Cross-Age Reference Coding for Age-Invariant Face Recognition and Retrieval

Bor-Chun Chen[1], Chu-Song Chen[1], and Winston H. Hsu[2]

[1] Institute of Information Science, Academia Sinica, Taipei, Taiwan
[2] National Taiwan University, Taipei, Taiwan

Abstract. Recently, promising results have been shown on face recognition researches. However, face recognition and retrieval across age is still challenging. Unlike prior methods using complex models with strong parametric assumptions to model the aging process, we use a data-driven method to address this problem. We propose a novel coding framework called Cross-Age Reference Coding (CARC). By leveraging a large-scale image dataset freely available on the Internet as a reference set, CARC is able to encode the low-level feature of a face image with an age-invariant reference space. In the testing phase, the proposed method only requires a linear projection to encode the feature and therefore it is highly scalable. To thoroughly evaluate our work, we introduce a new large-scale dataset for face recognition and retrieval across age called Cross-Age Celebrity Dataset (CACD). The dataset contains more than 160,000 images of 2,000 celebrities with age ranging from 16 to 62. To the best of our knowledge, it is by far the largest publicly available cross-age face dataset. Experimental results show that the proposed method can achieve state-of-the-art performance on both our dataset as well as the other widely used dataset for face recognition across age, MORPH dataset.

Keywords: Face Recognition, Aging.

1 Introduction

Face related problems (e.g., face detection, face recognition) are important but challenging, and they have drawn many computer vision researchers' attention for decades. For matching faces, there are four key factors that compromise the accuracy: pose, illumination, expression, and aging [12]. Many researches had been dedicated to solve the face recognition problem with the existence of one or more types of these variations. Recently, due to the improvement of the face and facial landmark detection accuracy as well as the increase of the computational power, researches [4,27,2] show that we can achieve near-human performance on face verification benchmark taken in the unconstrainted environments such as Labeled Faces in the Wild dataset (LFW) [11]. However, as LFW dataset contains large variations in pose, illumination, and expression, it contains little variation in aging. As can be seen in Figure 3 that faces across age can be very different, therefore, face matching with age variation is still very challenging.

D. Fleet et al. (Eds.): ECCV 2014, Part VI, LNCS 8694, pp. 768–783, 2014.

Face recognition and retrieval across age has a wide range of applications. For example, finding missing persons and child trafficking in forensic applications, and automatic photo annotation in personal media. However, as most age-related works on face image analysis focus on age estimation and simulation, works focusing on face recognition and retrieval across age are limited.

By taking advantage of widely available celebrity images on the Internet, we propose a new approach to address this problem with a different angle from prior works. Instead of modeling the aging process with strong parametric assumptions, we adopt a data driven approach and introduce a novel coding method called Cross-Age Reference Coding (CARC). Our basic assumption is that if two people look alike when they are young, they might also look similar when they both grow older. Based on this assumption, CARC leverages a set of reference images available freely from the Internet to encode the low-level features of a face image with an averaged representation in reference space. As shown in Figure 1, two images of the same person will have similar representations using CARC because they both look similar to certain reference people (with different ages), and experimental results with CARC shown in section 5 support this assumption. Since images downloaded from Internet could be noisy, CARC is designed to be robust against such noise. Note that although the idea of using a reference set for face recognition was proposed in other literatures such as [13,39], they did not consider the age variation. The proposed method is essentially different because we incorporate the age information of the reference set into the coding framework.

We notice that benchmarks for evaluating age-invariant face recognition and retrieval are limited because it is hard to collect images of the same person with different ages. In order to thoroughly evaluate our work, we introduce a new cross-age face dataset called Cross-Age Celebrity Dataset (CACD) by collecting celebrity images on the Internet[1]. Because many celebrities are active for a long period, we can easily obtain images of them with different ages. CACD contains more than 160,000 face images of 2,000 celebrities across ten years with age ranging from 16 to 62. To our best knowledge, this is the largest publicly available cross-age face dataset. Examples of the dataset can be found in Figure 3. By conducting extensive experiments, we show that the proposed method can outperform state-of-the-art methods on both MORPH [26] and CACD.

To sum up, contributions of this paper include:

- We propose a new coding framework called CARC that leverages a reference image set (available from Internet) for age-invariant face recognition and retrieval.
- We introduce a new large-scale face dataset, CACD, for evaluating face recognition and retrieval across age. The dataset contains more than 160,000 images with 2,000 people and is made publicly available[2].

[1] Note that although celebrity images were used in other benchmarks such as LFW [11] or Pubfig [13], none of them focus on collecting celebrity images across age.

