beta||DA^k||^2$. The minus transfers the maximize problem to minimization by:

$$D = \begin{bmatrix} I\ 0 \cdots 0\ -I \end{bmatrix} \in R^{p\times(mp)}. \tag{8}$$

Additional cost for extreme point: We also can force the encoded results into first three distributions by giving additional cost for extreme point. The good distributions should have only one extreme point while the bad distribution may have several ones. So we give the extreme point additional cost for occurrence. This constraint can also be written as $-\beta||DA^k||^2$ with new D:

(a) Distribution 1   (b) Distribution 2   (c) Distribution 3   (d) Distribution 4
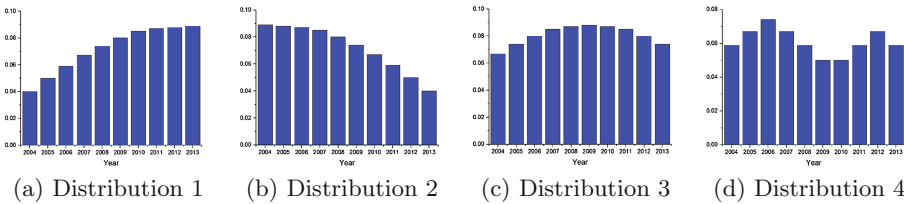
Fig. 4. With smoothness constraint, we may get above 4 distributions, but only the first three are good distributions in our work. Distribution 4 means all the possible distributions with more than one extreme point.

$$D = \begin{bmatrix} I & 0 & -I & 0 & \cdots & 0 & 0 & 0 \\ 0 & I & 0 & -I & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & I & 0 & -I \end{bmatrix} \in R^{(m-2)p \times (mp)}. \tag{9}$$

We should also notice that the locality constraint and distribution constraint both represent the distance relationship between original feature space and encoded feature space. For the convenience of learning parameters, they are supposed to be learned together. The combined constraint is $\lambda(||LA^k||^2 - \beta||DA^k||^2)$.

4.3   Optimization

To combine all the constraints, we denote $X^k = [(x^k)^T, ..., (x^k)^T]^T \in R^{mp}$ and matrix F:

$$F^k = \begin{bmatrix} \check{E}_1^k & 0 & \cdots & 0 \\ 0 & \check{E}_2^k & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \check{E}_m^k \end{bmatrix} \in R^{(md) \times (mp)}, \tag{10}$$

so the final optimization function is:

$$\underset{A^k}{\text{minimize}} ||X^k - F^k A^k||^2 + \lambda_1 ||A^k||^2 + \lambda_2(||LA^k||^2 - \beta||DA^k||^2), \forall k. \tag{11}$$

It's easy to obtain $A^k = ((F^k)^T F^k + \lambda_1 I + \lambda_2 L^T L - \lambda_2 \beta D^T D)^{-1}(F^k)^T X^k, \forall k$. We define P^k to be a projection matrix, $P^k = ((F^k)^T F^k + \lambda_1 I + \lambda_2 L^T L - \lambda_2 \beta D^T D)^{-1}(F^k)^T$. HD-LBP face features can be easily transferred into reference space by multiplying with P^k.

4.4   Max Pooling

After sparse coding is applied, we use maximum pooling to achieve age-invariance. If a face is young, it should be more similar to the younger part of each eigen face reference, which means that the corresponding younger age has a little bit larger $\alpha_{l,j}^k$ than elder one. The maximum pooling will pool out this value. Two faces from same person at different age will pool out different ages but their maximum values are similar compared with the faces from different people. This is the reason why we can use maximum pooling to achieve age-invariance.

$A^k \in R^{mp}$ is a vector containing $\alpha_{l,j}^k$ as elements. We calculate the absolute maximum of each eigen face l on different ages, $max(|\alpha_{l,1}^k|, |\alpha_{l,2}^k|, ..., |\alpha_{l,m}^k|), \forall l, k$. The final age-invariant face feature is a p dimensional vector, p is the number of eigen faces we use in constructing eigen-aging reference.

5   Experiments

5.1   Cross-Age Face Dataset

Compared with traditional human face datasets, the cross-age datasets are extremely rare because it needs to track individuals over decades. The most popular datasets in this field are FG-NET, MORPH [20] and CACD [1]. CACD is published in recent years but its quality has already been proved. All the images of CACD are celebrity images captured in various unconstrained environments, compared with other cross-age datasets, images in CACD are more close to the complicated real-world environment.

