A Reference-Based Framework for Pose Invariant Face Recognition

Mehran Kafai[1] and Kave Eshghi[2] and Le An[3] and Bir Bhanu[4]

[1] HP Labs, Palo Alto, CA 94304, USA
[2] Google Inc., Mountain View, CA 94043, USA
[3] BRIC, University of North Carolina at Chapel Hill, NC 27599, USA
[4] Center for Research in Intelligent Systems, University of California, Riverside, CA 92521, USA

mehran.kafai@hp.com, kave@google.com, lan004@unc.edu, bhanu@cris.ucr.edu

Abstract— While face recognition technology has made significant progress in recent years, practical pose invariant face recognition remains a challenge. This paper describes a reference-based framework for solving this problem. The similarity between a face image and a set of reference individuals defines the reference-based descriptor for a face image. Recognition is performed using the reference-based descriptors of probe and gallery images. The dimensionality of the face descriptor generated by the accompanying face recognition algorithm is reduced to the number of individuals in the reference set. The proposed framework is a generalization of previous recognition methods that use indirect similarity and reference-based descriptors. Results are shown on a combination of seven publicly available face databases (LFW, FEI, RaFD, FERET, FacePix, CMU-PIE, and Multi-PIE). The proposed approach achieves good accuracy as compared to popular state-of-the-art algorithms, and it is computationally efficient.

I. INTRODUCTION

Face recognition is an important computer vision and pattern recognition technology that is used in many different applications, from organizing photo albums to surveillance. In its most common form named *identification*, a database of images of individuals, one or more images per individual, called the gallery is available. There is also a set of images, called probes, taken of individuals some of whom may be the individuals in the gallery. The task is to identify which probe and gallery images correspond to the same individual. If a probe individual is not in the gallery (open-set identification) then the system should be able to infer it.

When the probe and gallery images are taken under the same pose (for example frontal or side-view), the task essentially boils down to a pattern recognition problem: the gallery and probe images are first processed to generate face descriptors (e.g., Local Binary Patterns (LBP) [1]). The probe descriptors are compared with the gallery descriptors using some distance measure, such as the Euclidean distance. The gallery images which are closest in the distance measure to the probe image are deemed to be of the same individual.

Face recognition across pose is the problem of recognizing a face from a new viewpoint which has not previously been seen (e.g., the probe image is a profile/side-view image, whereas the gallery image is frontal). The problem is that, using descriptors such as LBP and an associated distance

measure, two images of the same person in different poses are typically more distant from each other than the images of two different individuals in the same pose. As a result, poor recognition accuracy may be observed. Most of the solutions to this problem assume that for each individual, multiple images in different poses are available (either in the gallery or among the probes). This assumption is not always valid, e.g., in surveillance [2], where only one image may be available in the gallery, and one image, in a possibly different pose, as the probe. There are algorithms that can match a single gallery image with a given probe image in a different pose after generating a 3-D model for every gallery image (e.g., [3]), but these algorithms are computationally expensive, particularly when the size of gallery is large.

In this paper, we describe a new framework for face recognition across pose. The proposed approach is a computationally inexpensive solution to the two major challenges in face recognition: face recognition across pose, and face recognition with a single image per person gallery. This approach relies on generating a *reference-based (RB) descriptor* for each gallery and probe face image based on their similarity to a reference set. The reference set is utilized as a basis to generate RB descriptors for the probe and gallery images. Once the RB descriptors are computed, the similarity between the probe and gallery images is rated on the RB descriptors. The cosine measure is utilized for this purpose.

To make sure that the proposed approach is not database-specific and is not tuned to a certain database, we use a combination of seven face databases to obtain the face images. The seven databases include: LFW [4], FEI [5], RaFD [6], FERET [7], FacePix [8], CMU-PIE [9], and Multi-PIE [10]. We have chosen the gallery to contain only frontal face images; however, this is not a requirement and the gallery may contain any arbitrary pose image.

In the rest of this paper, Section II describes the related work and contributions of this work. Section III introduces the reference set, presents the overall system framework, and discusses the technical approach. Section IV introduces the face databases and presents the experimental results. Concluding remarks are provided in Section V. For better understanding of the terms and symbols, we summarize the definitions of the important terms in Table I, and define the symbols in Table II.

TABLE I

IMPORTANT TERM DEFINITIONS

Term	Definition
Profile face image	The side view image of a face.
Reference set	A set of images of multiple individuals. Various poses are stored for each individual.
Reference face	An individual from the reference set. The reference set contains N reference faces.
Gallery	A set of face images of multiple individuals. Each individual is represented with a **single** image.
Probe	A query face image which matches with **one** individual from the gallery.
First Level Descriptor (FLD)	A vector representing any gallery, probe, or reference set image. Here we use LBP as the FLD.
First Level Similarity (FLS)	Similarity between the probe or gallery images and the reference set images using an existing method.
Reference-based descriptor	A vector of FLS scores representing the similarity between a given face image and reference faces.
Second Level Similarity (SLS)	Similarity between probe and gallery reference-based descriptors.