[2] Available at `http://bcsiriuschen.github.io/CARC/`

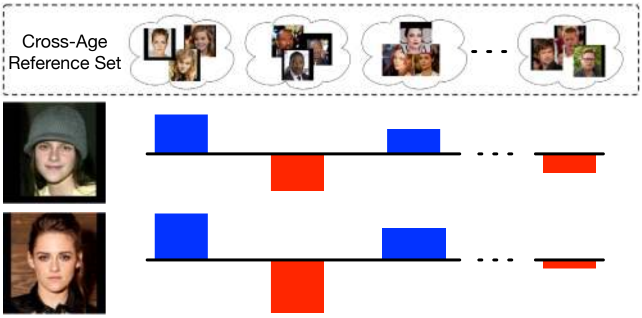


Fig. 1. Each cluster on the top represents the images of one reference person. We use images of n different people as our reference set, and encode each local feature of testing image as an n dimensional feature. Because the reference set contains images with different ages, we can convert each local feature into age-invariant representation using the proposed method. Two images of the same person with different ages will have similar features in the new reference space and therefore we can achieve high accuracy for face recognition and retrieval across age.

– We conduct extensive experiments on MORPH and CACD and show that CARC can outperform state-of-the-art methods on both datasets.

The rest of the paper is organized as follow: section 2 discusses the related work. Section 3 describes the proposed coding framework, Cross-Age Reference Coding. Section 4 introduces our dataset, Cross-Age Celebrity Dataset. Section 5 gives the experimental results, and section 6 concludes this paper.

2 Related Work

2.1 Face Recognition and Retrieval

Face recognition has been investigated for decades by many researchers. Turk and Pentland introduce the idea of eigenface [31] in 1991, which is one of earliest successes in the face recognition research; Ahonen et al. [1] successfully apply texture descriptor, local binary pattern (LBP), on the face recognition problem. Wright et al. [35] propose to use sparse representation derived from training images for face recognition. The method is proved to be robust against occlusions for face recognition. Recently, Chen et al. [4] use a high dimensional version of LBP and achieve near-human performance on the LFW dataset.

Some researches also use a reference set to improve the accuracy of face recognition and retrieval. Kumar et al. [13] propose to use attribute and simile classifier, SVM classifier trained on reference set, for face verification. Berg et al. [3] further improve the method by using "Tom-vs-Pete" classifier. Yin et al. [39] propose an associate-predict model using 200 identities in Multi-PIE dataset [9] as a reference set. Wu et al. [36] propose an identity-based quantization using a dictionary constructed by 270 identities for large-scale face image retrieval.

Although these methods achieve salient performance on face recognition, they do not work well when the age variation exists because they do not consider the age information in the reference set.

2.2 Age-Invariant Face Recognition

Most existing age-related works for face image analysis focus on age estimation [14,15,40,25,34,7,38,6,10,22,21] and age simulation [16,30,28,29,24]. In recent years, researchers have started to focus on face recognition across age. One of the approaches is to construct 2D or 3D aging models [16,7,24] to reduce the age variation in face matching. Such models usually rely on strong parametric assumptions, accurate age estimation, as well as clean training data, and therefore they do not work well in unconstrainted environments. Some other works focus on discriminative approaches. Ling et al. [19] use gradient orientation pyramid with support vector machine for face verification across age progression. Li et al. [17] use multi-feature discriminant analysis for close-set face identification. Gong et al. [8] propose to separate the feature into identity and age components using hidden factor analysis. Different from the above methods, we propose to adopt a data-driven approach to address this problem. By taking advantage of a cross-age reference set freely available on the Internet, and using a novel coding framework called CARC, we are able to achieve high accuracy for face recognition and retrieval with age variation.

2.3 Face Dataset

There are many face datasets available for researches in face recognition. Among all the datasets, LFW [11] is one of the most popular dataset for face verification task in unconstrainted environments, and it contains 13,233 images of 5,749 people extracted from the news program. Pubfig [13] is another dataset collected in the unconstrainted environments. It aims to improve the LFW dataset by providing more images for each individual, and it contains 58,797 images with 200 people. For age estimation and face recognition across age, FG-NET [20] and MORPH [26] are the two most widely used datasets. FG-NET contains 1,002 images of 82 people with age range from 0 to 69. MORPH contains 55,134 images of 13,618 people with age range from 16 to 77. Information and comparison of these datasets can be found in Table 1 and Figure 5. Compared to existing datasets, our dataset contains a larger number of images of different people in different ages.

3 Cross-Age Reference Coding (CARC)

3.1 System Overview

Figure 2 shows the system overview of the proposed method. For every image in the database, we first apply a face detection algorithm to find the locations of the