The difference among FG-NET, MORPH and CACD is shown in Table 1. FG-NET is an early dataset. It contains only one thousand images from no more than one hundred individuals. Because of its limitation of data size, it's hard to support a general method. MORPH contains 55,134 images of 13,618 people with age range from 16–77, but they are in clear background environment. Another demerit of MORPH is that there are about 4 images in average for each individual and only one image in each certain age. It's hard to cover different PIE conditions. The CACD contains 163,446 images of 2,000 celebrities with age ranging from 16 to 62. All the collected images are from 2004 to 2013, and each individual has 80 images in average, which means about 8 images in every certain age per individual. It ensures that face information of each single individual in each age can be fully extracted, so CACD is the best choice for us to build a robust and effective reference. In our experiment, all the training and testing images come from CACD.

Table 1. The difference among FG-NET, MORPH and CACD.

Dataset	# of images	# of individuals	# images/individual	Age gap
FG-NET	1,002	82	12.2	0–45
MORPH [20]	55,134	13,618	4.1	0–5
CACD [1]	163,446	2,000	81.7	0–10

5.2   Similarity Measurement

We use cosine similarity to measure the difference between two age-invariant face features in our experiment. Because the feature from each landmark is calculate independently, we add up these similarities from all the landmarks, and use the sum to represent the total similarity between two faces.

5.3   Training Data and Parameters Selection

To compare our experiment results to the previous state-of-the-art methods in CACD, we organize our training data and test data in the same way. There are

Table 2. The coding parameters of EARC in three cases.

Case	λ_1	λ_2	β
EARC	10^0	10^1	-
EARC-maximize boundary difference	10^{-2}	10^4	10^{-1}
EARC-additional extreme point cost	10^{-1}	10^4	10^3

2,000 celebrities in total. 1,800 of them come from internet without annotation. (1) 1,200 of these 1,800 individuals are used to calculate reference representations; (2) 600 are used to calculate PCA subspace in High-Dimensional LBP. The rest images of 200 individuals have already been manually annotated, which means that their quality can be guaranteed. (3) We use 80 of them to learn parameters. (4) 120 of them test our experiment results.

To learn our parameters, we use 80 qualified celebrities. The parameters include the dimension of PCA subspace in high-dimensional LBP, the regularization parameters $\lambda_1, \lambda_2, \beta$, and the number of eigen faces we use. The images captured in 2004–2006 are collected as gallery set while those captured in 2013 are query set. The reason why we don't use images from 2007–2012 as gallery set to learn parameters is that a larger gap of age between gallery and query set is more meaningful for a cross-age face recognition system.

– In high-dimensional LBP, we use PCA to reduce the dimension while remove some variations. The original landmark dimension is 4,720. We try to reduce it to the range from 100 to 1,500. According to their performance and computational cost, we choose 800 dimension in our experiment. 900 and 1,000 dimension can increase a little bit performance, but the required computer memory and computational cost are extremely expensive.
– To learn regularization parameters λ_1, λ_2 and β, we greedily train them one by one from value 10^{-6} to 10^6. We record the parameters in three cases (see Table 2): EARC without distribution constraint, EARC with maximize boundary difference constraint, EARC with additional cost for extreme point constraint.
– In order to find a proper number of eigen faces, We first use 600 training individuals to train and select eigen face components from 10 to 100. It shows that the best performance is achieved at 70 eigen face components, since the lower ranked eigen components may have more noise and less structure information. Then we change the number of training individuals from 600 to 1200. The best performance achieves at 70, 60, 60, 50 when training individuals are 600, 800, 1000, 1200 (see Fig. 5(a)). It seems like the larger training data we use, the less eigen components we will need. So we make use of all the 1200 training individuals and choose top 50 eigen components as reference. Under this condition, it will have the best performance and the lowest computational expense of projection. CARC can't make full use of all 1200 training individuals due to its expensive projection cost.
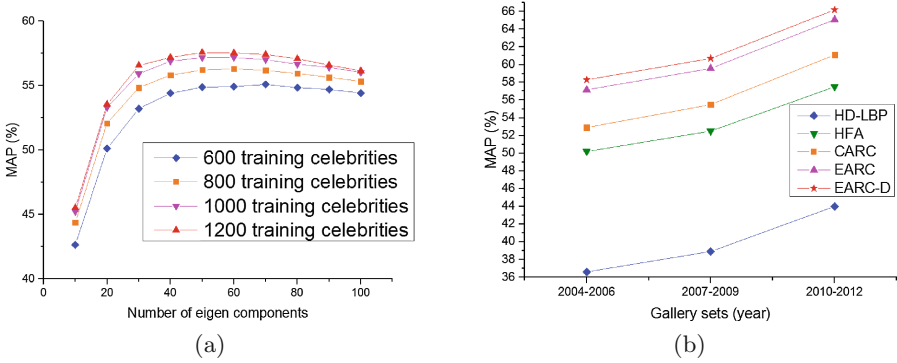
Fig. 5. (a) How does the retrieval performance change with the number of training individuals and number of eigen components. (b) The retrieval results of the proposed methods and previous state-of-the-art approaches. It shows that EARC-D achieves best performance at all the three age gaps.