TABLE II

DEFINITION OF SYMBOLS USED IN THIS PAPER

Symbol	Definition
A, B	Any individual from the gallery or probe set
I_j^A	Face image of individual A under pose j
N	Number of individuals in the reference set
R_u	The uth individual in the reference set. Individuals are labeled 1 to N.
$\text{FLS}(I_j^A, I_i^{R_u})$	FLS between image of individual A under pose j and image of individual u from the reference set under pose i
F^A	An N-dimensional reference-based descriptor for individual A
f_u^A	Maximum FLS between individual A and all poses of individual u from the reference set
$SLS(F^A, F^B)$	Second level similarity between F^A and F^B
γ	Lower incomplete gamma function
k	Degrees of freedom of the Chi-sq distribution
Γ	Gamma function with closed form values
P, O	LBP histograms of an image
$\chi^2(P, O)$	Chi-square distance between P and O

II. RELATED WORK AND OUR CONTRIBUTIONS

A. Related work

2-D recognition algorithms cover three main subcategories: multi-view galleries [11], using pose tolerant control points [12], and probabilistic graphical models [13]. In multi-view gallery approaches, the gallery holds multiple views of every individual with different poses. In multi-view galleries that contain real images, multiple poses are available for each individual. This differs from the work being addressed in this paper for single-view galleries.

For pose invariant face recognition, both 2-D and 3-D based approaches have been studied. 2-D faces with different poses can be synthetically generated [11]; however, this is computationally intense due to the need of generating virtual synthetic views for different poses for all individuals within the gallery. Approaches based on Markov random fields have also been developed for pose-invariant face recognition [13]; however, they are computationally expensive due to the iterative optimization stage. A warped average face (WAF) [14] is proposed for video based face recognition in which the small pose variations are rectified and the frontal view of a subject is generated by iterative averaging and template warping. Li *et al.* [15] learn a domain-specific cross-view classifier

by using a regressor with a coupled bias-variance tradeoff, and stabilize the regressor against the pose variation for face recognition across pose. Li *et al.* [16] first extract local features from densely sampled image patches and then apply a Gaussian mixture model (GMM) to capture the spatial-appearance distribution of all face images during training. Each mixture component of the GMM is confined to be a spherical Gaussian to balance the influence of the appearance and the location terms. Zhang *et al.* [17] propose a high-level feature learning scheme to extract pose-invariant identity features for face recognition.

3-D face recognition algorithms typically use a 3-D model to address across-pose recognition. The model is either obtained by a 3-D scan or generated using facial features/textures from 2-D/3-D images. Asthana *et al.* [18] propose an automatic 3-D face recognition system via 3-D pose normalization. The proposed system generates a synthetic frontal view face image for any given gallery or probe image, then performs recognition based on the frontal views. Yi *et al.* [19] first build a 3D deformable model is built and this model is used to estimate the pose of the face image. Then a group of Gabor filters are used to extract pose adaptive features. Taigman *et al.* [20] apply 3D face model to perform piecewise affine transformation. Then a deep network is trained with abundant training data, achieving human-level performance on Labeled Faces in the Wild (LFW) databases [4]. The 3-D pose normalization approach has major drawbacks; it only handles pose variation up to $\pm 45°$ in yaw and $\pm 30°$ in pitch angles. Further, it requires accurate detection of landmark points (eyes, nose, and mouth), and it is computationally expensive. Overall, due to their complexity, 3-D face recognition algorithms are not used typically for real-time face recognition across pose.

Recognition and retrieval via reference-based descriptors and indirect similarity has been well explored in the field of computer vision. Duin *et al.* [21] discuss the basic idea of a representation set in which a dissimilarity space is a data-dependent mapping using a representation set. The representation set can be chosen externally or from the available training data. Guo *et al.* [22] use exemplar-based embedding for vehicle matching. An *et al.* [23] use a reference set to bypass the direct matching in person re-identification. Liu *et al.* [24] represent human actions by a set of attributes, and perform activity recognition via visual
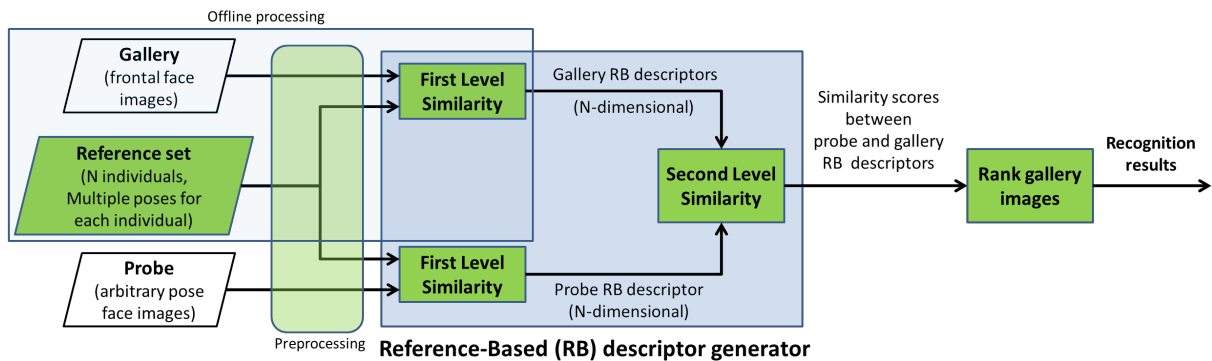
Fig. 1. The overall system diagram for reference-based face recognition across pose. After preprocessing, the reference-based descriptors for the gallery images are computed by measuring first level similarity between the gallery and reference set feature vectors. The reference-based descriptor for the probe image is generated by computing the first level similarity between the probe and reference set images. The gallery images are ranked based on the second level similarity scores between reference-based descriptors of the probe and gallery images.

characteristics symbolizing the spatial-temporal evolution of actions in a video. Schroff *et al.* [25] describe a face image by an ordered list of similar faces from a face library, i.e., a rank list representation (Doppelgänger list) is generated for each image. Kumar *et al.* [26] use attribute and simile classifiers for verification, where face-pairs are compared via their similes and attributes rather than a direct comparison. Experiments are performed on PubFig and LFW only; however, these databases do not contain images with large pose variation. Also, attribute classifiers require extensive training for recognition across pose. Likelihood-predict and associate-predict models are proposed in [27], utilizing a third data set called the identity data set as a bridge between any two images that ought to be matched.