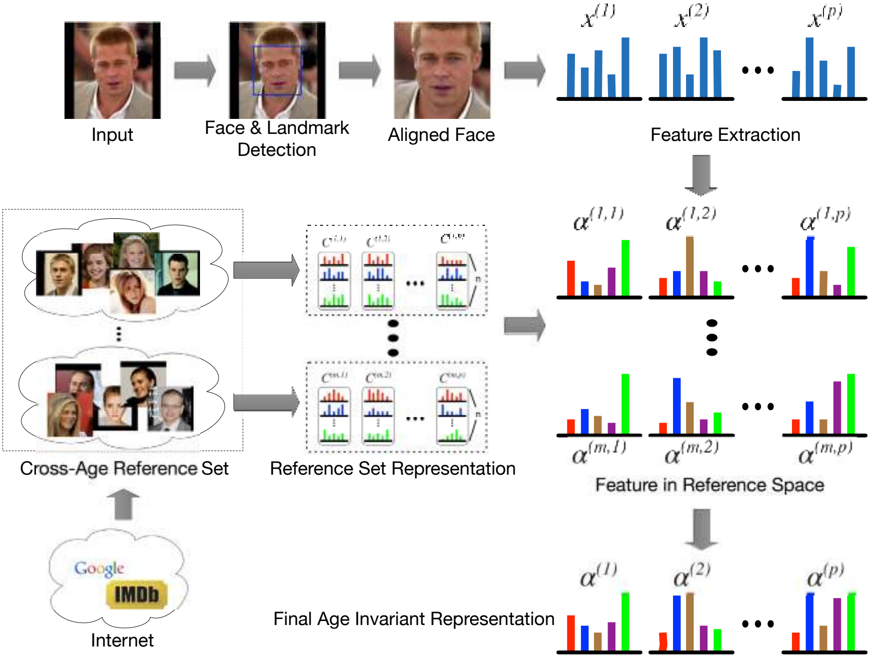


Fig. 2. System overview of the proposed method. For each image, we first apply face and facial landmarks detection. We then extract local feature (high-dimensional LBP) from each landmark, and use CARC to encode the local features into the age-invariant representation with three steps. First, by using a cross-age reference set collected from Internet, we compute the reference set representations. Second, we map the local feature extracted from the images into reference space in each of the m years. Finally, we aggregate the m features from different years into a final age-invariant representation for the input image. The final representation is $n \times p$ dimensions where n is the number of reference people and p is the number of facial landmarks. Details of CARC are described in section 3.

faces in the image. We adopt the widely used Viola-Jones face detector [32] for the task. For each face, we then locate sixteen different facial landmarks using face alignment algorithm. Xiong et al. [37] recently propose a supervised decent method for face alignment. Their method uses supervised learning to replace the expensive computation in second order optimization schemes and can efficiently locate the landmarks with high accuracy; therefore we adopt their method to locate the facial landmarks. After landmarks finding, the face is aligned using similarity transform to make the eyes located at the same horizontal positions for all images.

After the face is aligned, we extract local features from each landmark. Among all kinds of different local features, high-dimensional local binary pattern [4] has shown promising results in face verification task. Therefore, we adopt a similar pipeline to extract local features from face images. Around each of these sixteen landmarks, we crop a fixed-size patch with 5 different scales. Each patch is then divided into 4 × 4 cells, and we extract a 59-dimensional uniform local binary pattern [23] from each cell. Features extracted from the same landmarks are then concatenated together as a descriptor for the landmark. The feature dimension for each landmark is 4,720. We use principal component analysis (PCA) to reduce the dimension to 500 for each landmark for further processing.

We then apply CARC to encode the local features into age-invariant representation. CARC contains three main steps: (1) computing reference set representations for different reference people in different years using age-varying reference images obtained from the Internet (cf. section 4), (2) encoding local features into reference space using the reference set representations, and (3) aggregating the features found in step 2 to yield a final age-invariant representation. The following sections will describe each step in detail.

3.2 Reference Set Representations

Using the local features extracted from images of the reference people, we can compute the reference set representations using the follow equation:

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

where $C_i^{(j,k)} \in R^d$ is the reference representation of the person i in year j at landmark k and n is the number of reference people, m is the number of years, p is the number of landmarks. It is computed by averaging over all the features ($x^{(k)}$) from the same reference person in the same years, N_{ij} is the total number of such images. Because the reference set is directly obtained from the Internet, it might contain noise. Taking average is helpful to compute a representation more robust to such noisy data.

3.3 Encoding Feature into the Reference Space

Given a set of n reference person representation $C^{(j,k)} = [C_1^{(j,k)}, C_2^{(j,k)}, \ldots C_n^{(j,k)}]$ and a new feature $x^{(k)}$ extracted at landmark k, we want to use the reference representation to encode the new feature. To achieve this goal, we first define a vector $\alpha^{(j,k)} \in R^{n\times 1}$, which represents the relationship to n reference people (as shown in Figure 1) in year j for the feature extracted at landmark k. We know that $\alpha_i^{(j,k)}$ should be big if the testing feature $x^{(k)}$ is close to the i_{th} reference person, and small otherwise. Here we consider finding such representation by solving a least squared problem with Tikhonov regularization:

$$\underset{\alpha^{(j,k)}}{minimize} \left\| x^{(k)} - C^{(j,k)}\alpha^{(j,k)} \right\|^2 + \lambda \left\| \alpha^{(j,k)} \right\|^2, \quad \forall j, k, \tag{2}$$