5.4   Retrieval Experiments

In order to test retrieval performance of our proposed method, the rest annotated 120 celebrities are used. Because all the images are labelled the captured year from 2004–2013, we organize the gallery sets and query set by year. The images captured in 2013 will be gathered as query set while images captured in 2004–2006, 2007–2009 and 2010–2012 will be collected into 3 gallery sets.

The Mean Average Precision (MAP) is used to evaluate retrieval performance. It is widely used in information and image retrieval. If there is a query set contains Q query images, for each query image, it will compute its averaged precision (AP) of retrieval results at every recall level. MAP is the final average of these APs.

We compare our retrieval results (with and without distribution constraint) with the state-of-the-art methods CARC [1] and HFA [9]. They are proved to be very effective in CACD dataset. The results are shown in Fig. 5(b). We find that both maximize boundary difference constraint and additional cost for extreme point constraint have similar results, so we only use one EARC-D to represent distribution constraint. And it indeed improves the performance of original EARC.

5.5   Verification Experiments

Verification experiments are conducted under a verification subset of CACD called CACD-VS. It contains 4,000 image pairs, which have been manually checked to guarantee the quality. Half of these 4,000 image pairs are positive (come from the same person), the rest are negative (come from different people).

The CACD-VS contains two human performance benchmarks: average human performance and human voting performance. The former is the average of human accuracy, which is 85.7%. The human voting has a much better

result, which is 94.2% accuracy. It uses decisions from 9 human beings to do each verification. The majority decision will be considered as final decision. For now, human voting result still shows the best accuracy over all the published methods.

In the proposed method, we separate data into 10 folds, 200 positive pairs and 200 negative pairs for each. 9 of them are used to learn PCA subspace and 1 to test. It will run 10 times and compute average results. After calculating cosine similarities, a simple threshold is used to classify. Both two kinds of distribution constraints have similar results, so we use EARC-D and EARC to represent the result with or without distribution constraint. We compare our methods with CARC [1], HFA [9] and HD-LBP [2] (see Table 3). It achieves the best accuracy, 91.2%, which is very close to human voting result.

Table 3. The verification results.

Method	Verification accuracy
High-dimensional LBP [2]	81.6%
Hidden factor analysis [9]	84.4%
Average human	85.7%
Cross-age reference coding [1]	87.6%
Eigen-aging reference coding	90.6%
Eigen-aging reference coding-D	91.2%
Human voting	94.2%

5.6 Computational Cost

Beside of good performance EARC and EARC-D achieve, they also significantly speed up the computation by reducing the encoded dimension. For a real-world face recognition system, it always contains huge amount of face information in the dataset. A computational expensive method has less practical application value. Compared with CARC, we only use 1/12 of the dimension to represent an age-invariant face and significantly improve the efficiency. The comparison of encoded dimension and computation time between CARC and EARC will be shown in Table 4. To measure the retrieval efficiency, we retrieve 100 face images in a gallery set with 10,000 images. The retrieval speed shows the computation

Table 4. The encoded dimension and computation time of CARC, EARC and EARC-D.

Method	Dimension of landmark	Dimension of face	Retrieval speed (ms)
CARC [1]	600	9600	9237
EARC	50	800	1352
EARC-D	50	800	1347

time of each method. This experiment is running under a computer with Intel(R) Core(TM) i7-4720HQ CPU @ 2.60 GHz 2.60 GHz, 16.0 GB RAM and MATLAB R2013a.

6   Conclusions

In this paper, we mainly propose an eigen face component based reference to encode the faces into an age-invariant space. It performs better than specific individual based reference and requires less computation time. We also present the distribution constraint to improve sparse coding. It further optimizes our method without costing additional computational consumption. Although the proposed two kinds of distribution constraint terms are based on different assumption, their mathematical similarity results in similar results.

In spite of the state-of-the-art result we achieve in CACD, it's still not as good as the human voting result. We suppose cross-dataset voting could improve our performance, because it might increase the stability of our system. We assume cross-dataset voting can lead to higher accuracy, because human voting is better than human average. Limited by the rareness of public cross-age dataset, we only try separating CACD into several small datasets. Although it improves a little bit, it doesn't make a mentionable difference. This is because the same dataset doesn't have enough appearance difference. If there are several large facial aging datasets with good quality and different appearance distributions, a more robust cross-dataset voting method may exceed human voting result.

