Prototype-Based Discriminative Feature Learning for Kinship Verification

Haibin Yan, Jiwen Lu, Member, IEEE, and Xiuzhuang Zhou, Member, IEEE

Abstract—In this paper, we propose a new prototype-based discriminative feature learning (PDFL) method for kinship verification. Unlike most previous kinship verification methods which employ low-level hand-crafted descriptors such as local binary pattern and Gabor features for face representation, this paper aims to learn discriminative mid-level features to better characterize the kin relation of face images for kinship verification. To achieve this, we construct a set of face samples with unlabeled kin relation from the labeled face in the wild dataset as the reference set. Then, each sample in the training face kinship dataset is represented as a mid-level feature vector, where each entry is the corresponding decision value from one support vector machine hyperplane. Subsequently, we formulate an optimization function by minimizing the intraclass samples (with a kin relation) and maximizing the neighboring interclass samples (without a kin relation) with the mid-level features. To better use multiple low-level features for mid-level feature learning, we further propose a multiview PDFL method to learn multiple mid-level features to improve the verification performance. Experimental results on four publicly available kinship datasets show the superior performance of the proposed methods over both the state-of-the-art kinship verification methods and human ability in our kinship verification task.

Index Terms—Discriminative learning, feature learning, kinship verification, mid-level feature representation, prototype, soft biometrics.

I. INTRODUCTION

R ECENT advances in psychology and cognitive sciences [2], [6], [7], [24] have revealed that human face is an important cue for kin similarity measure as children usually look like their parents more than other adults because children and their parents are biologically related and have overlapped genetics. Inspired by this finding, there have been some seminal attempts on kinship verification via hand-crafted feature descriptors [15], [16], [18], [26], [42], [45], [49], [50], [52], [56], [57]. While there are many potential applications for kinship verification such as missing children searching and social media mining, it is still challenging to develop a robust kinship verification system for real applications because there are usually large variations on pose, illumination, expression, and aging on facial images, especially when face images are captured in unconstrained environments. While the past five years have witnessed encouraging progress in this area [15], [16], [18], [26], [42], [45], [49], [50], [52], [56], [57], the problem of kinship verification still remains unsolved because it is extremely challenging to extract kin-related features from human ages, especially when face images are captured in the wild.

In this paper, we propose a new prototype-based discriminative feature learning (PDFL) method for kinship verification. Fig. 1 shows the pipeline of our proposed approach. Unlike most previous kinship verification works where low-level hand-crafted feature descriptors [15], [16], [18], [26], [42], [45], [49], [50], [52], [56], [57] such as local binary pattern (LBP) [1], [9] and Gabor features [29], [57] are employed for face representation, we learn discriminative mid-level features to better characterize the relation of face images for kinship verification. To achieve this, we construct a set of face samples with unlabeled kinship relation from the labeled face in the wild (LFW) dataset [22] as the reference set. Then, each sample in the training set with a labeled kin relation is represented as a mid-level feature vector. Then, we formulate an objective function by minimizing the intraclass samples (with a kinship relation) and maximizing the interclass samples with the mid-level features. To make better use of multiple low-level features for mid-level feature learning, we further propose a multiview PDFL (MPDFL) method to learn multiple mid-level features to improve the verification performance. Experimental results on four publicly available kinship datasets show the superior performance of the proposed methods over the state-of-the-art methods. We also compare our methods with human ability in kinship verification and experimental results have shown that our methods achieve better performance than humans in our kinship verification task.

The rest of this paper is organized as follows. Section II reviews the related work. Section III presents the proposed approach. Section IV shows the experimental results, and Section V concludes this paper.
Fig. 1. Pipeline of our proposed kinship verification approach. First, we construct a set of face samples from the LFW dataset as the prototypes and represent each face image from the kinship dataset as a combination of these prototypes in the hyperplane space. Then, we use the labeled kinship information and learn mid-level features in the hyperplane space to extract more semantic information for feature representation. Lastly, the learned hyperplane parameters are used to represent face images in both the training and testing sets as a discriminative mid-level feature for kinship verification.

<table>
<thead>
<tr>
<th>Method</th>
<th>Feature representation</th>
<th>Classifier</th>
<th>Dataset</th>
<th>Accuracy (%)</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fang et al. [16]</td>
<td>Local features from different face parts</td>
<td>KNN</td>
<td>Cornell KinFace</td>
<td>70.7</td>
<td>2010</td>
</tr>
<tr>
<td>Zhou et al. [56]</td>
<td>Spatial pyramid local feature descriptor</td>
<td>SVM</td>
<td>400 pairs (N.A.)</td>
<td>67.8</td>
<td>2011</td>
</tr>
<tr>
<td>Xia et al. [50]</td>
<td>Context feature with transfer learning</td>
<td>KNN</td>
<td>UB KinFace</td>
<td>68.5</td>
<td>2012</td>
</tr>
<tr>
<td>Guo and Wang [18]</td>
<td>DAISY descriptors from semantic parts</td>
<td>Bayes</td>
<td>200 pairs (N.A.)</td>
<td>75.0</td>
<td>2012</td>
</tr>
<tr>
<td>Zhou et al. [57]</td>
<td>Gabor gradient orientation pyramid</td>
<td>SVM</td>
<td>400 pairs (N.A.)</td>
<td>69.8</td>
<td>2012</td>
</tr>
<tr>
<td>Kohli et al. [36]</td>
<td>Self similarity of Weber face</td>
<td>SVM</td>
<td>IITD KinFace</td>
<td>[64.2, 74.1]</td>
<td>2012</td>
</tr>
<tr>
<td>Lu et al. [42]</td>
<td>Local feature with metric learning</td>
<td>SVM</td>
<td>KinFace-WI</td>
<td>69.9</td>
<td>2014</td>
</tr>
<tr>
<td>Guo et al. [17]</td>
<td>LBP feature</td>
<td>Logistic regressor + fusion</td>
<td>322 pairs (available upon request)</td>
<td>69.3</td>
<td>2014</td>
</tr>
<tr>
<td>Our method</td>
<td>Mid-level features by discriminative learning</td>
<td>SVM</td>
<td>KinFace-WI</td>
<td>70.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>KinFace-WI</td>
<td>77.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Cornell KinFace</td>
<td>71.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>UB KinFace</td>
<td>67.3</td>
<td></td>
</tr>
</tbody>
</table>