However, it does not consider the temporal relationship between representations across different years, whereas one person is similar to a reference person at year j, he/she is most likely similar to the same reference person at adjacent years $j-1$ and $j+1$. Therefore, we add a temporal constraint to reflect this issue in our coding scheme. We first define a tridiagonal matrix L as follow:

$$L = \begin{bmatrix} 1 & -2 & 1 & 0 \dots & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 \dots & 1 & -2 & 1 \end{bmatrix} \in R^{(m-2)\times m}. \quad (3)$$

L is a smoothness operator for the temporal constraint to make $\alpha_i^{(j,k)}$ similar to $\alpha_i^{(j+1,k)}$ and $\alpha_i^{(j-1,k)}$ by minimizing their difference. Let:

$$A^{(k)} = [\alpha^{(1,k)}, \alpha^{(2,k)}, \dots, \alpha^{(m,k)}] \in R^{n\times m}, \quad \forall k. \quad (4)$$

The testing features $x^{(k)}$ can now be cast to the new reference space by minimizing the following objective function with Tikhonov regularization and temporal smoothing:

$$\underset{A^{(k)}}{minimize} \sum_{j=1}^{m} \left(\left\| x^{(k)} - C^{(j,k)}\alpha^{(j,k)} \right\|^2 + \lambda_1 \left\| \alpha^{(j,k)} \right\|^2 \right) + \lambda_2 \left\| LA^{(k)T} \right\|^2, \quad \forall k. \quad (5)$$

The first term in the above equation is to make sure the reconstruction error in reference space is small, and the second term is to let the coefficients of the same reference person across adjacent years become similar.

Solving Equation 5 is simple because it is a $l2$-regularized least-squared problem. We first define new matrices $\hat{C}^{(k)}$ and $\hat{L}$ as follows:

$$\hat{C}^{(k)} = \begin{bmatrix} C^{(1,k)} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & C^{(2,k)} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & C^{(m,k)} \end{bmatrix} \in R^{md\times mn}, \quad \forall k \quad (6)$$

$$\hat{L} = \begin{bmatrix} I & -2I & I & \mathbf{0} \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I & -2I & I \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \dots & I & -2I & I \end{bmatrix} \in R^{(m-2)n\times mn}, \quad (7)$$

and we define the vector $\hat{\alpha}^{(k)} = [\alpha^{(1,k)T}, \alpha^{(2,k)T}, \dots, \alpha^{(m,k)T}]^T \in R^{mn}$ and $\hat{x}^{(k)} = [x^{(k)T}, \dots, x^{(k)T}] \in R^{md}$. We can now rewrite Equation 5 as:

$$\underset{\hat{\alpha}^{(k)}}{minimize} \left\| \hat{x}^{(k)} - \hat{C}^{(k)}\hat{\alpha}^{(k)} \right\|^2 + \hat{\lambda}_1 \left\| \hat{\alpha}^{(k)} \right\|^2 + \hat{\lambda}_2 \left\| \hat{L}\hat{\alpha}^{(k)} \right\|^2, \quad \forall k \quad (8)$$

which is a standard regularized least square problem with a closed-form solution:

$$\hat{\alpha}^{(k)} = \left(\hat{C}^{(k)T}\hat{C}^{(k)} + \hat{\lambda}_1 I + \hat{\lambda}_2 \hat{L}^T \hat{L}\right)^{-1} \hat{C}^{(k)T}\hat{x}^{(k)}, \quad \forall k. \tag{9}$$

Let $\hat{P}^{(k)} = \left(\hat{C}^{(k)T}\hat{C}^{(k)} + \hat{\lambda}_1 I + \hat{\lambda}_2 \hat{L}^T \hat{L}\right)^{-1} \hat{C}^{(k)T}$, we can precompute $\hat{P}^{(k)}$ as a projection matrix so that when a query image comes, we can efficiently map it to the reference set space via a linear projection.

3.4 Aggregating Representation across Different Years

We want to aggregate the representations in reference space across different years. Here we propose to use max pooling to achieve the goal:

$$\alpha_i^{(k)} = \max\left(\alpha_i^{(1,k)}, \alpha_i^{(2,k)}, \ldots, \alpha_i^{(m,k)}\right), \quad \forall i, k. \tag{10}$$

By using max pooling, the final representation will have a high response to one reference person as long as it has a high response to the person in any year. So when there are two images of the same person at different ages, the younger image might have a high response at a certain reference celebrity in an early year, while the older image might have a high response at the same celebrity but in a later year. The final representations for these two images will both have high response at that specific reference person because of the max pooling aggregation. Therefore, we can achieve age-invariant face recognition and retrieval. Note that [18] provides a theoretical support for the use of max pooling with reference images.

After we obtain the final representation, we use cosine similarity to compute the matching scores between images for face recognition and retrieval.

