Face Recognition and Retrieval using Cross-Age Reference Coding with Cross-Age Celebrity Dataset

Bor-Chun Chen, Chu-Song Chen, Winston H. Hsu

***Abstract*—This paper introduces a new method for face recognition across age and also a dataset containing variations of age in the wild. Instead of using complex models with strong parametric assumptions to model the aging process, we use a data-driven method to address the cross-age face recognition problem, called Cross-Age Reference Coding (CARC). By leveraging a large-scale image dataset freely available on the Internet as a reference set, CARC can encode the low-level feature of a face image with an age-invariant reference space. In the retrieval phase, our method only requires a linear projection to encode the feature and thus it is highly scalable. To evaluate our method, we introduce a new large-scale dataset 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 our best knowledge, it is by far the largest publicly available cross-age face dataset. Experimental results show that our method can achieve state-of-the-art performance on both CACD and the other widely used dataset for face recognition across age. In order to understand the difficulties of face recognition across age, we further construct a verification subset from the CACD called CACD-VS and conduct human evaluation using Amazon Mechanical Turk. CACD-VS contains 2,000 positive pairs and 2,000 negative pairs and is carefully annotated by checking both of the associated image and web contents. Our experiments show that although state-of-the-art methods can achieve competitive performance compared to average human performance, majority votes of several humans can achieve much higher performance on this task. The gap between machine and human would imply possible directions for further improvement of cross-age face recognition in the future.**

***Index Terms*—Face Recognition, face image retrieval, cross-age face recognition**

I. Introduction

FACE recognition or retrieval has long been an important topic in computer vision and multimedia. There are four key factors affecting the accuracy: pose, illumination, expression, and aging [1]. Many previous studies have devoted to solving the face recognition problem with respect to one or more types of these factors. Recently, due to the improvement of face and facial landmark detection accuracies and increase of the computational power, many researches [2], [3], [4] show that near-human performance can be achieved on face verification benchmark taken in the unconstrained environments such as Labeled Faces in the Wild dataset (LFW) [5]. However, as LFW dataset contains large variations in pose, illumination, and expression, it contains little variation in aging. As can be seen in Figure 1 that faces across age can be very different, therefore, face matching or retrieval under aging changes is still very challenging. Besides, as most age-related works focus on age estimation and simulation, works focusing on face recognition and retrieval across age are still few.

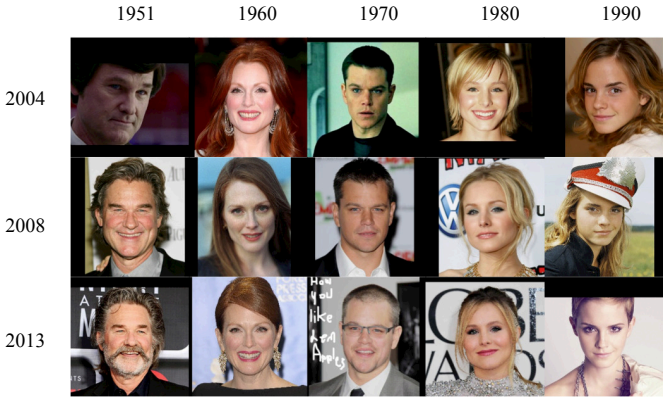

Fig. 1. Examples of face images across age. Top row numbers are the birth years of the celebrities, and left column numbers indicate the years in which the images were taken. Images in the same column are of the same celebrity.

By taking advantage of widely available celebrity images on the Internet, we introduce a new approach to address this problem in a different way from previous studies. 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 2, 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 V 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 [6], [7], they did not consider the age variation. The proposed method is essentially different because we incorporate the age information of the

B.-C. Chen is with Institute of Information Science, Academia Sinica (e-mail: sirius@umd.edu). C.-S. Chen is with the Research Center for Information Technology Innovation (CITI) and Institute of Information Science (IIS), Academia Sinica, Taipei, Taiwan (e-mail: song@iis.sinica.edu.tw). W. H. Hsu is with the Graduate Institute of Networking and Multimedia and the Department of Computer Science and Information Engineering, National Taiwan University, Taipei 10617, Taiwan (e-mail: winston@csie.ntu.edu.tw).

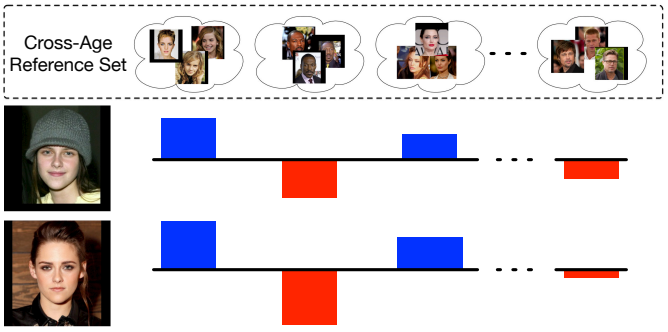

Fig. 2. Each cluster on the top represents the images of one reference person. Our approach uses images of n different people as a reference set, and encode the local features of a testing image as an n dimensional feature. Because the reference set contains images with different ages, we can convert each local feature into an 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 our approach can achieve high accuracy for face recognition and retrieval across age.

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 . 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 1. By conducting extensive experiments, we show that the proposed method can outperform state-of-the-art methods on both MORPH [8] and CACD datasets.

We further develop a verification subset called CACD-VS from CACD for evaluating human performance on cross-age face recognition. CACD-VS contains 2,000 positive pairs (images of the same person across age) and 2,000 negative pairs. Since it is sometimes hard even for humans to recognize people across age, CACD-VS is carefully annotated based on both image and the associated web contents. From our experiments, the proposed method can achieve similar performance compared to average human performance on face verification. However, by aggregating the decisions from multiple humans together, human can achieve higher performance and therefore it suggests that there is still a room to improve for the task of face recognition across age.

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[1].
- We conduct extensive experiments on MORPH and CACD and show that CARC can outperform state-of-the-art methods on both datasets.
- We further construct a verification subset from CACD called CACD-VS, and conduct human experiments on CACD-VS using Amazon Mechanical Turk. Our studies show several interesting findings including (1) relationship between acquaintance to the subject and human recognition performance, and (2) recognition performance by aggregating multiple human results. We also show the performance comparisons between human experiments and state-of-the-art age-invariant face-retrieval methods, which provide useful hints for future improvements.

The rest of the paper is organized as follow: section II discusses the related work. Section III describes the proposed coding framework, CARC. Section IV introduces our dataset, CACD. Section V gives the experimental results, including those on MORPH and CACD, as well as human performance on CACD-VS, and section VI concludes this paper.

II. Related Work

A. Face Recognition and Retrieval

Face recognition and retrieval have been investigated for a long time in many studies. A thorough survey of this topic is beyond the scope of this paper, and we refer the readers to the survey papers/books [9], [10] for a comprehensive review of this problem. Below we only give a concise survey of several important methods related to our work. Turk and Pentland introduce the idea of eigenface [11] in 1991, which is one of earliest successes in the face recognition research; Ahonen et al. [12] successfully apply the texture descriptor, local binary pattern (LBP), on the face recognition problem. Wright et al. [13] 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. [2] 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. [6] propose to use attribute and simile classifiers, SVM classifiers trained on reference set, for face verification. Berg et al. [14] further improve the method by using "Tom-vs-Pete" classifier. Yin et al. [7] propose an associate-predict model using 200 identities in Multi-PIE dataset [15] as a reference set. Wu et al. [16] 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.