In our framework, for each reference subject, multiple images representing different poses were enrolled and the similarity between a query subject and a reference subject is determined by the maximum similarity among different samples of this reference subject. In this way, the pose variation can be handled and it is proven to be effective for pose invariant face recognition in the experiments.

Our proposed approach can handle both closed-set and open-set identification. If used for open-set identification, the query's identity will be determined as unknown when its similarity to the gallery is below a certain threshold. The threshold can be learned from the reference set itself by evaluation via cross validation.

B. Contributions of this paper

The contributions of this paper are:

1) *Unified reference-based face recognition*: We propose a unified reference-based face recognition framework. We introduce the *reference set* as the basis set for similarity matching. The reference set contains multiple poses of different individuals. A reference-based (RB) descriptor is computed for every gallery and probe image based on their similarity to the reference set images. We call this similarity, the first level similarity (FLS).

2) *Compatible with other face recognition algorithms and face descriptors*: Our reference-based face recognition framework is compatible with any existing face recognition algorithm. In other words, it can use any current recognition algorithm to compute the first level similarity between a given gallery/probe image and the reference set images. In addition to the descriptors used in this paper, other descriptors can also be used to improve the first level similarity.

3) *Practicality*: The proposed reference-based recognition framework is simple, efficient, and can be used for real-world face recognition applications. Also, it can be extended for applications other than face recognition across pose (e.g., facial expression recognition, non-facial image recognition, etc.).

III. TECHNICAL APPROACH

Figure 1 presents the overall diagram for reference-based face recognition across pose. We start with two image sets: the gallery, and the reference set. The gallery consists of frontal face images. The goal is to rank the gallery images based on their similarity to a given probe image.

The reference set contains multiple individuals with several images under different poses for each individual. The reference set individuals are different from the individuals in the gallery and all probes; thus, no person from the gallery or probe set is present in the reference set. The reference set is built only once and it is used in conjunction with any gallery or probe set. The reference set images are collected and prepared offline before the testing (online) stage. In this paper, we use images from FEI [5] and Multi-PIE [10] to build the reference set for all experiments except the experiment on the LFW database [4], for which we use a reference set with images from the WDRef dataset [28].

We refer to offline processing as all that is done before a probe enters the system and it is ready to be recognized. During offline processing, the gallery images are compared, using first level similarity (FLS) with the reference set images and a reference-based descriptor is generated for each gallery image. The dimensionality of each RB descriptor is equal to the number of individuals in the reference set. The
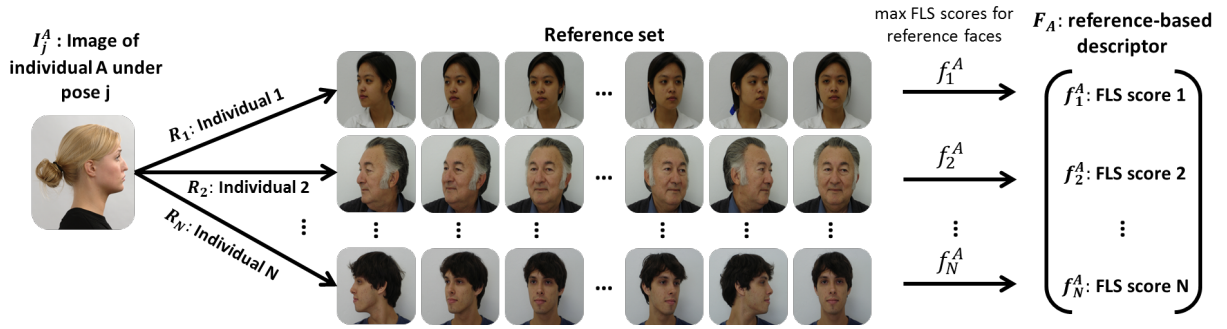
Fig. 2. Computing the reference-based descriptor. Any given gallery/probe image I_j^A is compared to all images of each reference face $R_1, R_2, \ldots, R_N$. The FLS score f_u^A is computed for I_j^A and each reference face R_u. f_u^A is the maximum similarity between I_j^A and all images of R_u. The RB descriptor for I_j^A, F^A, is defined as $< f_1^A, f_2^A, \ldots, f_N^A >$.

offline processing stage is complete after the reference-based descriptors are generated for all gallery images.

During online processing a probe image enters the system and its RB descriptor is generated. At this stage, RB descriptors are available for each of the gallery and the probe images. Recognition is performed using a similarity measure between the gallery and probe RB descriptors. We refer to this similarity as the *Second Level Similarity* (SLS).

For first level descriptors (FLD), we chose to use LBP due to the fact that LBP features are widely used in computer vision applications, especially face recognition [1]. Note that the reference-based algorithm described in this paper is independent of any specific preprocessing steps. For example, using a different facial feature extraction algorithm or replacing LBP with another face descriptor does not change how the reference-based algorithm operates.