4 Cross-Age Celebrity Dataset (CACD)

4.1 Celebrity Name Collection

We first form a list of celebrity names that we want to include in the dataset. In order to create a large-scale dataset with diversity in ages, we identify two important criteria to decide whose name should be on the list: (1) the people on the list should have different ages, and (2) these people must have many images available on the Internet for us to collect. We select our names from an online celebrity database, IMDb.com[3], and the former criterion is satisfied by collecting names with different birth years; while the later one is satisfied by collecting names of popular celebrities. In detail, we collect names of celebrities whose birth date is from 1951 to 1990. In this 40 years period, we collect the names of top 50 popular celebrities from each birth year with 2,000 names in total. Similar approach is adopted in [33] to collect celebrity names.

[3] IMDb.com is one of the largest online movie database, and it contains profiles of millions of movies and celebrities.

Fig. 3. Examples of images collected. Numbers on the top are the birth years of the celebrities, and numbers on the left indicate the years in which the images were taken. Images in the same column are of the same celebrity.

4.2 Image Collection

We use Google Image Search to collect images. In order to collect celebrities images across different ages, we use a combination of celebrity name and year as keywords. For example, we use “Emma Watson 2004” as keywords to retrieve Emma Watson’s images taken in 2004. These might include photos taken in an event held in 2004 or images from a 2004 movie such as “Harry Potter and the Prisoner of Azkaban.” For each celebrity, we collect images across ten years from 2004 to 2013. Since we already know the birth years of the celebrities, we can calculate the ages of celebrities in the images by simply subtract the birth year from the year of which the photo was taken. Examples of images collected can be found in Figure 3. Note that the dataset might contain noise because we could accidentally collect images of other celebrities in the same event or movie, and some of the celebrities might retire from the public during some periods between 2004 to 2013 and have little photos available at that time, which could make year few labels incorrect. Nevertheless, the proposed coding method is robust to such noise and proved to have good performance in our experiments (cf. section 5).

4.3 Dataset Statistics

After applying face detection [32] to all images, there are more than 200,000 images containing faces for all 2,000 celebrities. However, some of the images are duplicated. We use a simple duplicate detection algorithm based on low-level features to remove such images. After removing duplicate images, we have around 160,000 face images left. For a subset of 200 celebrities, we manually check

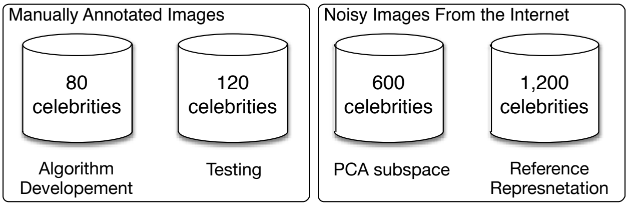


Fig. 4. CACD contains images of 2,000 celebrities. 200 of them are manually annotated. We randomly separate these 200 celebrities into two subset: 80 of them are used for algorithm development and parameter selection, and the other 120 are used for testing. 600 out of 1,800 celebrities without annotation are used for computing the PCA subspace. Another 1,200 is used as reference set.

the images and remove the noisy images in the dataset[4]. These 200 celebrities are used for algorithm development and evaluation. More specifically, we further separate images of these 200 celebrities into two subsets. One of them contains 80 celebrities and should be used for algorithm development and parameter selection; the other 120 celebrities are for testing. The protocol for using the dataset in our experiments are shown in Figure 4. The dataset contains 163,446 images of 2,000 celebrities after removing the noisy images, which is the largest publicly available cross-age dataset to the best of our knowledge. Table 1 shows the statistics of the dataset and comparison to other existing face datasets. From this table, our dataset has the largest amount of images and contains age information. Compared to MORPH dataset, age gaps for CACD between images of the same person are larger. FG-NET has larger age gaps but it only contains few images from a limited number of people. Figure 5 shows the distribution of the datasets with different ages. Both MORPH and our dataset do not contain images with age of 10 or younger, while FG-NET has more images of younger ages. However, our dataset has more images for all other ages.

5 Experiments

5.1 Experiment on Cross-Age Celebrity Dataset

In Cross-Age Celebrity Dataset, there are images of a total of 2,000 celebrities. We separate the dataset into four parts as shown in Figure 4: images of 200 celebrities with manual annotations are used for evaluating the algorithms: (1) 80 out of these 200 celebrities are used for parameters selection and (2) the other 120 are used for reporting testing results; (3) images of another 600 celebrities are used for computing the PCA subspace; (4) the final images of 1,200 celebrities are used for reference representations.

[4] Note that we manually removed noisy images from these 200 celebrities by checking the image content. However, since some of the images are hard to identify even for humans, the subset might contain noises. Also, we only employ simple duplicate detection method, thus the dataset might still have near-duplicate images.

Table 1. The comparison between existing datasets. Our dataset has the largest amount of images and contains age information. Compared to MORPH dataset, age gaps between images of the same person are larger. FG-NET has larger age gaps but it only contains a small amount of images from a limited number of people.