B. Age-Invariant Face Recognition

Most existing age-related works for face image analysis focus on age estimation [17], [18], [19] and age simulation [20], [21], [22]. In recent years, researchers have started to

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

focus on face recognition across age. One of the approaches is to construct 2D or 3D aging models [23], [24], [22] 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 unconstrained environments. In [25], Wu et al. propose to use a relative craniofacial growth model to model the face shapes for cross-age face recognition and it yields good performance on FG-NET dataset. However, their approach requires age information to predict the new shapes, which is not always available. Some other works focus on discriminative approaches. Ling et al. [26] use gradient orientation pyramid with SVM for face verification across age progression. Li et al. [27] use multi-feature discriminant analysis for close-set face identification. Gong et al. [28] 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 in face recognition and retrieval with age variation.

The preliminary results have been published in [29], and the contributions are presented here as a whole. The extensions in this work include: (1) Verification dataset: We construct a verification subset called CACD-VS. CACD-VS contains 4,000 image pairs across ages, and it is constructed by carefully checking both web and image contents. (2) Human evaluations for cross-age face recognition: We use Amazon Mechanical Turk to conduct experiments on face verification in CACD-VS, in order to understand the difficulties of face recognition across age. (3) More in-depth experiments in face retrieval, verification, and identification and the comparison between human and machine performance.

C. Face Dataset

Face datasets can be roughly divided into two categories, datasets collected in controlled environments and datasets from unconstrained environments. In both categories, there are many face datasets available for researches in face recognition. For datasets in controlled environments, FERET [30], Yale, and CMU PIE are some of the popular datasets. For datasets in unconstrained environment, LFW [5] is one of the most popular datasets for face verification task, and it contains 13,233 images of 5,749 people extracted from news programs. Pubfig [6] is another dataset collected in the unconstrained 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 [31] and MORPH [8] 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 I and Table II. Compared to existing datasets, our dataset contains a larger number of images of different people in different ages.

III. Cross-Age Reference Coding (CARC)

A. System Overview

Figure 3 gives an illustration of the proposed method. For every image in the database, we first apply a face detection algorithm to find the face regions 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 a face alignment algorithm. Xiong et al. [33] 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. Sixteen landmarks including eyes' corners, nose tip, mouth corners, are used. After landmark detection, we use the eye locations to align the face images. Images are first rotated so that the two eyes are horizontally even. We then compute the distance between two eyes, and use the obtained distance d to crop a rectangular region of the rotated face with $4 \times d$ in width and height (cf. Figure 3).

After face alignment, we extract local features from each landmark. Among all kinds of different local features, high-dimensional local binary pattern [2] has shown promising results in face verification. 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 $\times$ 4 cells, and we extract a 59-dimensional uniform local binary pattern [34] 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 d = 4,720. We use principal component analysis (PCA) to reduce the dimension to 500 for each landmark for further processing. More details of our implementation of high-dimensional LBP features can be found in [35].

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 IV), (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.

B. Reference Set Representations

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

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

where $C_i^{(j,k)} \in R^d$ is the reference representation of the person i in year j at landmark k, d is the feature dimension,

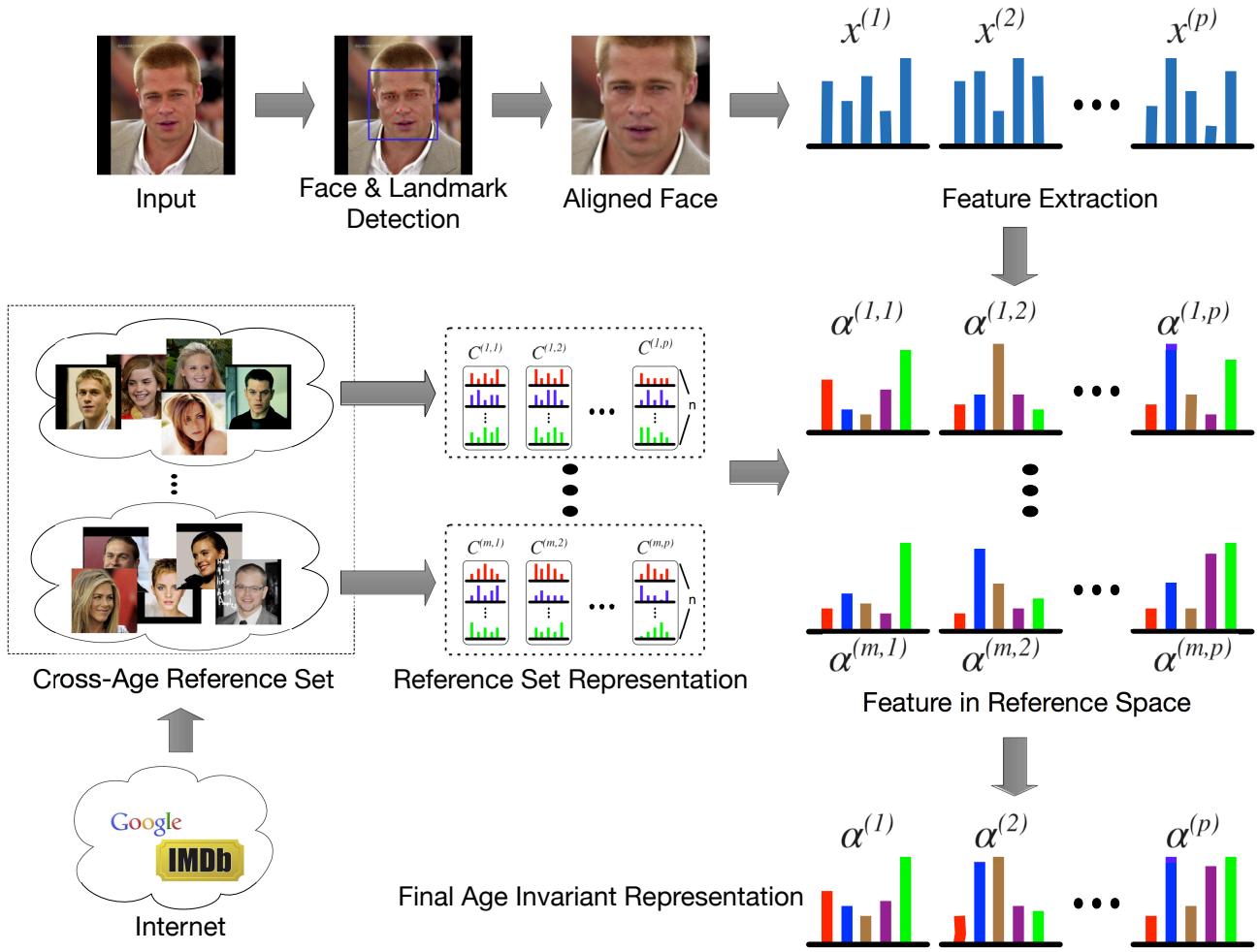

Fig. 3. System overview of our method. For each image, our system first applies face and facial landmarks detection. We then extract local features (high-dimensional LBP) from each landmark, and use CARC to encode the local features into the age-invariant representation with three steps. First, by using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

and n, m, p are the numbers of reference people, range of years, and number of landmarks, respectively. It is computed by averaging over all the features ($x^{(k)}$) from the same reference person in the same year, 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 an entry of representation more robust to such noisy data.