References

1. Chen, B.-C., Chen, C.-S., Hsu, W.H.: Cross-age reference coding for age-invariant face recognition and retrieval. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8694, pp. 768–783. Springer, Heidelberg (2014). doi:10.1007/978-3-319-10599-4_49
2. Chen, D., Cao, X., Wen, F., Sun, J.: Blessing of dimensionality: high-dimensional feature and its efficient compression for face verification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3025–3032 (2013)
3. Barkan, O., Weill, J., Wolf, L., Aronowitz, H.: Fast high dimensional vector multiplication face recognition. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1960–1967 (2013)
4. Zhu, K., Gong, D., Li, Z., Tang, X.: Orthogonal Gaussian process for automatic age estimation. In: Proceedings of the ACM International Conference on Multimedia, pp. 857–860. ACM (2014)
5. Lanitis, A., Taylor, C.J., Cootes, T.F.: Toward automatic simulation of aging effects on face images. IEEE Trans. Pattern Anal. Mach. Intell. **24**, 442–455 (2002)

6. Du, J.X., Zhai, C.M., Ye, Y.Q.: Face aging simulation and recognition based on NMF algorithm with sparseness constraints. Neurocomputing **116**, 250–259 (2013)
7. Park, U., Tong, Y., Jain, A.K.: Age-invariant face recognition. IEEE Trans. Pattern Anal. Mach. Intell. **32**, 947–954 (2010)
8. Li, Z., Gong, D., Li, X., Tao, D.: Aging face recognition: a hierarchical learning model based on local patterns selection. IEEE Trans. Image Process. **25**, 2146–2154 (2016)
9. Gong, D., Li, Z., Lin, D., Liu, J., Tang, X.: Hidden factor analysis for age invariant face recognition. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2872–2879 (2013)
10. Yin, Q., Tang, X., Sun, J.: An associate-predict model for face recognition. In: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 497–504. IEEE (2011)
11. Kumar, N., Berg, A.C., Belhumeur, P.N., Nayar, S.K.: Attribute and simile classifiers for face verification. In: 2009 IEEE 12th International Conference on Computer Vision, pp. 365–372. IEEE (2009)
12. Li, Z., Gong, D., Li, X., Tao, D.: Learning compact feature descriptor and adaptive matching framework for face recognition. IEEE Trans. Image Process. **24**, 2736–2745 (2015)
13. Li, Y., Meng, L., Feng, J., Wu, J.: Downsampling sparse representation and discriminant information aided occluded face recognition. Sci. Chin. Inf. Sci. **57**, 1–8 (2014)
14. Fu, Y., Huang, T.S.: Human age estimation with regression on discriminative aging manifold. IEEE Trans. Multimedia **10**, 578–584 (2008)
15. Geng, X., Zhou, Z.H., Smith-Miles, K.: Automatic age estimation based on facial aging patterns. IEEE Trans. Pattern Anal. Mach. Intell. **29**, 2234–2240 (2007)
16. Suo, J., Chen, X., Shan, S., Gao, W.: Learning long term face aging patterns from partially dense aging databases. In: 2009 IEEE 12th International Conference on Computer Vision, pp. 622–629. IEEE (2009)
17. Tsumura, N., Ojima, N., Sato, K., Shiraishi, M., Shimizu, H., Nabeshima, H., Akazaki, S., Hori, K., Miyake, Y.: Image-based skin color and texture analysis/synthesis by extracting hemoglobin and melanin information in the skin. ACM Trans. Graph. (TOG) **22**, 770–779 (2003)
18. Ling, H., Soatto, S., Ramanathan, N., Jacobs, D.W.: Face verification across age progression using discriminative methods. IEEE Trans. Inf. Forensics Secur. **5**, 82–91 (2010)
19. Klare, B., Jain, A.K.: Face recognition across time lapse: on learning feature subspaces. In: 2011 International Joint Conference on Biometrics (IJCB), pp. 1–8. IEEE (2011)
20. Ricanek Jr., K., Tesafaye, T.: MORPH: a longitudinal image database of normal adult age-progression. In: 2006 7th International Conference on Automatic Face and Gesture Recognition, FGR 2006, pp, 341–345. IEEE (2006)
21. Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: a database for studying face recognition in unconstrained environments. Technical report 07–49, University of Massachusetts, Amherst (2007)
22. Heseltine, T., Pears, N., Austin, J.: Evaluation of image preprocessing techniques for eigenface-based face recognition. In: Second International Conference on Image and Graphics, pp. 677–685. International Society for Optics and Photonics (2002)