*If the method was presented in a conference version and then extended to a journal, we only list its journal version for comparison.

II. RELATED WORK

In this section, we briefly review two related topics: 1) kinship verification and 2) feature learning.

A. Kinship Verification

Kinship verification via facial image analysis is an interesting problem in computer vision. In recent years, there have been a few seminal studies in [15]–[18], [42], [45], [49], [50], [52], [56], and [57]. Existing kinship verification methods can be mainly categorized into two classes: 1) feature-based [11], [16], [18], [56], [57] and 2) model-based [17], [42], [50], [52]. Generally, feature-based methods aim to extract discriminative feature descriptors to effectively represent facial images so that stable kin-related characteristics can be well preserved. Existing feature representation methods include skin color [16], histogram of gradient [16], [45], [56], Gabor wavelet [13], [45], [50], [57], gradient orientation pyramid [57], LBP [8], [42], scale-invariant feature transform (SIFT) [42], [45], [52], salient part [18], [49], self-similarity [26], and dynamic features combined with spatio-temporal appearance descriptor [11]. Model-based methods usually apply some statistical learning techniques to learn an effective classifier, such as subspace learning [50], metric learning [42], transfer learning [50], multiple kernel learning [57] and graph-based fusion [17]. Table I briefly reviews and compares existing kinship verification methods for facial kinship modeling over the past five years, where the performance of these methods is evaluated by the mean verification rate. While the verification rate of different methods in this table cannot be directly compared due to different datasets and experimental protocols, we still see that a major progress of kinship verification have been obtained in recent years. More recently, Fang et al. [15] extended kinship verification to kinship classification. In their work, they proposed a kinship classification approach by reconstructing the query face from a sparse set of samples among the candidates for family classification, and 15% rank-one classification rate was achieved on a dataset consisting of 50 families. Unlike previous kinship verification works [15]–[18], [42], [45], [49], [50], [52], [56], [57], this paper presents a new feature learning method to learn mid-level features to better characterize
facial images for kinship verification, so that more discriminative information can be exploited than the original low-level features.

B. Feature Learning

Recently, there has been growing interest in unsupervised feature learning and deep learning in computer vision and machine learning, and a variety of feature learning methods have been proposed in [12], [19], [21], [23], [25], [27], [28], [34], [44], [46], [51], and [53] to learn feature representations from raw pixels. Generally, feature learning methods exploit some prior knowledge such as smoothness, sparsity, and temporal and spatial coherence [3]. Representative feature learning methods include sparse auto-encoder [44], [46], restricted Boltzmann machines [19], independent subspace analysis [10], [28] and convolutional neural networks [27].

These methods have been successfully applied in many computer vision tasks such as image classification [27], human action recognition [23], face recognition [21], and visual computer vision tasks such as image classification [27], human analysis [10], [28] and convolutional neural networks [27]. Restricted Boltzmann machines [19], independent subspace analysis [10], [28] and convolutional neural networks [27].

Representative feature learning methods include sparse auto-encoder [44], [46], restricted Boltzmann machines [19], independent subspace analysis [10], [28] and convolutional neural networks [27].

III. PROPOSED APPROACH

Low-level feature descriptors such as LBP [1] and SIFT [30] are usually ambiguous, which are not discriminative enough for kinship verification, especially when face images were captured in the wild because there are large variations on face images captured in such scenarios. To extract more semantic information from low-level features, we learn mid-level discriminative features with low-level descriptor, where each entry in the mid-level feature vector is the corresponding decision value from one support vector machine (SVM) hyperplane. We formulate an optimization objective function on the learned features so that face samples with a kin relation are expected to have similar decision values from these hyperplanes. Hence, our method is complementary to the exiting feature learning methods.

A. PDFL

Let \( Z = [z_1, z_2, \ldots, z_N] \in \mathbb{R}^{d \times N} \) be a unlabeled reference image set, where \( N \) and \( d \) are the number of samples and feature dimension of each sample, respectively. Assume \( S = (x_1, y_1), \ldots, (x_i, y_i), \ldots, (x_M, y_M) \) be the training set containing \( M \) pairs of face images with kinship relation (positive image pairs), where \( x_i \) and \( y_i \) are face images of the \( i \)th pair, and \( x_i \in \mathbb{R}^d \) and \( y_i \in \mathbb{R}^d \). Different from most existing feature learning methods which learn feature representations from raw pixels, we aim to learn a set of mid-level features which are obtained from a set of prototype hyperplanes. For each training sample in \( S \), we apply the linear SVM to learn the weight vector \( w \) to represent it as follows:

\[
\begin{align*}
    w &= \sum_{n=1}^{N} \alpha_n z_n = \sum_{n=1}^{N} \beta_n z_n = Z \beta \\
\end{align*}
\]

where \( \alpha_n \) is the label of the unlabeled data \( z_n \), \( \beta_n = \alpha_n l_n \) is the combination coefficient, \( \alpha_n \) is the dual variable, and \( \beta = [\beta_1, \beta_2, \ldots, \beta_M] \in \mathbb{R}^{N \times 1} \) is the coefficient vector. Specifically, if \( \beta_n \) is nonzero, it means that the sample \( z_k \) in the unlabeled reference set is selected as a support vector of the learned SVM model, and \( l_n = 1 \) if \( \beta_n \) is positive. Otherwise, \( l_n = -1 \). Motivated by the maximal margin principle of SVM, we only need to select a sparse set of support vectors to learn the SVM hyperplane. Hence, \( \beta \) should be a sparse vector, where \( \| \beta \|_1 \leq \gamma \), and \( \gamma \) is a parameter to control the sparsity of \( \beta \).

Having learned the SVM hyperplane, each training sample \( x_i \) and \( y_i \) can be represented as

\[
\begin{align*}
    f(x_i) &= w^T x_i = x_i^T Z \beta \\
    f(y_i) &= w^T y_i = y_i^T Z \beta. \\
\end{align*}
\]

Assume we have learned \( K \) linear SVM hyperplanes, then the mid-level feature representations of \( x_i \) and \( y_i \) can be represented as

\[
\begin{align*}
    f(x_i) &= [x_i^T Z \beta_1, x_i^T Z \beta_2, \ldots, x_i^T Z \beta_K] = B^1 Z^1 x_i \\
    f(y_i) &= [y_i^T Z \beta_1, y_i^T Z \beta_2, \ldots, y_i^T Z \beta_K] = B^1 Z^1 y_i \\
\end{align*}
\]

where \( B = [\beta_1, \beta_2, \ldots, \beta_K] \) is the coefficient matrix.

Now, we propose the following optimization criterion to learn the coefficient matrix \( B \) with the sparsity constraint:

\[
\begin{align*}
    \max H(B) &= H_1(B) + H_2(B) - H_3(B) \\
    &= \frac{1}{M_k} \sum_{i=1}^{M} \sum_{t=1}^{k} \| f(x_{it}) - f(y_{it}) \|_2^2 \\
    &\quad + \frac{1}{M_k} \sum_{i=1}^{M} \sum_{t=1}^{k} \| f(x_{it}) - f(y_{it}) \|_2^2 \\
    &\quad - \frac{1}{M} \sum_{i=1}^{M} \| f(x_i) - f(y_i) \|_2^2 \\
    \text{subject to } &\| \beta_k \|_1 \leq \gamma, k = 1, 2, \ldots, K. \\
\end{align*}
\]

where \( y_{it} \) represents the \( t \)th \( k \)-nearest neighbor of \( y_i \) and \( x_{it} \) denotes the \( t \)th \( k \)-nearest neighbor of \( x_i \), respectively. The objectives of \( H_1 \) and \( H_2 \) in (6) is to make the mid-level feature representations of \( y_{it} \) and \( x_i \) and \( y_{it} \) and \( y_i \) as far as possible if they are originally near to each other in the low-level feature space. The physical meaning of \( H_3 \) in (6) is to expect that \( x_i \) and \( y_i \) are close to each other in the mid-level feature space. We enforce the sparsity constraint on \( \beta_k \) such that only a sparse set of support vectors from the unlabeled reference dataset are selected to learn the hyperplane because we assume each sample can be sparsely reconstructed the reference set, which is inspired by this paper in [25]. In this paper, we apply the same parameter \( \gamma \) to reduce the number...
of parameters in our proposed model so that the complexity of the proposed approach is reduced.

Combining (4)–(6), we simplify $H_1(B)$ to the following form:

$$H_1(B) = \frac{1}{M} \sum_{i=1}^{M} \sum_{t_1=1}^{k} \| B^T Z^T x_i - B^T Z^T y_{i1} \|_2^2$$

$$= \frac{1}{M} tr \left( \sum_{i=1}^{M} \sum_{t_1=1}^{k} B^T Z^T (x_i - y_{i1}) (x_i - y_{i1})^T Z \right)$$

$$= tr (B^T F_1 B)$$

(7)

where

$$F_1 = \frac{1}{M} \sum_{i=1}^{M} \sum_{t_1=1}^{k} Z^T (x_i - y_{i1}) (x_i - y_{i1})^T Z.$$  

Similarly, $H_2(B)$ and $H_3(B)$ can be simplified as

$$H_2(B) = tr (B^T F_2 B), \quad H_3(B) = tr (B^T F_3 B)$$

(9)

where

$$F_2 = \frac{1}{M} \sum_{i=1}^{M} \sum_{t_2=1}^{k} Z^T (x_{i2} - y_i)(x_{i2} - y_i)^T Z$$

$$F_3 = \sum_{i=1}^{M} Z^T (x_i - y_i) (x_i - y_i)^T Z.$$  

Based on (7)–(11), the proposed PDFL model can be formulated as follows:

$$\max H(B) = tr \left[ B^T (F_1 + F_2 - F_3) B \right]$$

subject to $B^T B = I$, $\| \beta_k \|_1 \leq \gamma, k = 1, 2, \ldots, K$  

(12)

where $B^T B = I$ is a constraint to control the scale of $B$ so that the optimization problem in (12) is well-posed with respect to $B$.