C. Encoding Feature into the Reference Space

Let $C^{j,k}$ be a $d \times n$ matrix consisting of n reference person representations, $C^{(j,k)} = [C_1^{(j,k)}, C_2^{(j,k)}, \dots C_n^{(j,k)}]$. Given a new feature $x^{(k)}$ extracted at landmark k, we want to use the reference representation to encode the new feature. For this purpose, 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 2) in year j for feature extracted at landmark k. It is concerned that $\alpha_i^{(j,k)}$ should be large if the testing feature $x^{(k)}$ is close to the i_{th} reference person, and small otherwise. Here we consider finding such a representation by solving a least squared problem with Tikhonov regularization:

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

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 consistency term to reflect this temporal constraint in our coding scheme. We first define a tridiagonal matrix L as follow:

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

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

their difference. Let

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

The testing features $x^{(k)}$ can now be converted to the new reference space by minimizing the following objective function with additional temporal smoothness constraint:

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

The first term in the above equation is to ensure the reconstruction error in reference space is small, and the second term is to make the coefficients of the same reference person across adjacent years 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:

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

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

and we define the vector $\hat{\alpha}^{(k)} = [\alpha^{(1,k)T}, \alpha^{(2,k)T}, \ldots, \alpha^{(m,k)T}]^T \in R^{mn}$ and $\hat{x}^{(k)} = [x^{(k)T}, \ldots, x^{(k)T}] \in R^{md}$. 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 has an analytic solution:

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

Denote $\hat{P}^{(k)} = (\hat{C}^{(k)T} \hat{C}^{(k)} + \hat{\lambda}_1 I + \hat{\lambda}_2 \hat{L}^T \hat{L})^{-1} \hat{C}^{(k)T}$. Note that $\hat{P}^{(k)}$ can be obtained offline as a projection matrix. Hence, when a query image comes, our system can efficiently cast it to the reference set space via the precomputed projection matrix $\hat{P}^{(k)}$.

D. 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:

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

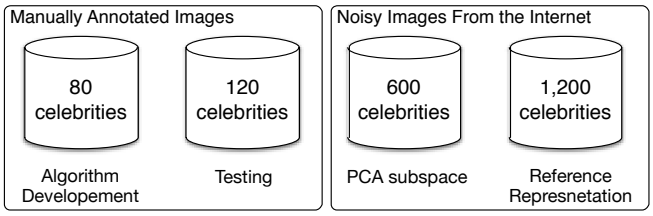

Fig. 4. CACD contains images of 2,000 celebrities. 200 of them are manually annotated by viewing the image contents. These 200 celebrities are divided into two subset: 80 are used for algorithm development and parameter selection, and the other 120 are used for testing. 600 out of 1,800 celebrities without annotation are employed for the PCA subspace computation. The other 1,200 are used as reference set.

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: 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 in a later year. Hence the final representations for these two images will both have high response at that specific reference person because of the max pooling aggregation. Our system can therefore achieve age-invariant face recognition and retrieval.

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

IV. Cross-Age Celebrity Dataset (CACD)

A. Celebrity Name Collection

In order to create a dataset with large gaps of ages, two important criteria are adopted to decide whose images should be included in the dataset: (1) the people in the dataset should have different ages, and (2) these people must have many images available on the Internet to gather. A list of celebrity names are first collected for the dataset construction. We select our names from an online celebrity database, IMDb.com[2], 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 dates are 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. A similar approach is adopted in [36] to collect celebrity names.

B. Image Collection

We use Google Image Search to collect images. We specify to collect "face image" (an option in Google image search). Therefore, most images only contain single face. In the case of multiple face detected, we simply extract the largest face in the image. Note that there might be still some false detections. However, we manually remove these images for the evaluation part of the dataset. In order to collect celebrities images across

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

different ages, we use a combination of celebrity name and year as keywords. For example, "Emma Watson 2004" is employed 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." Then, we collect images across ten years from 2004 to 2013 for each celebrity. Since we already know the birth years of the celebrities, the ages of celebrities in the images can be calculated by simply subtracting the birth year with the year of which the photo was taken. Figure 1 shows some sample images collected. Note that the dataset might contain noise because we could accidentally collect images of other celebrities in the same event or movie. Nevertheless, the proposed coding method is robust to such noise and proved to achieve good performance in our experiments that will be demonstrated in section V.

C. Dataset Statistics

We perform face detection [32] to all images and find more than 200,000 images containing faces for all 2,000 celebrities. We use a simple duplicate detection algorithm based on low-level features to remove the duplicated images. After that, we have around 160,000 face images left. For a subset of 200 celebrities, we manually check the images and remove the noisy images in the dataset by viewing the image content. 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 are preserved for algorithm development and parameter selection; the other 120 celebrities are for testing and performance evaluation. Figure 4 shows the protocol for using the dataset in our experiments. 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 our knowledge. The statistics of the dataset and comparison to other existing face datasets are shown in Table I, and our dataset has the largest amount of images with age variation. Compared to MORPH dataset, age gaps for CACD between images of the same person are larger. FG-NET has larger age gaps but there are only few images from a limited number of people contained in this dataset. Table II shows the distribution of the datasets with different ages. Both MORPH and CACD do not contain images with age of 10 or younger, while FG-NET has more images of younger ages. However, CACD has more images for all other ages. To further understand the dataset, we also run race and gender detection using a commercial system [37] to analyze the gender and race distributions of the datasets. The results are shown in Figure 5. We can see that the numbers of male and female in CACD are roughly equal but MORPH dataset consists of mostly male subject. CACD consists of mostly White people because the name list is obtained from IMDB while MORPH dataset consists of mostly Black people. Note that CACD currently only contains images from 2004-2013. More images, including images taken after 2013 and images before 2004, can be easily added to extend the dataset, as the celebrity names are publicly available.