A. Reference set and motivation

1) Reference set: The reference set is a set of faces belonging to N individuals. We call these individuals reference individuals and represent them with $R_1, R_2, \ldots, R_N$. We use two reference sets in this paper. The first reference set includes 200 individuals from the FEI database [5] and 300 individuals from Multi-PIE [10]. For each individual from FEI, we select 13 images under various poses from full profile left to full profile right. Thus, each pose from FEI has one image. For Multi-PIE, we randomly choose 300 individuals from the total of 337. Multi-PIE includes 15 viewpoints per individual. We randomly chose 13 of them so that the number of poses per individual in our reference set would be similar for all individuals (not a requirement).

For the experiment on LFW (Section IV-C.4), we use a reference set with images from the WDRef [28] dataset which includes 99,773 images of 2995 individuals. LFW and WDRef both contain unconstrained face images collected from the internet; thus, we test our framework on LFW with a reference set from WDRef. We choose 500 individuals (which have 40 images each) from WDRef to build our reference set. There are no common individuals between the LFW test individuals and the reference set.

2) Motivation for reference-based descriptors: Our technique is based on the following two intuitions:

- A face image can be described by its degree of similarity to the images of a set of reference individuals. In this context, the reference individuals form a coordinate system in an N-dimensional space and each input face corresponds to a point in this space, the coordinates of which are determined by the similarity of the face image to the reference individuals.

- Two faces which are visually similar in one pose, for example profile, are also, to some extent, similar in other poses, for example frontal. In other words, we assume that visual similarity follows from underlying physical similarity in the real world. We take advantage of this phenomenon in the following way: compare B, a gallery/probe face image, with the images of all available poses of a reference individual from the reference set. The degree of similarity with the best matching image of the reference individual is a degree of the similarity of the two faces, B and the reference individual. By repeating this procedure for each one of the reference individuals, we create a descriptor for the face image B that reflects the degree of similarity between B and the reference set faces.

B. Reference-based face recognition

Let us first define some terminology, and then we will describe the reference-based approach to face recognition across pose. In what follows, we use the following terminology: We use the term *input face* to denote the face for which we want to create a descriptor which may belong to the gallery or probe set. *Input image* denotes the image of the input face that we have at our disposal. For a given face A, we use the notation I_j^A to refer to an image of A where j is used to differentiate this image from other images of A. For example, j could refer to the pose under which the image was taken.

Reference-based face recognition consists of two main stages: computing reference-based descriptors and performing recognition using these descriptors.

1) Computing reference-based descriptors: Figure 2 illustrates the setup of generating the reference-based descriptor.

Typically, we have an image I_j^A (a probe or gallery image)

of individual A for which we wish to construct a reference-based descriptor, which is an N-dimensional real vector. Let

$$F^A = <f_1^A, f_2^A, \ldots, f_N^A> \qquad (1)$$

be the RB descriptor for A. The idea is that the elements $f_1^A, f_2^A, \ldots, f_N^A$ represent the similarity of A to the reference individuals $R_1, R_2, \ldots, R_N$. We estimate the similarity of A to a given reference individual by selecting the maximum FLS score between A and all images of a given reference individual R_u. In other words, to compute f_u^A, we compare the image A with each one of the images of the reference individual R_u in all different poses by computing $\text{FLS}(I_j^A, I_i^{R_u})$, which is a measure of the similarity of the two images. Thereafter, the largest FLS value obtained from this comparison is used as the value of f_u^A. The point is that the reference set would have some images that are close to the variation that the input image presents. Our framework does not require that the most similar reference set face image to a probe image has to be under the same pose. Thus,

$$f_u^A = \max(\text{FLS}(I_j^A, I_1^{R_u}), \ldots, \text{FLS}(I_j^A, I_K^{R_u})). \qquad (2)$$

Equation (2) is computed for all reference individuals, i.e., given A, f_u^A is computed for $u = 1 \ldots N$. These N values represent the N-dimensional RB descriptor for A.

2) Face recognition using RB descriptors: Images from probe and gallery are compared using the SLS measure between RB descriptors. In this paper, we use the cosine between the RB descriptors as the SLS measure, i.e.,

$$SLS(F^A, F^B) = \cos(F^A, F^B) = \frac{F^A \cdot F^B}{||F^A|| \times ||F^B||}. \qquad (3)$$

In other words, given two images I_u^A and I_v^B, and their corresponding RB descriptors F^A and F^B, $\cos(F^A, F^B)$ represents the second level similarity between the two images. As mentioned previously, SLS is the measure we use for ranking the gallery images for their degree of similarity to the probe image.

C. The first level similarity variants

We have investigated three variants for the first level similarity measure in this paper. The second level similarity computation is the same for all these variants. The three FLS variants are PCA-cosine method, Chi-square method, and pose-specific Chi-square method. All three variants have the following initial steps in common. The PCA-cosine method does not require pose estimation and achieves the highest accuracy in our experiments. The other two methods are proposed as a baseline for comparison.

After preprocessing, LBP features are computed for each image. Before extracting the LBP features, the face images are cropped and resized to 200×200 and divided to 20×20 regions. For each region the uniform LBP with a radius of 1 and 8 neighboring pixels are computed, and a histogram of 59 possible patterns is generated [1]. Histograms of all 100 regions are concatenated in a single histogram of length 5900 to describe the face image.