Since there is a sparsity constraint for each $\beta_k$, we cannot obtain $B$ by solving a standard eigenvalue equation. To address this, we propose an alternating optimization method in [43] by reformulating the optimization problem as a regression problem. Let $F \triangleq F_1 + F_2 - F_3$. We perform singular value decomposition (SVD) on $F = G^T G$, where $G \in \mathbb{R}^{N \times K}$. Following [43], we reformulate a regression problem by using an intermediate matrix $A = [a_1, a_2, \ldots, a_K] \in \mathbb{R}^{N \times K}$ (see Theorem 1 in [43])

$$\min H(A, B) = \sum_{k=1}^{K} \| G a_k - G \beta_k \|_2^2 + \lambda \sum_{k=1}^{K} \| \beta_k \|_1^2$$

subject to $A^T A = I_{K \times K}$, $\| \beta_k \|_1 \leq \gamma, k = 1, 2, \ldots, K$.  

(13)

We employ an alternating optimization method [43] to optimize $A$ and $B$ iteratively.

**Algorithm 1: PDFL**

**Input:** Reference set: $Z = [z_1, z_2, \ldots, z_N] \in \mathbb{R}^{d \times N}$, training set: $S = \{(x_i, y_i) | i = 1, 2, \ldots, M\}$, $x_i \in \mathbb{R}^{d}$ and $y_i \in \mathbb{R}^{d}$.

**Output:** Coefficient matrix $B = [\beta_1, \beta_2, \ldots, \beta_K]$.

**Step 1 (Initialization):**

Initialize $A \in \mathbb{R}^{N \times K}$ and $B \in \mathbb{R}^{N \times K}$ where each entry of them is set as 1.

**Step 2 (Local optimization):**

For $t = 1, 2, \ldots, T$ repeat

1. Compute $B$ according to (15).
2. Compute $A$ according to (16)-(17).
3. If $t > 2$ and $\| B_t - B_{t-1} \|_F \leq \epsilon$ (offset is set as 0.001 in our experiments), go to Step 3.

**Step 3 (Output coefficient matrix):**

Output coefficient matrix $B = B_t$.

1) **Fix $A$, optimize $B$:** For a given $A$, we solve the following problem to obtain $B$:

$$\min H(B) = \sum_{k=1}^{K} \| G a_k - G \beta_k \|_2^2 + \lambda \sum_{k=1}^{K} \| \beta_k \|_1^2$$

subject to $\| \beta_k \|_1 \leq \gamma, k = 1, 2, \ldots, K.$  

(14)

Considering that $\beta_k$ are independent in (14), we individually obtain $\beta_k$ by solving the following optimization problem:

$$\min H(\beta_k) = \| h_k - G \beta_k \|_2^2$$

subject to $\| \beta_k \|_1 \leq \gamma.$  

(15)

where $h_k = G a_k$, $g_k = [h_k^T, 0_K^T]^T$, $P = [G^T, \sqrt{\lambda} I_K]^T$, and $\beta_k$ can be easily obtained by using the conventional least angle regression solver [14].

2) **Fix $B$, optimize $A$:** For a given $B$, we solve the following problem to obtain $A$:

$$\min H(A) = \| G a - G B \|_2^2$$

subject to $A^T A = I_{K \times K}$.

(16)

And $A$ can be obtained by using SVD, namely

$$G^T G B = U S V^T, \quad A = U \tilde{V}^T$$  

(17)

where $\tilde{U} = [u_1, u_2, \ldots, u_K]$ be the top $K$ leading eigenvectors of $\tilde{U} = [u_1, u_2, \ldots, u_K]$.

We repeat the above two steps until the algorithm meets a certain convergence condition. The proposed PDFL algorithm is summarized in Algorithm 1.

**B. MPDFL**

Different feature descriptors usually capture complementary information to describe face images from different aspects [5] and it is helpful for us to improve the kinship verification performance with multiple feature descriptors. A nature solution for feature learning with multiview data is concatenating multiple features first and then applying existing feature learning methods on the concatenated features. However, it is not physically meaningful to directly combine different features because they usually show different statistical characteristics.


and such a concatenation cannot well exploit the feature diversity. In this paper, we introduce a MPDFL method to learn a common coefficient matrix with multiple low-level descriptors for mid-level feature representation for kinship verification.

Given the training set $S$, we first extract $L$ feature descriptors denoted as $S^l_1, \ldots, S^l_d$, where $S^l_i = (x^l_i, y^l_i), \ldots, (x^l_{M^l}, y^l_{M^l})$ is the $i$th feature representation, $1 \leq i \leq L$, $x^l_i \in R^{d^l}$ and $y^l_i \in R^{d^l}$ are the $i$th parent and child faces in the $i$th feature space, $l = 1, 2, \ldots, L$. MPDFL aims to learn a shared coefficient matrix $B$ with the sparsity constraint so that the intraclass variations are minimized and the interclass variations are maximized in the mid-level feature spaces.

To exploit complementary information from facial images, we introduce a nonnegative weighted vector $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_L]$ to weight each feature space of PDFL. Generally, the larger $\alpha_i$, the more contribution it is to learn the sparse coefficient matrix. MPDFL is formulated as the following objective function by using an intermediate matrix $A = [a_1, a_2, \ldots, a_K] \in R^{K \times K}$:

$$
\max_{B, \alpha} \sum_{l=1}^{L} \alpha_l tr \left[ B^T (F^l_1 + F^l_2 - F^l_3) B \right]
$$

subject to $B^T B = I$, $\| \beta_k \|_1 \leq \gamma, k = 1, 2, \ldots, K$, $\sum_{l=1}^{L} \alpha_l = 1, \alpha_l \geq 0$ (18)

where $F^l_1$, $F^l_2$, and $F^l_3$ are the expressions of $F_1$, $F_2$, and $F_3$ in the $i$th feature space, and $1 \leq l \leq L$.