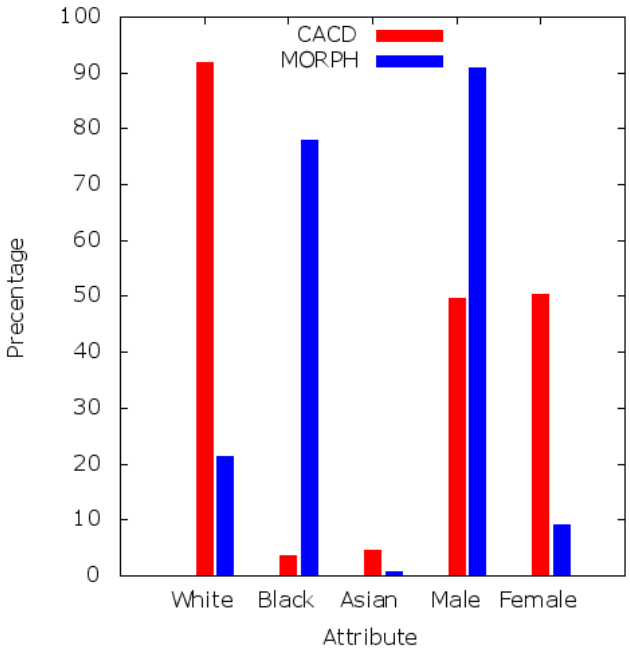

Fig. 5. The gender and race distributions of CACD and MORPH dataset. MORPH dataset contains mostly male subjects while CACD have roughly equal numbers of male and female subjects. MORPH dataset contains mostly Black people while CACD contains mostly White people because the name is collected from IMDB.

D. Verification Subset (CACD-VS)

To understand how human performs on the task of cross-age face recognition. We follow the protocol similar to [5] and construct a verification subset (CACD-VS) from CACD for face verification.

Since CACD is constructed from Internet images, it contains some noisy images. Although we manually remove these noisy images by checking the image content for 200 celebrities, it might still contain a small amount of noisy images because face images across ages are sometimes hard to recognize even for humans. In order to avoid these noisy images, CACD-VS is more carefully annotated by checking both image and surrounding web contents to ensure that the images have correct identity tags. By annotating 4,000 images from 2,000 celebrities with each celebrity having two images, CACD-VS is constructed with 4,000 image pairs, including 2,000 positive pairs and 2,000 negative pairs, for the verification task in the following section.

V. Experiments

A. Experiments on Cross-Age Celebrity Dataset

We separate the CACD 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 as a validation set for parameter 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.

TABLE I
THE COMPARISON BETWEEN EXISTING DATASETS. THE AGE GAP HERE MEANS THE GAP BETWEEN IMAGES OF THE SAME IDENTITY IN THE DATASET. OUR DATASET HAS THE LARGEST AMOUNT OF IMAGES AND CONTAINS AGE INFORMATION. COMPARED TO MORPH DATASET, AGE GAPS BETWEEN IMAGES OF THE SAME PERSON IN CACD 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 [5]	13,233	5,749	2.3	No	-
Pubfig [6]	58,797	200	293.9	No	-
FGNet [31]	1,002	82	12.2	Yes	0-45
MORPH [8]	55,134	13,618	4.1	Yes	0-5
Ours (CACD)	163,446	2,000	81.7	Yes	0-10

TABLE II
THE DISTRIBUTION OF THE DATASETS WITH DIFFERENT AGES.

Dataset	(0-10)	(10-20)	(20-30)	(30-40)	(40-50)	(50-60)	60+
FGNET [31]	411	319	143	69	39	14	7
MORPH [8]	0	7,469	16,325	15,357	12,050	3,593	340
CACD	0	7,057	39,069	43,104	40,344	30,960	2,912

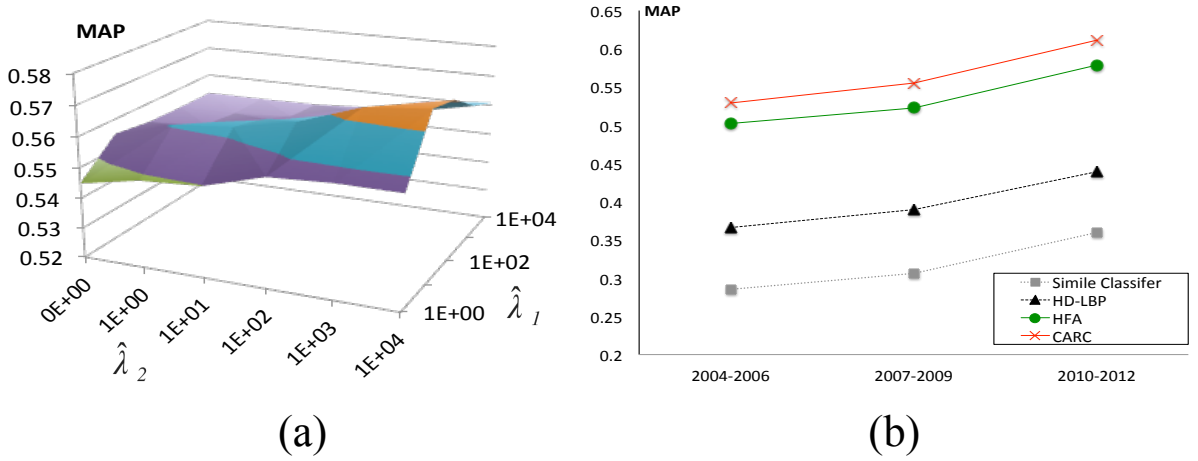

Fig. 6. (a) Validation results of Cross-Age Reference Coding on our CACD using different parameters. The results reveal that using temporal constraint can improve the performance. The parameters $(\hat{\lambda}_1, \hat{\lambda}_2) = (10^1, 10^4)$ are chosen. (b) The retrieval results on CACD in comparison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

1) Evaluation Metrics: Mean average precision (MAP) has been widely used as an evaluation metric in information retrieval and image retrieval tasks [38], and it is calculated as follows: For the retrieval results of each query image, precision at every recall level is computed and averaged to get the average precision (AP). MAP is then computed by averaging the APs of all query images. More specifically, let $q_j \in Q$ be the query images, and the positive images (images of the same person as query) in the database for q_j are $\{I_1, I_2, \ldots, I_{m_j}\}$. We rank the retrieval results of q_j in a descending order, and let R_{jk} be the ranked retrieval results from the top to the image I_k. The MAP is calculated as:

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

where $Precision(X)$ is the precision (the ratio of positive images in X). We adopt MAP as the evaluation metric in the following experiments.

2) Parameters Selection: To select 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. In our method, there are several parameters we need to decide, including the PCA feature dimensions d, the regularization parameters in our coding framework $\hat{\lambda}_1$, $\hat{\lambda}_2$, and the number of reference celebrities, n.

- PCA feature dimensions d: we run experiments from 100 to 1,000 and find that the performance stops to improve after 500, and thus we fix d as 500 in the further experiments[3].

[3]Note that 500 dimension in PCA subspace retains 68.7% total variance while 1,000 dimension retains 81.0% in our experiments.

- Regularization parameters: 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). As can be seen, adding temporal constraint by increasing $\hat{\lambda}_2$ in our coding framework is helpful to the performance enhancement. We set $(\hat{\lambda}_1, \hat{\lambda}_2) = (10^1, 10^4)$ where they achieve the best performance in the validation set.
- Size of reference set n: we then randomly select reference celebrities from 40 to 1,200 and find that the performance stops to improve after 600 for the validation set. Hence, we fix the number of reference celebrities to 600 for testing.