1) PCA-cosine method: PCA [29] is commonly used for decorrelation and dimensionality reduction. Utilizing PCA, the principal components are generated and then used to convert the high-dimensional feature subspace of the reference/probe/gallery face images to a smaller eigenspace subspace. We use it to speed up and improve the accuracy of the first level similarity calculation.

We use the LBP descriptors of the images in the reference set to derive the PCA coefficient matrix. Once the coefficient matrix is derived, it is applied to the LBP descriptors of all the images: the reference set, probes, and gallery. Thereafter, the reduced dimension vectors are used for computing first level similarity.

The cosine similarity measure is utilized to measure FLS in the PCA-cosine method. After projecting the LBP image histograms into eigenspace, they are no longer region-based histograms. Therefore, we choose to use cosine similarity instead of the commonly used Chi-square distance measure to measure FLS. The cosine measure works very well as the distance measure in this context, and is significantly cheaper to compute than Chi-square on the original LBP vectors.

2) Chi-square method: In the Chi-square method, unlike the PCA-cosine method, the original LBP feature vectors for the reference set, gallery, and probe face images are directly used to compute the FLS. Initially, to compute the FLS scores, the Chi-square distance

$$x = \chi^2(P, Q) = \sum_{b,w} \frac{(P_{b,w} - Q_{b,w})^2}{P_{b,w} + Q_{b,w}} \qquad (4)$$

is measured between reference set LBP histograms and gallery/probe LBP histograms. In Equation (4), P and Q are the LBP histograms, b denotes the histogram bin number, and w refers to the local region. We chose Chi-square distance because it is the most commonly used distance measure for LBP features [1] in the literature and has proven to be accurate.

The distance x is converted to similarity score using the Chi-square cumulative distribution function

$$\text{FLS} = 1 - \frac{\gamma(\frac{k}{2}, \frac{x}{2})}{\Gamma(\frac{k}{2})}, \qquad (5)$$

where $\Gamma(\frac{k}{2})$ is the Gamma function with closed form values for half-integer arguments, γ is the lower incomplete gamma function, and k refers to the degrees of freedom (in this case 59). Figure 3 shows the plot for FLS computed for a sample distance vector with values from 1 to 200.
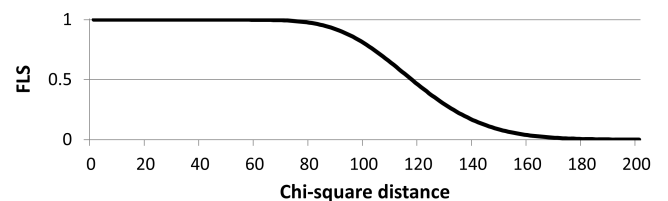

Fig. 3. FLS plot for a sample distance vector.

A limitation of the Chi-square method is its speed. The computation of the Chi-square distance is expensive, and yet it must be carried out to compare each probe/gallery image to all the images in the reference set. To overcome this problem, we investigate the pose specific Chi-square method. More detail on the processing times is discussed in Section IV-D.

3) Pose specific Chi-square method: In the pose specific Chi-square method, each probe/gallery face image is only compared to a single image of each reference individual. The image chosen from each individual for generating the similarity feature descriptor should be of the same pose as the probe/gallery image for which the similarity face descriptor is to be generated. Doing such accelerates the processing time dramatically because it does not require comparing the probe/gallery images with every image of each reference individual. Another difference between this method and the Chi-square method is that when generating the reference-based descriptor, only one similarity score per reference individual will be available to choose from. This similarity score is used for the RB descriptor of the probe/gallery images. To use this technique, the pose of the probe and gallery images needs to be estimated.

IV. EXPERIMENTAL RESULTS

A. Data

We use face images from seven face databases: LFW [4], FEI [5], RaFD [6], FacePix [8], CMU-PIE [9], Multi-PIE [10], and the Facial Recognition Technology (FERET) database [7]. The FEI [5] face database consists of 200 individuals. For each individual 14 images are available, 11 different poses from full profile left to full profile right, one image with a smiling expression, and 2 images with different illumination conditions. The RaFD face database [6] includes multiple images with various camera angles, expressions, and gaze directions for 67 individuals. For each individual we use 5 images with neutral expression and the following poses: profile left, half left, full frontal, half right, and profile right. The Multi-PIE [10] is a collection of $755,370$ images from 337 individuals. Images are taken from 15 viewpoints under 19 illumination settings. The CMU-PIE database [9] includes 41368 images of 68 individuals, with 13 poses, 43 illumination conditions, and 4 facial expressions. The FacePix [8] database holds images of 30 individuals, with images spanning the spectrum of $180°$ with increments of $1°$. The FERET face database [7] contains 14126 images of 1199 individuals and is commonly used within the face recognition research community. The LFW database [4] includes 13233 images of 5749 individuals from the web.

The gallery consists of frontal face images chosen from all aforementioned databases, except for the LFW database which is discussed in Section IV-C.4. In this paper, we only use frontal face images in the gallery; however, we have validated our approach (with no changes) on a gallery with frontal and non-frontal face images and the results are similar. This is due to the fact that the reference set treats frontal and non-frontal images in the same manner.

Faces are first normalized in illumination using the method in [30]. Afterwards, face alignment is executed in two steps: pose estimation and template alignment. Head pose estimation is performed using the method in [31]. The face images are aligned with the template corresponding to their estimated pose using a set of extracted control points (eyes, mouth, and nose).