Since the solution to (18) is $\alpha_l = 1$, which corresponds to the maximal $tr[B^T (F^l_1 + F^l_2 - F^l_3) B]$ over different feature descriptors, and $\alpha_p = 0$ otherwise. To address this, we revisit $\alpha_l$ as $\alpha^r_l$ ($r > 1$) and redefine the following optimization function as:

$$
\max_{B, \alpha} \sum_{l=1}^{L} \alpha^r_l tr \left[ B^T (F^l_1 + F^l_2 - F^l_3) B \right]
$$

subject to $B^T B = I$, $\| \beta_k \|_1 \leq \gamma, k = 1, 2, \ldots, K$, $\sum_{l=1}^{L} \alpha_l = 1, \alpha_l \geq 0$. (19)

Similar to PDFL, we also reformulate MPDFL as the following regression problem:

$$
\min_{A, B, \alpha} \sum_{l=1}^{L} \sum_{k=1}^{K} \alpha^r_l \| G_l a_k - G_l \beta_k \|^2 + \lambda \sum_{k=1}^{K} \beta_k^T \beta_k
$$

subject to $A^T A = I_{K \times K}$, $\| \beta_k \|_1 \leq \gamma, k = 1, 2, \ldots, K$. (20)

where $F^l_i = G^T_l G^l_i$, and $F^l_i = F^l_1 + F^l_2 - F^l_3$.

Since (20) is nonconvex with respect to $A$, $B$, and $\alpha$, we solve it iteratively similar to PDFL by using an alternating optimization method.

1) Fix $A$ and $B$, optimize $\alpha$: For the given $A$ and $B$, we construct a Lagrange function

$$
L(\alpha, \eta) = \sum_{l=1}^{L} \alpha^r_l tr \left[ B^T (F^l_1 + F^l_2 - F^l_3) B \right] - \zeta \left( \sum_{l=1}^{L} \alpha_l - 1 \right)
$$

Let $(\partial L(\alpha, \eta)/\partial \alpha_l) = 0$ and $(\partial L(\alpha, \eta)/\partial \zeta) = 0$, we have

$$
\alpha^r_l tr \left[ B^T (F^l_1 + F^l_2 - F^l_3) B \right] - \zeta = 0
$$

$$
\sum_{l=1}^{L} \alpha_l - 1 = 0
$$

Combining (22) and (23), we can obtain $\alpha_l$ as follows:

$$
\alpha_l = \frac{(1/\text{tr}[B^T (F^l_1 + F^l_2 - F^l_3) B])^{(r-1)}}{\sum_{l=1}^{L} (1/\text{tr}[B^T (F^l_1 + F^l_2 - F^l_3) B])^{(r-1)}}
$$

2) Fix $A$ and $\alpha$, optimize $B$: For the given $A$ and $\alpha$, we solve the following problem to obtain $B$:

$$
\min_{B, \alpha} H(B) = \sum_{l=1}^{L} \sum_{k=1}^{K} \alpha^r_l \| G_l a_k - G_l \beta_k \|^2 + \lambda \sum_{k=1}^{K} \beta_k^T \beta_k
$$

subject to $\| \beta_k \|_1 \leq \gamma, k = 1, 2, \ldots, K$. (25)

Similar to PDFL, we individually obtain $\beta_k$ by solving the following optimization problem:

$$
\min_{B, \alpha} H(\beta_k) = \sum_{l=1}^{L} \alpha^r_l \| G_l a_k - G_l \beta_k \|^2 + \lambda \sum_{k=1}^{K} \beta_k^T \beta_k
$$

subject to $\| \beta_k \|_1 \leq \gamma$.

where $h_k = \sum_{l=1}^{L} \alpha^r_l G_l a_k$, $g_k = [h_k^T, 0_K^L]^T$, $P = [\sum_{l=1}^{L} \alpha^r_l G_l, \sqrt{\lambda} I_K]$ and $\beta_k$ can be obtained by using the conventional least angle regression solver [14].

3) Fix $B$ and $\alpha$, optimize $A$: For the given $B$ and $\alpha$, we solve the following problem to obtain $A$:

$$
\min_{A, B, \alpha} H(A) = \sum_{l=1}^{L} \alpha^r_l \| G_l A - G_l B \|^2
$$

subject to $A^T A = I_{K \times K}$. (27)

And $A$ can be obtained by using SVD, namely

$$
\left( \sum_{l=1}^{L} \alpha^r_l G_l^T G_l \right) B = U S V^T, \quad A = U \hat{V}^T
$$

where $\hat{U} = [u_1, u_2, \ldots, u_K]$ be the top $K$ leading eigenvectors of $U = [u_1, u_2, \ldots, u_K]$.