3) Compared Algorithms: We compare CARC to several state-of-the-art methods, including:

- High-dimensional local binary pattern [2] (HD-LBP): it achieves top performance for the LFW dataset, the most popular environment-unconstrained dataset. HD-LBP is also the local feature adopted for CARC, and we use PCA to reduce the dimension to 500 for the features obtained from each landmark.
- Simile Classifier [6]: we train a linear-SVM for each reference celebrity and use the sign distance to the decision boundary as the feature. We use LIBLINEAR package [39] to carry out the training and the number of reference celebrities is also set as 600.
- Hidden Factor Analysis [28] (HFA): a state-of-the-art method for age-invariant face recognition. We use HD-LBP as input feature and the parameters are tuned to the best setting according to that paper.

4) Results and Discussions: The images of 120 celebrities are used for testing. We conduct experiments with three different groups. In all three groups, images taken in 2013 are used as query images. The database contains images taken in 2004-2006, 2007-2009, and 2010-2012 for each of the three groups, respectively. For all methods, cosine similarity is employed as the similarity measure. The performance is shown in Figure 6 (b). As can be seen, the proposed method outperforms other methods in all three groups. 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 difference is larger, which reveals the difficulty of face retrieval with age variation. Nevertheless, both HFA and the proposed method, CARC, can achieve higher performance on the group of larger age difference compared to baseline features on the group with smaller age difference. It demonstrates the effectiveness of the age-invariant methods. CARC achieves higher performance than HFA, which reveals that CARC can better utilize the noisy reference set and is more robust to age variation. Figure 7 shows some top-10 retrieval results using the proposed method.

B. Experiments on MORPH Dataset

We also use CARC for face recognition experiments on MORPH dataset to show its efficacy. For this dataset, we follow the experimental setting in [27] for close set face identification. 10,000 subjects are randomly selected from the MORPH dataset and the youngest images of these subjects are used to construct the gallery set, and the oldest images of the subjects are used as the probe set. Both gallery and probe sets consist of 10,000 images from 10,000 different subjects. We then randomly select another 600 subjects from the MORPH dataset as the reference set for our algorithm. Images of subjects outside these 10,600 subjects are used for building PCA and LDA subspaces as having been done in [28]. We follow [28] to reduce the dimension of PCA and LDA to 1,001 and 1,000 for features from each landmark. However, as we found out in the experiments, the PCA dimension has little impact on the performance after it reaches 500.

TABLE III
RANK-1 IDENTIFICATION RESULTS ON MORPH DATASET. CARC ACHIEVES THE HIGHEST RATE COMPARED TO OTHER STATE-OF-THE-ART METHODS.

Method	Recognition Rate
Generative model (Park et al., 2010) [22]	79.8%
Discriminative model (Li et al., 2011) [27]	83.9%
HFA (Gong et al., 2013) [28]	91.1%
CARC (Ours)	**92.8%**
CARC (Cross-dataset)	83.4%

Our algorithm is compared to several state-of-the-art methods including, (1) a generative aging model [22], (2) a discriminative model for age-invariant face recognition [27], and (3) HFA, currently the best result on the dataset [28]. The results in terms of rank-1 recognition rate of our algorithm compared to other methods can be found in Table III, which shows that the proposed method can achieve better performance compared to other state-of-the-art methods. Figure 8 shows some incorrect matching examples. Although our system can achieve higher than 92% accuracy, it still fails in some cases, particularly when the probe and gallery are significantly different. Some of these cases are really hard even for humans to recognize. For some applications, we do not need to have perfect rank-1 accuracy, but only need to find the correct match in the top-k results. For instance, in crime investigation, the law enforcement agency only needs to go through top-20 list to find out the suspect with the help of the face recognition system. Correct matches can be found in top-20 result in 98% of the probe images using our approach, and we can achieve 94.5% MAP in the MORPH dataset.

C. Cross-dataset Experiments

In order to find out the generalizability of CARC and the CACD dataset, we also run experiments under a cross-dataset setting. We use the same protocol as the previous section to run experiments on MORPH dataset. However, instead of using a reference set drawn from the training set of MORPH, we directly use celebrity images from CACD as the reference set. The result are shown in the last row of Table III. The performance drops from 92.8% to 83.4% when using celebrity images as the reference set. This is probably because the distributions of two datasets are quite different: CACD consists mostly of White people while MORPH consists mostly of