B. Recognition results

The results in this section are reported on a gallery with 1000 frontal face images, and 200 randomly selected probe images. The results are averaged after repeating the experiments 10 times, each time randomly selecting 200 new probe images.

1) Reference-based vs. direct LBP: The Cumulative Match Characteristic (CMC) curve in Figure 4 compares the PCA-cosine RB method with direct LBP. The plot clearly shows that our proposed method greatly improves the recognition rate compared to using LBP directly and extends LBP's capability to perform face recognition across pose.
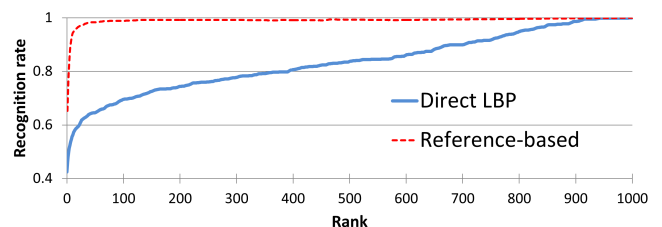

Fig. 4. CMC curve comparing the PCA-cosine RB vs. direct LBP.

2) Comparison of reference-based methods: CMC curves comparing the three proposed methods are presented in Figure 5. The results show that the PCA-cosine based method has better performance than the other two methods. The pose-specific Chi-square method reports lower recognition rates than the Chi-square method. This is because the pose estimation has great influence on the final recognition rate. The reference set images are labeled with the pose ID; however, the ID and the angle it corresponds to may be inaccurate. Therefore, when comparing the probe with the reference set, comparison may be performed between images with different poses which affects the recognition rate.
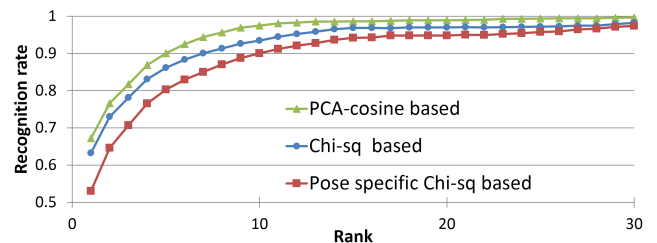

Fig. 5. Comparison of the three proposed RB methods.

C. Comparison with other methods

To evaluate our framework's performance, we compare our results with multiple state-of-the-art algorithms. Also, Sec-

tion IV-C.4 demonstrates how the proposed framework performs on the LFW database [4].

1) Comparison with Doppelgänger Lists [25]: Table III presents results comparing our proposed method with the Doppelgänger list approach [25]. For a fair comparison, the experimental setup is similar to that performed in [25]. For this experiment, images are selected from the FacePix database [8], FPLBP features [32] are generated for all images, and 10 test sets each containing 500 positive and 500 negative pairs are used to perform face verification (pair-wise matching). The results show that reference-based recognition is superior to Doppelgänger list comparison for all experimented pose ranges. On the Multi-PIE dataset [10], Doppelgänger list comparison reports $74.5\% \pm 2.6$ whereas our reference-based recognition achieves $80.3\% \pm 1.1$.

TABLE III
CLASSIFICATION ACCURACY OVER VARIOUS POSE RANGES

Probe 1	Probe 2	Doppelgänger list [25]	Reference-based (this paper)
$-30°$ to $30°$	$-30°$ to $30°$	$74.5\% \pm 2.6$	$80.3\% \pm 0.9$
$-90°$ to $90°$	$-90°$ to $90°$	$68.3\% \pm 1.0$	$75.2\% \pm 0.8$
$-10°$ to $10°$	angle $> \lvert 70° \rvert$	$66.9\% \pm 1.0$	$74.4\% \pm 0.8$

2) Comparison with associate-predict model [27]: We compare reference-based recognition with the associate predict (AP) model [27] using images from the Multi-PIE database. The results in Figure 6 demonstrate that reference-based (RB) recognition has better performance than both AP and likelihood-predict (LP) models.
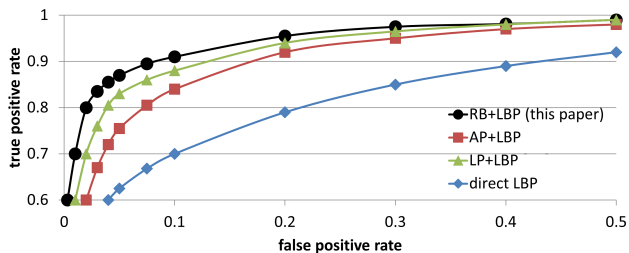

Fig. 6. Reference-based (RB) vs. AP [27], LP [27], and direct LBP [1].

3) Comparison with 3D pose normalization [18]: We also compare unified reference-based recognition with 3D pose normalization introduced in [18]. 3D pose normalization is limited to poses between $-45°$ to $45°$ in yaw, and is incapable of processing full profile face images. The results in Table IV show that 3D pose normalization performs better than the proposed reference-based recognition for close to frontal poses (i.e., $-30°$ to $30°$); however, for larger pose angles (i.e., $-45°$ to $-31°$ and $31°$ to $45°$) reference-based recognition has better performance than 3D pose normalization. The reason for such results is that 3D pose normalization is unable to accurately reconstruct the frontal face when the pose angle is large and facial features differ significantly.