Dataset	# of images	# of people	# images/person	Age info.	Age gap
LFW [11]	13,233	5,749	2.3	No	-
Pubfig [13]	58,797	200	293.9	No	-
FGNet [20]	1,002	82	12.2	Yes	0-45
MORPH [26]	55,134	13,618	4.1	Yes	0-5
Ours (CACD)	163,446	2,000	81.7	Yes	0-10

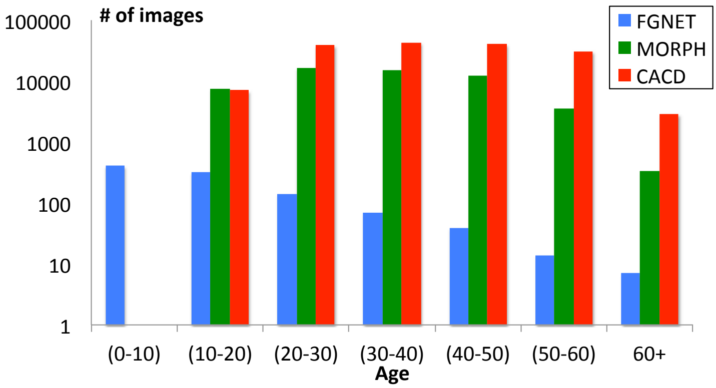
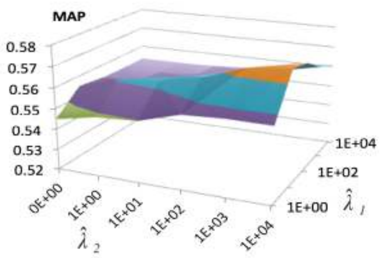


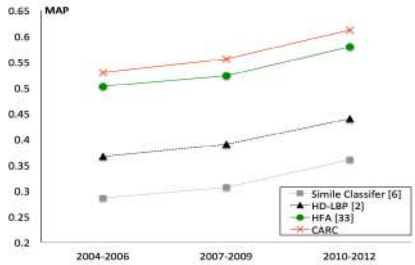
Fig. 5. The distribution of the dataset with different ages. Both MORPH and our dataset do not contain images with age of 10 or younger, while FG-NET has more images of younger ages. However, our dataset has more images in all other ages.

Parameters Selection. For selecting the parameters in our algorithm, we use images taken in 2013 as query images and images taken in other years (2004-2012) as database images. Mean average precision (MAP) is used as our evaluation metrics. For the retrieval results of each query image, precision at every recall level is computed and averaged to get average precision (AP). MAP is then computed by averaging AP from all query images. There are few parameters in the proposed method we need to decide: the PCA feature dimensions d, regularization parameters in the coding framework $\hat{\lambda}_1$, $\hat{\lambda}_2$, and number of reference celebrities used n.

For PCA feature dimensions d, we run experiments from 100 to 1,000 and find that the performance stops to improve after 500, so we fix our feature dimension d to 500 in the further experiments. For regularization parameters and number of celebrities, we first randomly select half reference celebrities and adjust $\hat{\lambda}_1$ and $\hat{\lambda}_2$ from 10^0 to 10^4. The results are shown in Figure 6 (a). We can see that adding temporal constraint in our coding framework can help the performance



(a)



(b)

Fig. 6. (a) Validation results of Cross-Age Reference Coding on our CACD using different parameters. The results show using temporal constraint can improve the performance. The parameters are set to $(\hat{\lambda}_1, \hat{\lambda}_2) = (10^1, 10^4)$. (b) The retrieval results on CACD compare to other state-of-the-art methods. The proposed method consistently shows the best performance across different years.

Table 2. Rank-1 identification results on MORPH dataset. The proposed method achieves the highest recognition rate compared to other state-of-the-art methods.

Method	Recognition rate
Park et al. (2010) [24]	79.8%
Li et al. (2011) [17]	83.9%
Gong et al. (2013) [8]	91.1%
Ours	**92.8%**

as it increases with $\hat{\lambda}_2$. We set $(\hat{\lambda}_1,\ \hat{\lambda}_2) = (10^1, 10^4)$ where they achieve the best performance in the validation set for testing. For the number of celebrities n, we then randomly select reference celebrities from 40 to 1,200 and find that the performance stops to improve after 600. Therefore we fix the number of celebrities to 600 for testing.