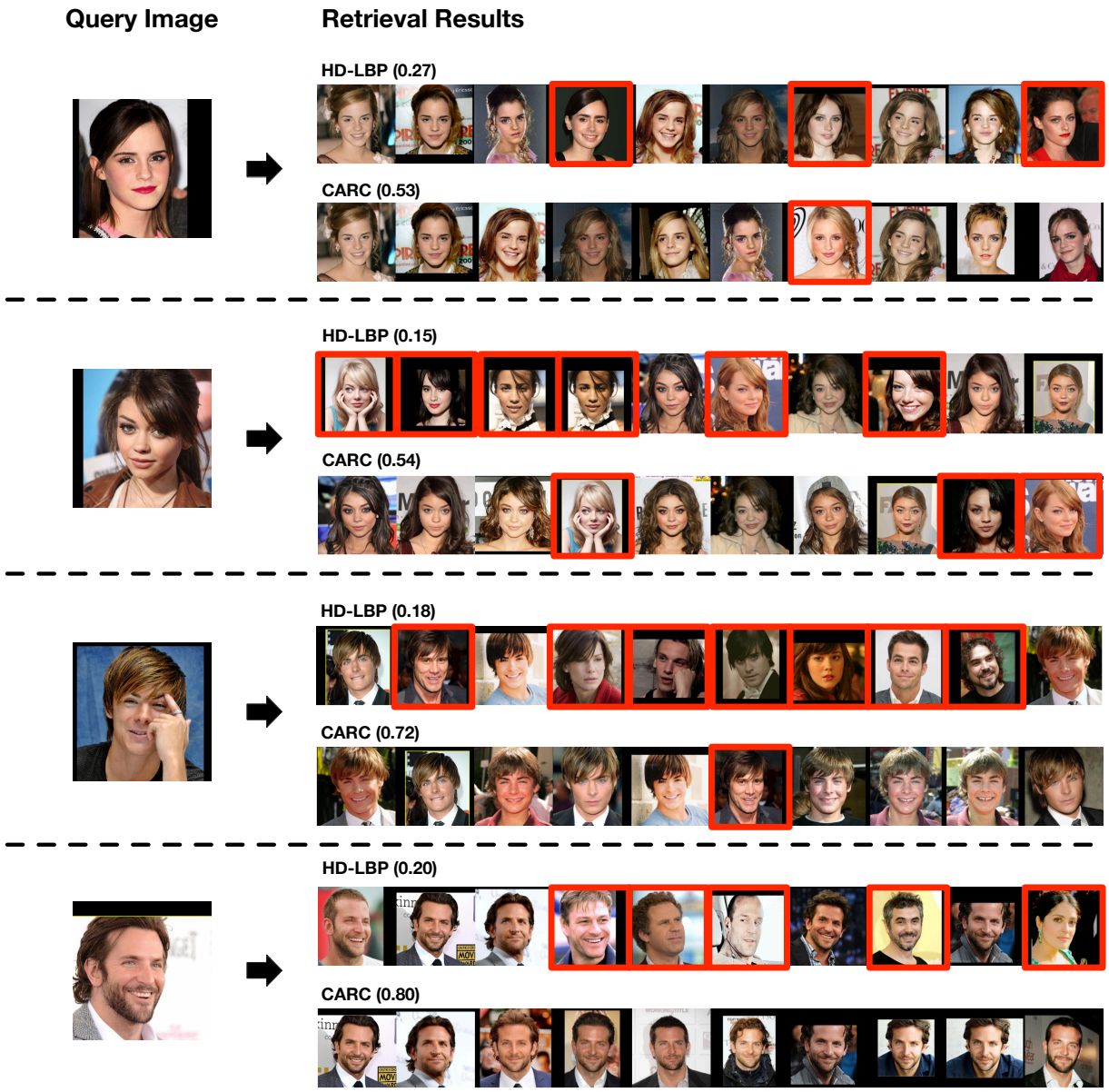

Fig. 7. Some examples of the retrieval results in CACD using HD-LBP and CARC. Left column contains the query images, and right column contains top ten retrieval results (from left to right) using the image in the left column as the query image. Red boxes indicate false positive, and the numbers in the parentheses are the AP for the result. CARC can retrieve images of the same person across ages and perform better than using the original feature (HD-LBP) space.

Black people (cf. Figure 5). Hence, it is hard to represent one dataset by the other.

D. Experiments on CACD-VS

1) CARC Performance: In order to conduct experiments on CACD-VS, we first divide the dataset into ten folds, and each fold contains 400 image pairs (200 positive pairs and 200 negative pairs) from 200 celebrities. Celebrities from each fold are mutually exclusive. We repeat the experiments ten times for each of the ten folds and report the average results. For each run, we use one of the ten fold for testing, and the other nine folds for computing PCA subspace. We adopt the parameters found in Section V-A for the experiments. For HFA and CARC, we use celebrity images from three folds

TABLE IV
VERIFICATION ACCURACY OF DIFFERENT METHODS ON CACD-VS. THE PROPOSED METHODS OUTPERFORM OTHER METHODS AND ACHIEVE THE ACCURACY THAT IS HIGHER THAN THE PERFORMANCE OF AVERAGE HUMAN. BY AGGREGATING RESULTS FROM MULTIPLE USERS, HUMAN CAN ACHIEVE THE ACCURACY UP TO 94.2%.

Method	Verification Accuracy
HD-LBP [2]	81.6%
HFA [28]	84.4%
CARC-NT	85.6%
CARC	**87.6%**
Human, Average	85.7%
Human, Voting	**94.2%**

Fig. 8. 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.

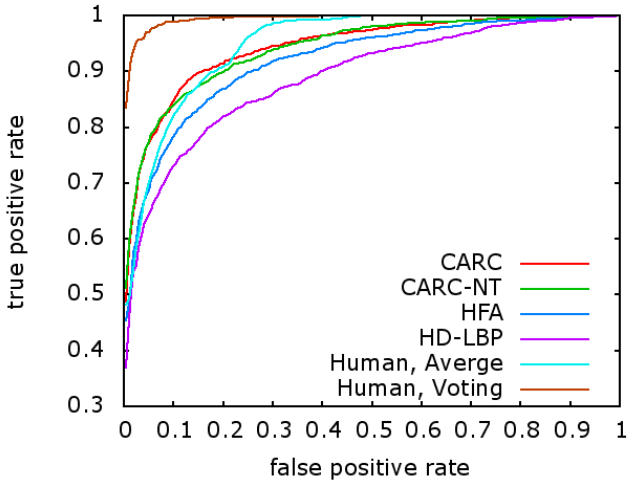

Fig. 9. ROC curves for different methods on CACD-VS. CARC performs better than HD-LBP and HFA, and it is competitive to average human. The bump on the purple curve suggests that human is better at rejecting negative pairs.

(600 celebrities) to compute the HFA model and reference representations, respectively. After finding the representation for each image in each method, we compute the cosine similarity between pairs and use a linear SVM to decide the classification threshold. Note that the linear SVM is only used to find the threshold on one-dimensional cosine similarities, so the parameter C does not have great influence on the accuracy in the experiments. Here we simply use the default parameter $C = 1$ for linear SVM. To show the effectiveness of the proposed method, we also run CARC without the temporal constraint. The result is denoted as CARC-NT. The verification accuracy of different methods including HD-LBP, HFA, CARC-NT, and CARC are shown in Table IV and the ROC curves are shown in Figure 9. Our method out-performs other state-of-the-art methods and achieves the highest accuracy of 87.6%.

2) Human Performance: We would like to know how human performs on the task of cross-age face recognition and how the acquaintance of the subjects affect the recognition performance. To this end, we follow the same procedure in [6] and collect data from Amazon Mechanical Turk. We ask ten users to answer three questions for each image pair, including whether two people in the images are the same person, how confident is he/she with his/her answer, and whether he/she has seen the person before. In order to make sure the users answer the questions legitimately, we require the users to have at least 95% approved rate. We asked each user to answer questions for 50 image pairs, and results show that 99% of the users achieve at least 60% accuracy on the questions, which is far beyond random guess and suggest that the users answer the questions in good faith. An example of the questions shown to the users can be found in Figure 10. We gather 40,000 data points from users and 34270 of them correctly answer the first question. Hence, the average human performs on the task of cross-age face verification is around 85.7%. Note that human performance on face verification in LFW shown in [6] is around 99.2%. This suggests that human performs worse on face verification when faces are across ages, and CACD-VS indeed is a more challenging dataset since it contains age variations. The proposed method, CARC, achieves an accuracy of 87.6%, which performs slightly better than the average human performance.

Although the proposed method performs better than average humans, we find that combining the decisions from multiple users can achieve better performance. Because this is like ensemble the results from multiple classifiers. Table V shows the majority voting results from the number of 1,3,5,7,9 users respectively. When combined results from nine users, human can achieve an accuracy of 94.2%. It suggests that there is still a gap between machine and human to improve on the task of cross-age face recognition. Table IV and Figure 9 show the accuracy and ROC curves of human performances. ROC curves on human performance are based on the confident scores they answer. Note that there is a bump on the ROC

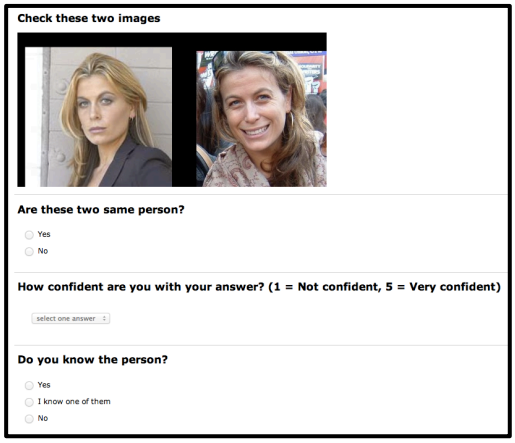

Fig. 10. Example questions shown to the users on Amazon Mechanical Turk. We collect 10 results from different users for each of the 4,000 image pairs in CACD-VS.