We repeat the above three steps until the algorithm converges to a local optimum. Algorithm 2 summarizes the proposed MPDFL algorithm.
in the similarity measure stage. Hence, our feature learning approach is expected to obtain better performance when it is combined with state-of-the-art metric learning methods. Previous metric learning methods \[20\], \[32\]–\[41\], \[54\], \[55\] aim to learn a discriminative distance metric under which the intraclass variation of low-level features is minimized, so that the distance metric is learned directly on the training samples. In our approach, we employ a reference sample set and the mid-level feature is learned based on the responses of the training samples over the training set, so that the response coefficients are considered as the mid-level feature, which is more related to the attribute-based feature representation. Moreover, our mid-level feature learning approach and existing metric learning methods exploit discriminative information from different stages of the kinship verification task. Specially, our feature learning approach exploits discriminative information in the feature extraction stage while metric learning methods exploit such information in the similarity measure stage. Hence, our feature learning approach is expected to obtain better performance when it is combined with state-of-the-art metric learning methods to further improve the verification performance. More empirical results will be presented in Section IV to illustrate this point.

C. Discussion

In this subsection, we discuss the relationship and difference between our proposed mid-level feature learning approach and existing metric learning methods. Generally, both our feature learning approach and previous supervised metric learning methods can exploit discriminative information for the verification task such as face verification and kinship verification. However, our approach is intrinsically different from previous metric learning methods \[42\], \[52\]. Previous metric learning methods \[20\], \[32\]–\[41\], \[54\], \[55\] aim to learn a discriminative distance metric under which the intraclass variation of low-level features is minimized and the interclass variation of low-level features is maximized, so that the distance metric is learned directly on the training samples. In our approach, we employ a reference sample set and the mid-level feature is learned based on the responses of the training samples over the training set, so that the response coefficients are considered as the mid-level feature, which is more related to the attribute-based feature representation. Moreover, our mid-level feature learning approach and existing metric learning methods exploit discriminative information from different stages of the kinship verification task. Specially, our feature learning approach exploits discriminative information in the feature extraction stage while metric learning methods exploit such information in the similarity measure stage. Hence, our feature learning approach is expected to obtain better performance when it is combined with state-of-the-art metric learning methods to further improve the verification performance. More empirical results will be presented in Section IV to illustrate this point.

IV. EXPERIMENTS

In this section, we conduct kinship verification experiments on four benchmark kinship datasets to show the efficacy of our proposed methods. The following details the results.
face images) and 2) set 2 (200 child and 200 old parent face images). Since there are large imbalances of the different kinship relations of the UB Kinface database (nearly 80% of them are the F-S relation), we have not separated different kinship relations on this dataset.

### B. Experimental Settings

We randomly selected 4000 face images from the LFW dataset to construct the reference set, which was used for all of the four kinship face datasets to learn the mid-level feature representations. We aligned each face image in all datasets into $64 \times 64$ pixels using the provided eyes positions and converted it into gray-scale image. We applied three different feature descriptors including LBP [1], spatial pyramid learning (SPLE) [56], and SIFT [30] to extract different and complementary information from each face image. The reason we selected these three features is that they have shown reasonably good performance in recent kinship verification studies [42], [56]. We followed the same parameter settings for these features in [42] so that a fair comparison can be obtained.

For the LBP feature, 256 bins were used to describe each face image because this setting yields better performance. For the SPLE method, we first constructed a sequence of grids at three different resolutions (0, 1, and 2), such that we have 21 cells in total. Then, each local feature in each cell was quantized into 200 bins and each face image was represented by a 4200-dimensional long feature vector. For the SIFT feature, we densely sampled and computed one 128-dimensional feature over each $16 \times 207$ patch, where the overlap between two neighboring patches is 8 pixels. Then, each SIFT descriptor was concatenated into a long feature vector. For these features, we applied principal component analysis to reduce each feature into 100 dimensions to remove some noise components.

The fivefold cross-validation strategy was used in our experiments. We tuned the parameters of our PDFL and MPDFL methods on the KinFaceW-II dataset because this dataset is the largest one such that it is more effective to tune parameters on this dataset than others. We divided the KinFaceW-II dataset into fivefold with an equal size, and applied fourfold to learn the coefficient matrix and the remaining one for testing. For the training samples, we used three of them to learn our models and the other onefold to tune the parameters of our methods. In our implementations, the parameters $r$, $\lambda$, $\gamma$, and $K$ were empirically set as 5, 1, 0.5, and 500, respectively. Finally, the SVM classifier with the RBF kernel is applied for verification.

### C. Results and Analysis

#### 1) Comparisons With Existing Low-Level Feature Descriptors:
We compared our PDFL and MPDFL methods with the existing low-level feature descriptors. The difference between our methods and the existing feature representations is that we use the mid-level features rather than the original low-level features for verification. Tables II–V tabulate the verification rate of different feature descriptors on the KinFaceW-I, KinFaceW-II, Cornell KinFace, and UB KinFace kinship datasets, respectively. From these tables, we see that our proposed PDFL and MPDFL outperform the best existing methods with the lowest gain in mean verification accuracy of 2.6% and 7.1%, 6.2% and 6.8%, 1.0% and 2.4%, 0.9% and 4.6% on the KinFaceW-I, KinFaceW-II, Cornell KinFace, and UB KinFace kinship datasets, respectively.