Compare to State-of-the-Art Methods. We compare CARC to several different state-of-the-art methods, including: (1) high-dimensional local binary pattern [4] (HD-LBP), the local feature used for cross-age celebrity coding. We also use PCA to reduce the dimension to 500 for features from each landmark. (2) Simile Classifier [13]: we train a linear-SVM for each reference celebrity and use the sign distance to the decision boundary as feature. We use LIBLINEAR package [5] to carry out the computation and the number of the reference celebrities is also set as 600. (3) Hidden Factor Analysis [8] (HFA), a state-of-the-art method for age-invariant face recognition. We use high-dimensional local binary pattern as input feature and the parameters are tuned to the best setting according to their paper. We use the images of 120 celebrities for testing. We conduct experiments with three different subsets. In all three subsets, images taken in 2013 are

Fig. 7. Some cases where the proposed method fails. The first row contains the probe images, the second row shows the rank-1 result using the proposed method, and the third row shows the correct match in the gallery. The number on the bottom shows the age of the image.

used as query images. The database contains images taken in 2004-2006, 2007-2009, 2010-2012 for each of the three subsets respectively. For all methods, we use cosine similarity as the similarity measure. The performance of all methods in terms of MAP is shown in Figure 6(b). We can see that the proposed method outperforms other methods in all three different subsets. Simile Classifier has the worst performance. It is because SVM classifier is not robust to noise and age variation in the training data. The performance drops on all methods when the age gap is larger, which reveals the difficulty of face retrieval with age variation. Nevertheless, both HFA and the proposed method can achieve higher performance on the subset with large age gap compared to baseline features on the subset with small age gap. It shows the effectiveness of the age-invariant methods. CARC achieves higher performance than HFA, which shows that CARC can better utilize the noisy reference set and is more robust to age variation.

5.2 Experiment on MORPH Dataset

To show the generalizability of the proposed method, we also conduct face recognition experiment on MORPH dataset. For this dataset, we follow the experimental setting in [17] for close set face identification. We randomly select 10,000 subjects in the MORPH dataset and use the youngest images of these 10,000 subjects to construct the gallery set, and the oldest images of the subjects are used as the probe set. Both gallery and probe consist of 10,000 images from 10,000 different subjects. We then randomly select another 600 subjects from the dataset as reference set for our algorithm. Images of subjects other than these 10,600 subjects are then used to compute PCA and LDA subspaces as in [8], and we reduce the dimension to 1,000 for features from each landmark.

We compare our algorithm to several state-of-the-art methods including, (1) a generative aging model [24], (2) a discriminative model for age-invariant face recognition [17], and (3) hidden factor analysis, currently the best result on the dataset [8]. The results in terms of rank-1 recognition rate of our algorithm compared to other methods are shown in Table 2. We can see that the proposed method can achieve better performance compared to other state-of-the-art methods. Some examples of incorrect matching are shown in Figure 7. Although our system can achieve an accuracy higher than 92%, it can still fail in some cases, especially when the probe and gallery are significantly different. Some of these cases are really hard even for human to recognize. For some applications, we do not need to have perfect rank-1 accuracy, and we only need to find the correct match in the top-k results. Our system can achieve more than 98% accuracy in top-20 results and 94.5% mean average precision in the MORPH dataset.

6 Conclusions

In this paper, we propose a new approach for age-invariant face recognition and retrieval called Cross-Age Reference Coding. Using the proposed method, we can map low-level feature into an age-invariant reference space. Experimental results show that the proposed method can outperform state-of-the-art methods on both MORPH and CACD datasets and achieve high accuracy in face recognition across age. We also introduce a large-scale face dataset, Cross-Age Celebrity Dataset, for the purpose of face recognition with age variation. To the best of our knowledge, the dataset is the largest publicly available cross-age face dataset, and we hope the dataset can help researchers to improve the result of face recognition. In the future, we want to investigate how to effectively choose a subset from the reference people and further improve the performance of age-invariant face recognition and retrieval.

References

1. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: Application to face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(12), 2037–2041 (2006)
2. Barkan, O., Weill, J., Wolf, L., Aronowitz, H.: Fast high dimensional vector multiplication face recognition. In: Proc. IEEE Int'l Conf. Computer Vision (2013)
3. Berg, T., Belhumeur, P.N.: Tom-vs-pete classifiers and identity-preserving alignment for face verification. In: BMVC, vol. 1, p. 5 (2012)
4. Chen, D., Cao, X., Wen, F., Sun, J.: Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3025–3032. IEEE (2013)
5. Fan, R.E., Chang, K.W., Hsieh, C.J., Wang, X.R., Lin, C.J.: Liblinear: A library for large linear classification. The Journal of Machine Learning Research 9, 1871–1874 (2008)
6. Fu, Y., Huang, T.S.: Human age estimation with regression on discriminative aging manifold. IEEE Transactions on Multimedia 10(4), 578–584 (2008)
7. Geng, X., Zhou, Z.H., Smith-Miles, K.: Automatic age estimation based on facial aging patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(12), 2234–2240 (2007)