TABLE V
VERIFICATION ACCURACY ON CACD-VS WHEN AGGREGATING RESULTS FROM MULTIPLE USERS. COMBINING RESULTS FROM MULTIPLE USERS ACHIEVE HIGHER ACCURACY SINCE IT IS LIKE ENSEMBLE WITH MULTIPLE CLASSIFIERS.

Number of Users	1	3	5	7	9
Accuracy	82.3%	89.8%	92.0%	93.1%	**94.2%**

curve of human performance. We hypothesize that the bumps suggest that humans perform better on rejecting negative pairs since the bump means most of the negative pairs are correctly classified with high confidence.

In order to support our hypothesis, We further analyze the percentage of false positive and false negative pairs in the verification results from both human and CARC. We find that the false positive rate of human is much smaller compared to CARC. The results of false positive rate and false negative rate are shown in Table VI. We find that while CARC achieve similar false negative rate compared to Human (Voting), false positive rate of CARC is much higher than the false positive rate of Human (Voting). This suggests that human and CARC perform similar at accepting positive pairs, but human performs much better than CARC at rejecting negative pairs. Therefore, we can focus on reducing the false positive rate in future researches.

Next, we want to analyze how the acquaintance to the subjects affect the recognition performance for human. For all 40,000 data points collected from users, 11,936 answer they know either one or both the subjects in the images. From these answers, 92.9% of them correctly classified the image pairs. On the other hand, 28,064 answer they do not know the subject in the images and only 82.6% of them correctly classified the image pairs. The results are intuitive: acquaintance with the subjects indeed affect the recognition performance, since knowing the subjects is like having extra training data on the subjects.

Finally, we show some examples misclassified by either human, CARC, or both. Figure 11 (a) shows false positive

TABLE VI
PERCENTAGES OF TRUE POSITIVE, TRUE NEGATIVE, FALSE POSITIVE AND FALSE NEGATIVE IN CARC AND HUMAN VERIFICATION RESULTS. HUMAN IS BETTER AT REJECTING NEGATIVE PAIRS, THEREFORE, IT HAS A LOWER FALSE POSITIVE RATE.

Method	True Positive	True Negative	False Positive	False Negative
CARC	44.05%	43.58%	6.43%	5.95%
Human, Average	39.11%	46.57%	3.43%	10.89%
Human, Voting	44.98%	49.20%	0.80%	5.03%

results and Figure 11 (b) shows false negative results. Red boxes in Figure 11 mean the images are of different people while blue boxes mean the images are of the same person. First columns in Figure 11 (a) and (b) show false positive and false negative examples that are misclassified by human but correctly classified by CARC, respectively; second columns of both figures show examples that are misclassified by CARC but correctly classified by human; and third columns show examples misclassified by both human and CARC.

VI. CONCLUSIONS

By utilizing a cross-age reference set obtained from the Internet, we propose a new coding method, CARC, which can help map low-level feature into an age-invariant reference space. The experimental results show that CARC outperform state-of-the-art methods and achieve high accuracy in face recognition and retrieval across age. We also introduce a large-scale face dataset, CACD, 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. Although our experiments show CARC can achieve superior performance in both CACD and MORPH datasets, the performance in cross-dataset setting drops considerably. The drop is probably caused by the huge difference between the appearance distributions of the two datasets. In the future work, we would like to address this problem by introducing domain adoption techniques. In order to evaluating human performance on the task of cross-age face recognition, we further constructed a carefully annotated verification subset called CACD-VS and conduct extensive experiments. Our experiments show that although the proposed methods performs better than average human, combing results from multiple human can achieve higher performance. Therefore, there are still a gap on the task. We also show that human performs better mostly on rejecting negative pairs, and acquaintance in the subject is helpful to human for recognition. In the future, we want to investigate how to effectively choose a subset from the reference people for further improving the performance of age-invariant face recognition and retrieval, and also how to reduce the false positive rate in the recognition process in order to achieve similar performance of human.

ACKNOWLEDGMENT

This work was supported in part by the Ministry of Science and Technology of Taiwan under Contract MOST 103-2221-E-002-105-MY3, Grant MOST 103-2221-E-001-010, and the

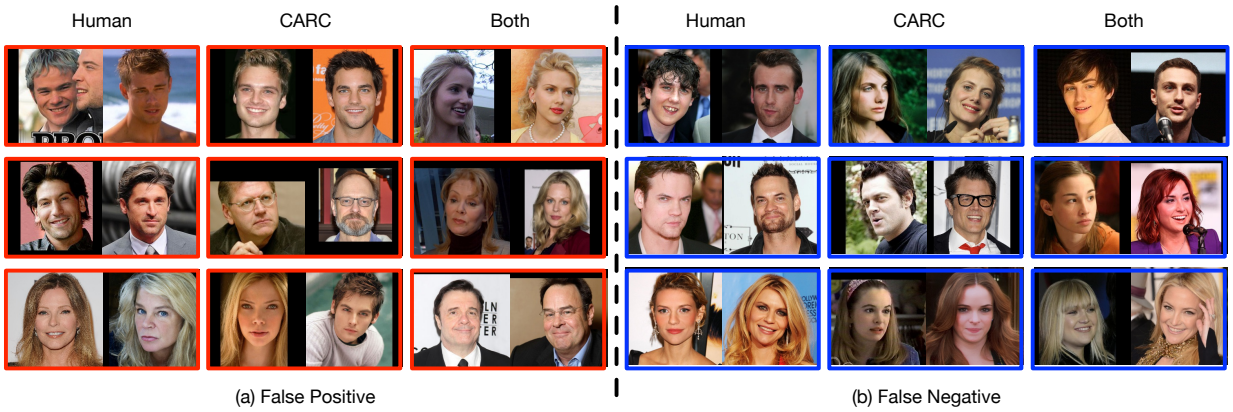

Fig. 11. (a) False positive examples and (b) False negative examples which are misclassified by human (first column), CARC (second column), or both (third column). Red boxes indicate different people in the image pairs and blue boxes indicate same person in the image pairs.

Excellent Research Projects of National Taiwan University under Grant 102R7762.