---

**TABLE II**

<table>
<thead>
<tr>
<th>Feature</th>
<th>F-S</th>
<th>F-D</th>
<th>M-S</th>
<th>M-D</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>LBP</td>
<td>62.7</td>
<td>60.2</td>
<td>54.4</td>
<td>61.4</td>
<td>59.7</td>
</tr>
<tr>
<td>LBP+PDFL</td>
<td>65.7</td>
<td>65.5</td>
<td>60.4</td>
<td>67.4</td>
<td>64.8</td>
</tr>
<tr>
<td>LE</td>
<td>66.1</td>
<td>59.1</td>
<td>58.9</td>
<td>68.0</td>
<td>63.0</td>
</tr>
<tr>
<td>LE+PDFL</td>
<td>68.2</td>
<td>63.5</td>
<td>61.3</td>
<td>69.5</td>
<td>65.6</td>
</tr>
<tr>
<td>SIFT</td>
<td>65.5</td>
<td>63.6</td>
<td>55.5</td>
<td>65.4</td>
<td>68.9</td>
</tr>
<tr>
<td>SIFT+PDFL</td>
<td>73.5</td>
<td>76.7</td>
<td>66.1</td>
<td>73.1</td>
<td>70.1</td>
</tr>
</tbody>
</table>

**TABLE III**

<table>
<thead>
<tr>
<th>Feature</th>
<th>F-S</th>
<th>F-D</th>
<th>M-S</th>
<th>M-D</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>LBP</td>
<td>64.0</td>
<td>63.5</td>
<td>62.8</td>
<td>63.0</td>
<td>63.3</td>
</tr>
<tr>
<td>LBP+PDFL</td>
<td>69.5</td>
<td>69.8</td>
<td>70.6</td>
<td>69.8</td>
<td>69.9</td>
</tr>
<tr>
<td>LE</td>
<td>69.8</td>
<td>66.1</td>
<td>72.8</td>
<td>72.0</td>
<td>70.2</td>
</tr>
<tr>
<td>LE+PDFL</td>
<td>77.0</td>
<td>74.3</td>
<td>77.0</td>
<td>77.2</td>
<td>76.4</td>
</tr>
<tr>
<td>SIFT</td>
<td>60.0</td>
<td>56.9</td>
<td>54.8</td>
<td>55.4</td>
<td>56.8</td>
</tr>
<tr>
<td>SIFT+PDFL</td>
<td>69.0</td>
<td>62.4</td>
<td>62.4</td>
<td>62.0</td>
<td>64.0</td>
</tr>
<tr>
<td>MPDFL (All)</td>
<td>77.3</td>
<td>74.7</td>
<td>77.8</td>
<td>78.0</td>
<td>77.0</td>
</tr>
</tbody>
</table>

**TABLE IV**

<table>
<thead>
<tr>
<th>Feature</th>
<th>F-S</th>
<th>F-D</th>
<th>M-S</th>
<th>M-D</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>LBP</td>
<td>67.1</td>
<td>63.8</td>
<td>73.0</td>
<td>60.0</td>
<td>66.5</td>
</tr>
<tr>
<td>LBP+PDFL</td>
<td>67.9</td>
<td>64.2</td>
<td>77.0</td>
<td>60.8</td>
<td>67.5</td>
</tr>
<tr>
<td>LE</td>
<td>72.7</td>
<td>66.8</td>
<td>75.4</td>
<td>63.2</td>
<td>69.5</td>
</tr>
<tr>
<td>LE+PDFL</td>
<td>73.7</td>
<td>67.8</td>
<td>76.4</td>
<td>64.2</td>
<td>70.5</td>
</tr>
<tr>
<td>SIFT</td>
<td>64.5</td>
<td>67.3</td>
<td>68.4</td>
<td>61.8</td>
<td>65.5</td>
</tr>
<tr>
<td>SIFT+PDFL</td>
<td>66.5</td>
<td>69.3</td>
<td>69.4</td>
<td>62.8</td>
<td>67.0</td>
</tr>
<tr>
<td>MPDFL (All)</td>
<td>74.8</td>
<td>69.1</td>
<td>77.5</td>
<td>66.1</td>
<td>71.9</td>
</tr>
</tbody>
</table>

**TABLE V**

<table>
<thead>
<tr>
<th>Feature</th>
<th>Set 1</th>
<th>Set 2</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>LBP</td>
<td>63.4</td>
<td>61.2</td>
<td>62.3</td>
</tr>
<tr>
<td>LBP+PDFL</td>
<td>64.0</td>
<td>62.2</td>
<td>63.1</td>
</tr>
<tr>
<td>LE</td>
<td>61.9</td>
<td>61.3</td>
<td>61.6</td>
</tr>
<tr>
<td>LE+PDFL</td>
<td>62.8</td>
<td>63.5</td>
<td>63.2</td>
</tr>
<tr>
<td>SIFT</td>
<td>62.5</td>
<td>62.8</td>
<td>62.7</td>
</tr>
<tr>
<td>SIFT+PDFL</td>
<td>63.8</td>
<td>63.4</td>
<td>63.6</td>
</tr>
<tr>
<td>MPDFL (All)</td>
<td>67.5</td>
<td>67.0</td>
<td>67.3</td>
</tr>
</tbody>
</table>