8. Gong, D., Li, Z., Lin, D., Liu, J., Tang, X.: Hidden factor analysis for age invariant face recognition. In: 2013 IEEE 14th International Conference on Computer Vision. IEEE (2013)
9. Gross, R., Matthews, I., Cohn, J., Kanade, T., Baker, S.: Multi-pie. Image and Vision Computing 28(5), 807–813 (2010)
10. Guo, G., Fu, Y., Dyer, C.R., Huang, T.S.: Image-based human age estimation by manifold learning and locally adjusted robust regression. IEEE Transactions on Image Processing 17(7), 1178–1188 (2008)
11. Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Tech. Rep. 07-49, University of Massachusetts, Amherst (October 2007)
12. Jain, A.K., Klare, B., Park, U.: Face matching and retrieval in forensics applications. IEEE MultiMedia 19(1), 20 (2012)
13. 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)
14. Kwon, Y.H., da Vitoria Lobo, N.: Age classification from facial images. In: Proceedings of the 1994 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 1994, pp. 762–767. IEEE (1994)
15. Lanitis, A., Draganova, C., Christodoulou, C.: Comparing different classifiers for automatic age estimation. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 34(1), 621–628 (2004)
16. Lanitis, A., Taylor, C.J., Cootes, T.F.: Toward automatic simulation of aging effects on face images. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(4), 442–455 (2002)
17. Li, Z., Park, U., Jain, A.K.: A discriminative model for age invariant face recognition. IEEE Transactions on Information Forensics and Security 6(3), 1028–1037 (2011)
18. Liao, Q., Leibo, J.Z., Mroueh, Y., Poggio, T.: Can a biologically-plausible hierarchy effectively replace face detection, alignment, and recognition pipelines? CoRR abs/1311.4082 (2013)
19. Ling, H., Soatto, S., Ramanathan, N., Jacobs, D.W.: Face verification across age progression using discriminative methods. IEEE Transactions on Information Forensics and Security 5(1), 82–91 (2010)
20. Face and Gesture Recognition Working group: FG-NET Aging Database (2000)
21. Montillo, A., Ling, H.: Age regression from faces using random forests. In: 2009 16th IEEE International Conference on Image Processing (ICIP), pp. 2465–2468. IEEE (2009)
22. Mu, G., Guo, G., Fu, Y., Huang, T.S.: Human age estimation using bio-inspired features. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009, pp. 112–119. IEEE (2009)
23. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(7), 971–987 (2002)
24. Park, U., Tong, Y., Jain, A.K.: Age-invariant face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 32(5), 947–954 (2010)
25. Ramanathan, N., Chellappa, R.: Face verification across age progression. IEEE Transactions on Image Processing 15(11), 3349–3361 (2006)
26. Ricanek, K., Tesafaye, T.: Morph: A longitudinal image database of normal adult age-progression. In: 7th International Conference on Automatic Face and Gesture Recognition, FGR 2006, pp. 341–345. IEEE (2006)

27. Simonyan, K., Parkhi, O.M., Vedaldi, A., Zisserman, A.: Fisher vector faces in the wild. In: Proc. BMVC, vol. 1, p. 7 (2013)
28. 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)
29. Suo, J., Zhu, S.C., Shan, S., Chen, X.: A compositional and dynamic model for face aging. IEEE Transactions on Pattern Analysis and Machine Intelligence 32(3), 385–401 (2010)
30. 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 Transactions on Graphics (TOG) 22, 770–779 (2003)
31. Turk, M.A., Pentland, A.P.: Face recognition using eigenfaces. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 1991, pp. 586–591. IEEE (1991)
32. Viola, P., Jones, M.J.: Robust real-time face detection. International Journal of Computer Vision 57(2), 137–154 (2004)
33. Wang, D., Hoi, S.C., He, Y., Zhu, J.: Retrieval-based face annotation by weak label regularized local coordinate coding. In: Proceedings of the 19th ACM International Conference on Multimedia, pp. 353–362. ACM (2011)
34. Wang, J., Shang, Y., Su, G., Lin, X.: Age simulation for face recognition. In: 18th International Conference on Pattern Recognition, ICPR 2006, vol. 3, pp. 913–916. IEEE (2006)
35. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y.: Robust face recognition via sparse representation. IEEE Transactions on Pattern Analysis and Machine Intelligence 31(2), 210–227 (2009)
36. Wu, Z., Ke, Q., Sun, J., Shum, H.Y.: Scalable face image retrieval with identity-based quantization and multireference reranking. IEEE Transactions on Pattern Analysis and Machine Intelligence 33(10), 1991–2001 (2011)
37. Xiong, X., De la Torre, F.: Supervised descent method and its applications to face alignment. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 532–539. IEEE (2013)
38. Yan, S., Wang, H., Tang, X., Huang, T.S.: Learning auto-structured regressor from uncertain nonnegative labels. In: IEEE 11th International Conference on Computer Vision, ICCV 2007, pp. 1–8. IEEE (2007)
39. 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)
40. Zhou, S.K., Georgescu, B., Zhou, X.S., Comaniciu, D.: Image based regression using boosting method. In: Tenth IEEE International Conference on Computer Vision, ICCV 2005, vol. 1, pp. 541–548. IEEE (2005)