REFERENCES

[1] A. K. Jain, B. Klare, and U. Park, "Face matching and retrieval in forensics applications," *IEEE Multimedia*, vol. 19, no. 1, p. 20, 2012.
[2] D. Chen, X. Cao, F. Wen, and J. Sun, "Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification," in *IEEE Conf. Computer Vision and Pattern Recognition*, 2013, pp. 3025–3032.
[3] K. Simonyan, O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Fisher vector faces in the wild," in *British Machine Vision Conf.*, vol. 1, 2013, p. 7.
[4] O. Barkan, J. Weill, L. Wolf, and H. Aronowitz, "Fast high dimensional vector multiplication face recognition," in *IEEE Int. Conf. Computer Vision*, 2013.
[5] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," University of Massachusetts, Amherst, Tech. Rep. 07-49, October 2007.
[6] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar, "Attribute and simile classifiers for face verification," in *IEEE Int. Conf. Computer Vision*, 2009, pp. 365–372.
[7] Q. Yin, X. Tang, and J. Sun, "An associate-predict model for face recognition," in *IEEE Conf. Computer Vision and Pattern Recognition*, 2011, pp. 497–504.
[8] K. Ricanek and T. Tesafaye, "Morph: A longitudinal image database of normal adult age-progression," in *IEEE Int. Conf. Automatic Face and Gesture Recognition*, 2006, pp. 341–345.
[9] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Computing Surveys*, vol. 35, no. 4, pp. 399–458, 2003.
[10] S. Z. Li and A. K. Jain, *Handbook of Face Recognition*, 2nd ed. Springer Publishing Company, Incorporated, 2011.
[11] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *IEEE Conf. Computer Vision and Pattern Recognition*, 1991, pp. 586–591.
[12] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2037–2041, 2006.
[13] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, 2009.
[14] T. Berg and P. N. Belhumeur, "Tom-vs-pete classifiers and identity-preserving alignment for face verification." in *British Machine Vision Conf.*, vol. 1, 2012, p. 5.
[15] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-pie," *Image and Vision Computing*, vol. 28, no. 5, pp. 807–813, 2010.
[16] Z. Wu, Q. Ke, J. Sun, and H.-Y. Shum, "Scalable face image retrieval with identity-based quantization and multireference reranking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 10, pp. 1991–2001, 2011.
[17] G. Mu, G. Guo, Y. Fu, and T. S. Huang, "Human age estimation using bio-inspired features," in *IEEE Conf. Computer Vision and Pattern Recognition*, 2009, pp. 112–119.
[18] A. Montillo and H. Ling, "Age regression from faces using random forests," in *IEEE Int. Conf. Image Processing*, 2009, pp. 2465–2468.
[19] K.-Y. Chang, C.-S. Chen, and Y.-P. Hung, "Ordinal hyperplanes ranker with cost sensitivities for age estimation," in *IEEE Conf. Computer Vision and Pattern Recognition*, 2011, pp. 585–592.
[20] J. Suo, X. Chen, S. Shan, and W. Gao, "Learning long term face aging patterns from partially dense aging databases," in *IEEE Int. Conf. Computer Vision*, 2009, pp. 622–629.
[21] J. Suo, S.-C. Zhu, S. Shan, and X. Chen, "A compositional and dynamic model for face aging," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 3, pp. 385–401, 2010.
[22] U. Park, Y. Tong, and A. K. Jain, "Age-invariant face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 5, pp. 947–954, 2010.
[23] A. Lanitis, C. J. Taylor, and T. F. Cootes, "Toward automatic simulation of aging effects on face images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 442–455, 2002.
[24] X. Geng, Z.-H. Zhou, and K. Smith-Miles, "Automatic age estimation based on facial aging patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 12, pp. 2234–2240, 2007.
[25] T. Wu and R. Chellappa, "Age invariant face verification with relative craniofacial growth model," in *European Conf. on Computer Vision*, 2012, pp. 58–71.
[26] H. Ling, S. Soatto, N. Ramanathan, and D. W. Jacobs, "Face verification across age progression using discriminative methods," *IEEE Trans. Inf. Forens. Security*, vol. 5, no. 1, pp. 82–91, 2010.
[27] Z. Li, U. Park, and A. K. Jain, "A discriminative model for age invariant face recognition," *IEEE Trans. Inf. Forens. Security*, vol. 6, no. 3, pp. 1028–1037, 2011.
[28] D. Gong, Z. Li, D. Lin, J. Liu, and X. Tang, "Hidden factor analysis for age invariant face recognition," in *IEEE Int. Conf. Computer Vision*, 2013.
[29] B.-C. Chen, C.-S. Chen, and W. H. Hsu, "Cross-age reference coding for age-invariant face recognition and retrieval," in *European Conf. on Computer Vision*, 2014.
[30] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The feret evaluation methodology for face-recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1090–1104, 2000.
[31] Face and Gesture Recognition Working group, "Fg-net aging database," 2000.
[32] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.
[33] X. Xiong and F. De la Torre, "Supervised descent method and its applications to face alignment," in *IEEE Conf. Computer Vision and Pattern Recognition*, 2013, pp. 532–539.
[34] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, 2002.
[35] B.-C. Chen, C.-S. Chen, and W. H. Hsu, "Review and implementation of high-dimensional local binary patterns and its application to face

recognition," Institute of Information Science, Academia Sinica, Tech. Rep. TR-IIS-14-003, 2014.
[36] D. Wang, S. C. Hoi, Y. He, and J. Zhu, "Retrieval-based face annotation by weak label regularized local coordinate coding," in *ACM Int. Conf. Multimedia*, 2011, pp. 353–362.
[37] H. Fan, Z. Cao, Y. Jiang, Q. Yin, and C. Doudou, "Learning deep face representation," *arXiv preprint arXiv:1403.2802*, 2014.
[38] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, "Object retrieval with large vocabularies and fast spatial matching," in *IEEE Conf. Computer Vision and Pattern Recognition*, 2007, pp. 1–8.
[39] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, "Liblinear: A library for large linear classification," *The Journal of Machine Learning Research*, vol. 9, pp. 1871–1874, 2008.

Bor-Chun Chen Bor-Chun Chen is a Ph.D. student at University of Maryland. He received the M.S. degree from the Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan, in 2012. His research interests include computer vision, multimedia analysis, large scale image retrieval, face image retrieval.

Chu-Song Chen Chu-Song Chen is a research fellow of Institute of Information Science (IIS), Academia Sinica, Taiwan, and also an adjunct professor of the Graduate Institute of Networking and Multimedia (GINM), National Taiwan University. In 2008-2014, he has been the deputy director of Research Center for Information Technology Innovation (CITI), Academia Sinica. In 2007-2008, he served as the Secretary-General of the IPPR Society, Taiwan, which is one of the regional societies of the International Association of Pattern Recognition (IAPR), and he is a governing board member of IPPR currently. Dr. Chen?s research interests include pattern recognition, computer vision, and signal/image processing. He served as an Area Chair of ACCV?10 and NBiS?10, the Program Chair of IMV?12 and IMV?13, the Tutorial Chair of ACCV?14, and the General Chair of IMEV?14, and will be the Workshop Chair of ACCV?16. He is on the editorial board of Journal of Multimedia (Academy Publisher), Machine Vision and Applications (Springer), and Journal of Information Science and Engineering (IIS).

Winston H. Hsu Winston H. Hsu (M07SM12) received the Ph.D. degree in electrical engineering from Columbia University, New York, NY, USA. He has been an Associate Professor in the Graduate Institute of Networking and Multimedia, National Taiwan University, Taipei, Taiwan, since February 2007. Prior to this, he was in the multimedia software industry for years. He is also with the Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan. His research interests include multimedia content analysis, image/video indexing and retrieval, machine learning, and mining over large-scale databases. Dr. Hsu serves in the Editorial Board for the IEEE Multimedia Magazine.