In Table IV, for fair comparison, we use Local Gabor Binary Patterns (LGBP) [33] as the feature to compare our results with the 3D pose normalization method [18]. The

TABLE IV
RANK-1 RECOGNITION RATES (%) FOR FACEPIX DATABASE

Pose $\rightarrow$	-90 to -46	-45 to -31	-30 to -16	-15 to -1
3D [18]	—	71.6%	90.0%	97.3%
RB+LGBP	70.6%	78.9%	84.7%	90.1%
RB+LGTF	75.1%%	83.6%	88.6%	94.4%

Pose $\rightarrow$	1 to 15	16 to 30	31 to 45	46 to 90
3D [18]	95.8%	92.7%	74.8%	—
RB+LGBP	89.9%	87.2%	81.3%	73.2%
RB+LGTF	94.2%	91.0%	85.8%	77.6%

performance of our reference-based framework relies on the first level similarity; thus, we also show results using the LGTF (LBP+Gabor+TPLBP+FPLBP) feature introduced in [32]. Although the reference-based framework is proposed for recognition across pose, it also performs well for near frontal faces. Using multiple databases, especially the LFW database, shows that the proposed reference-based framework, works well for unconstrained face images and frontal face images as well as images with large pose variations.

4) Comparison on LFW database [4]: The face images in the LFW database [4] have a large degree of variability in pose and illumination settings. Compared to the other databases used in this paper, LFW is the only database with images taken under unconstrained settings. We show our proposed method's performance on the LFW database by comparing it to the associate-predict (AP) and likelihood-predict (LP) models introduced in [27], reference-based verification with message passing model [34], and Cosine similarity metric learning [35]. For fair comparison, we have chosen those methods which use LBP features. We do not compare with other approaches using more advanced features (e.g., [36], [17]).

For the experiment on the LFW database [4] we use a reference set with images taken from the WDRef [28] dataset which includes 99,773 images of 2995 individuals. LFW and WDRef both contain unconstrained face images collected from the Internet; thus, we test our framework on the LFW database with a reference set from the WDRef dataset. We choose 500 individuals (which have 40 images each) from the WDRef dataset to build our reference set. There are no common individuals between the LFW test individuals and the reference set. Figure 7 presents the results.
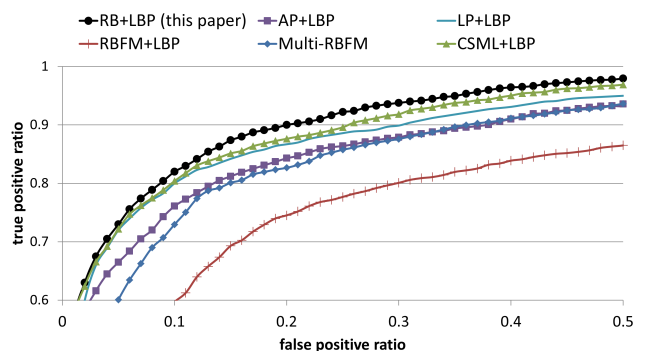

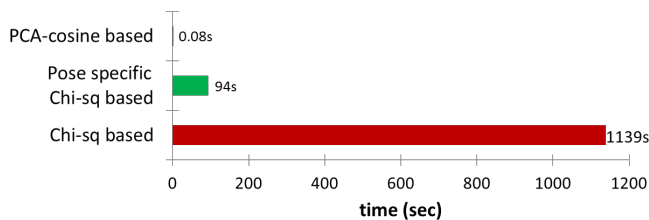Fig. 7. Reference-based (RB) vs. other algorithms on LFW database [4].

Fig. 8. Processing Time Comparison.

D. Processing time comparison

Figure 8 compares the processing time between the PCA-cosine RB and the other proposed RB recognition algorithms. The reported time in Figure 8 refers to the processing time required for computing the first level similarity scores and generating the RB descriptors. Clearly, using PCA not only improved recognition performance but also accelerated the recognition process dramatically. Our C++ and Matlab implementation was tested on an HP workstation laptop with Intel Core i7 CPU and 8GB of RAM. Feature extraction time is not included due to compartmentalization in our experiments. The reason for not reporting the entire processing time is because the time to generate the LBP histograms and the time to measure the second level similarity scores is the same for all three methods.

V. CONCLUSIONS

In recognition across pose, the probe and gallery have very different face descriptors which make it a challenging problem. We overcome this challenge by introducing a unified reference-based face recognition framework. The key idea is to create a reference-based descriptor for the probe and gallery face images by comparing them to a reference set, instead of comparing the probe and gallery images directly. The proposed algorithm was used in conjunction with various descriptors such as LBP, LGBP, and FPLBP, and it was compared with state-of-the-art across-pose recognition algorithms . In performing many comparisons and experiments on seven challenging databases (LFW, FERET, FEI, RaFD, FacePix, CMU-PIE, and Multi-PIE), we found that the proposed approach outperforms previous state-of-the-art methods in terms of identification rate and verification accuracy. Besides, reference-based recognition is computationally inexpensive and can be performed very efficiently.

REFERENCES

[1] Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: Application to face recognition. IEEE TPAMI **28** (2006) 2037–2041
[2] An, L., Kafai, M., Bhanu, B.: Dynamic Bayesian network for unconstrained face recognition in surveillance camera networks. IEEE JETCAS **3** (2013) 155–164
[3] Prabhu, U., Heo, J., Savvides, M.: Unconstrained pose-invariant face recognition using 3D generic elastic models. IEEE TPAMI **33** (2011) 1952–1961
[4] Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07–49, UMass Amherst (2007)
[5] Thomaz, C.E., Giraldi, G.A.: A new ranking method for principal components analysis and its application to face image analysis. Image and Vision Computing **28** (2010) 902 – 913
[6] Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D., Hawk, S., Van Knippenberg, A.: Presentation and validation of the Radboud faces database. Cognition and Emotion **24(8)** (2010) 1377–388
[7] Phillips, P., Moon, H., Rauss, P., Rizvi, S.: The FERET evaluation methodology for face-recognition algorithms. In: CVPR. (1997)
[8] Little, D., Krishna, S., Black, J., Panchanathan, S.: A methodology for evaluating robustness of face recognition algorithms with respect to variations in pose angle and illumination angle. In: ICASSP. (2005)
[9] Sim, T., Baker, S., Bsat, M.: The CMU pose, illumination, and expression (PIE) database. In: FG. (2002) 46–51
[10] Gross, R., Matthews, I., Cohn, J., Kanade, T., Baker, S.: Multi-PIE. Image and Vision Computing **28** (2010) 807–813
[11] Kahraman, F., Kurt, B., Gokmen, M.: Robust face alignment for illumination and pose invariant face recognition. In: CVPR. (2007)
[12] Huang, J., Yuen, P., Chen, W.S., Lai, J.H.: Choosing parameters of kernel subspace LDA for recognition of face images under pose and illumination variations. IEEE TSMC. B, Cybern. **37** (2007)
[13] Arashloo, S., Kittler, J.: Energy normalization for pose-invariant face recognition based on MRF model image matching. IEEE TPAMI **33** (2011) 1274–1280
[14] An, L., Bhanu, B., Yang, S.: Boosting face recognition in real-world surveillance videos. In: AVSS. (2012)
[15] Li, A., Shan, S., Gao, W.: Coupled bias-variance tradeoff for cross-pose face recognition. IEEE TIP (2012) 305–315
[16] Li, H., Hua, G., Lin, Z., Brandt, J., Yang, J.: Probabilistic elastic matching for pose variant face verification. In: CVPR. (2013)
[17] Zhang, Y., Shao, M., Wong, E., Fu, Y.: Random faces guided sparse many-to-one encoder for pose-invariant face recognition. In: ICCV. (2013) 2416–2423
[18] Asthana, A., Marks, T., Jones, M., Tieu, K., Rohith, M.: Fully automatic pose-invariant face recognition via 3D pose normalization. In: ICCV. (2011) 937–944
[19] Yi, D., Lei, Z., Li, S.: Towards pose robust face recognition. In: CVPR. (2013) 3539–3545
[20] Taigman, Y., Yang, M., Ranzato, M., Wolf, L.: Deepface: Closing the gap to human-level performance in face verification. In: CVPR. (2014) 1701–1708
[21] Duin, R.P., Pekalska, E.: The dissimilarity space: Bridging structural and statistical pattern recognition. Pattern Recognition Letters (2012)
[22] Guo, Y., Shan, Y., Sawhney, H., Kumar, R.: Peet: Prototype embedding and embedding transition for matching vehicles over disparate viewpoints. In: CVPR. (2007) 1–8
[23] An, L., Kafai, M., Yang, S., Bhanu, B.: Reference-based person re-identification. In: AVSS. (2013) 244–249
[24] Liu, J., Kuipers, B., Savarese, S.: Recognizing human actions by attributes. In: CVPR. (2011) 3337–3344
[25] Schroff, F., Treibitz, T., Kriegman, D., Belongie, S.: Pose, illumination and expression invariant pairwise face-similarity measure via Doppelgänger list comparison. In: ICCV. (2011) 2494–2501
[26] Kumar, N., Berg, A., Belhumeur, P., Nayar, S.: Attribute and simile classifiers for face verification. In: ICCV. (2009) 365–372
[27] Yin, Q., Tang, X., Sun, J.: An associate-predict model for face recognition. In: CVPR. (2011) 497–504
[28] Chen, D., Cao, X., Wang, L., Wen, F., Sun, J.: Bayesian face revisited: A joint formulation. In: ECCV. (2012) 566–579
[29] Wold, S., Esbensen, K., Geladi, P.: Principal component analysis. Chemometrics and Intelligent Laboratory Systems (1987) 37–52
[30] Tan, X., Triggs, B.: Enhanced local texture feature sets for face recognition under difficult lighting conditions. IEEE TIP (2010)
[31] Kafai, M., Bhanu, B., An, L.: Cluster-classification Bayesian networks for head pose estimation. In: ICPR. (2012)
[32] Wolf, L., Hassner, T., Taigman, Y.: Effective unconstrained face recognition by combining multiple descriptors and learned background statistics. IEEE TPAMI **33** (2011) 1978–1990
[33] Zhang, W., Shan, S., Gao, W., Chen, X., Zhang, H.: Local gabor binary pattern histogram sequence (LGBPHS): a novel non-statistical model for face representation and recognition. In: ICCV. (2005) 786–791
[34] Shen, W., Wang, B., Wang, Y., Bai, X., Latecki, L.J.: Face identification using reference-based features with message passing model. Neurocomputing **99** (2013) 339–346
[35] Nguyen, H.V., Bai, L.: Cosine similarity metric learning for face verification. In: ACCV. (2010) 709–720
[36] Chen, D., Cao, X., Wen, F., Sun, J.: Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification. In: CVPR. (2013) 3025–3032