**TABLE VI**

<table>
<thead>
<tr>
<th>Method</th>
<th>KinFaceW-I</th>
<th>KinFaceW-II</th>
<th>Cornell KinFace</th>
<th>UB KinFace</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method in [16]</td>
<td>N.A.</td>
<td>N.A.</td>
<td>70.7 (0.1)</td>
<td>N.A.</td>
</tr>
<tr>
<td>Method in [50]</td>
<td>N.A.</td>
<td>N.A.</td>
<td>69.5 (0.1)</td>
<td>65.6 (0.1)</td>
</tr>
<tr>
<td>NRML [42]</td>
<td>64.3 (0.1)</td>
<td>75.7 (0.1)</td>
<td>69.5 (0.1)</td>
<td>65.6 (0.1)</td>
</tr>
<tr>
<td>MNRMRL [42]</td>
<td>69.9 (0.0)</td>
<td>76.5 (0.0)</td>
<td>71.6 (0.0)</td>
<td>67.1 (0.0)</td>
</tr>
<tr>
<td>PDFL (best)</td>
<td>64.8</td>
<td>70.2</td>
<td>70.5</td>
<td>63.6</td>
</tr>
<tr>
<td>MPDFL</td>
<td>70.1</td>
<td>77.0</td>
<td>71.9</td>
<td>67.3</td>
</tr>
</tbody>
</table>
2) Comparison With State-of-the-Art Kinship Verification Methods: Table VI compares our PDFL and MPDFL methods with the state-of-the-art kinship verification methods presented in the past several years. To further investigate the performance differences between our feature learning approach and the other compared methods, we evaluate the verification results by using the null hypothesis statistical test based on Bernoulli model [4] to check whether the differences between the results of our approach and those of other methods are statistically significant. The results of the *p*-tests of PDFL and MPDFL are given in the brackets right after the verification rate of each method in each table, where the number, “1” represents significant difference and, “0” represents otherwise. There are two numbers in each bracket, where the first represents the significant difference of PDFL and the second represents that of MPDFL over previous methods. We see that PDFL achieves comparable accuracy with the existing state-of-the-art methods, and MPDFL obtains better performance than the existing kinship verification methods when the same kinship dataset was used for evaluation. Moreover, the improvement of MPDFL is significant for most comparisons.

Since our feature learning approach and previous metric learning methods exploit discriminative information in the feature extraction and similarity measure stages, respectively, we also conduct kinship verification experiments when both of them are used for our verification task. Table VII tabulates the verification performance when such discriminative information is exploited in different manners. We see that the performance of our feature learning approach can be further improved when the discriminative metric learning methods are applied.

3) Comparison With Different Classifiers: We investigated the performance of our PDFL (best single feature) and MPDFL with different classifiers. In our experiments, we evaluated two classifiers: 1) SVM and 2) nearest neighbor (NN). For the NN classifier, the cosine similarity of two face images is used. Tables VIII–IX tabulate the mean verification rate of our PDFL and MPDFL when different classifiers were used for verification. We see that our feature learning methods are not sensitive to the selection of the classifier.

4) Parameter Analysis: We took the KinFaceW-I dataset as an example to investigate the verification performance and training cost of our MPDFL versus varying values of *K* and *γ*. Figs. 3 and 4 show the mean verification accuracy and the training time of MPDFL versus different *K* and *γ*. We see that *K* and *γ* were set to 500 and 0.5 are good tradeoffs between the efficiency and effectiveness of our proposed method.

Fig. 5 shows the mean verification rates of PDFL and MPDFL versus different number of iteration on the KinFaceW-I dataset. We see that PDFL and MPDFL achieve stable verification performance in several iterations.

We investigated the effect of the parameter *r* in MPDFL. Fig. 6 shows the verification rate of MPDFL versus different number of *r* on different kinship datasets. We observe that our MPDFL method is in general robust to the parameter *r*, and the best verification performance can be obtained when *r* was set to 5.
Fig. 5. Mean verification rate of our PDFL and MPDFL versus different number of iterations, on the KinFaceW-I dataset.

Fig. 6. Mean verification rate of our MPDFL versus different values of \( r \) on different kinship face datasets.

### Table X

<table>
<thead>
<tr>
<th>Method</th>
<th>KinFaceW-I</th>
<th>KinFaceW-II</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F-S</td>
<td>F-D</td>
</tr>
<tr>
<td>LBP+PDFL</td>
<td>67.0</td>
<td>71.0</td>
</tr>
<tr>
<td>LE+PDFL</td>
<td>68.0</td>
<td>73.0</td>
</tr>
<tr>
<td>SIFT+PDFL</td>
<td>70.0</td>
<td>75.0</td>
</tr>
<tr>
<td>MPDFL</td>
<td>75.0</td>
<td>80.0</td>
</tr>
</tbody>
</table>

#### 5) Computational Time

We compare the computational time of the proposed PDFL and MPDFL methods with state-of-the-art metric learning based kinship verification methods including neighborhood repulsed metric learning (NRML) and multiview neighborhood repulsed metric learning (MNRML). Our hardware consists of a 2.4-GHz CPU and a 6 GB RAM. Table X shows the time spent on the training and the testing stages of different methods, where the MATLAB software, the KinFaceW-I database and the SVM classifier were used. We see that the computational time of our feature learning methods are comparable to those of NRML and MNRML.

#### 6) Comparison With Human Observers in Kinship Verification

Human ability in kinship verification was evaluated in [42]. We also compared our method with humans on the KinFaceW-I and KinFaceW-II datasets. For a comparison between human ability and our proposed approach, the training samples as well as their kin labels used in our approach were selected and presented to ten human observers (five males and females) who are 20–30 years old [42] to provide the prior knowledge to learn the kin relation from human face images. Then, the testing samples used in our experiments were presented to these to evaluate the performance of human ability in kinship verification. There are two evaluations for humans, HumanA and HumanB in [42], where the only face region and the whole original face were presented to human observers, respectively. Table XI shows the performance of these observers and our approach. We see that our proposed methods achieve even better kinship verification performance than human observers on most subsets of these two kinship datasets.

### D. Discussion

We make the following four observations from the above experimental results listed in Tables II–XI, and Figs. 3–6.

1) Learning discriminative mid-level feature achieves better verification performance than the original low-level feature. This is because the learned mid-level feature exploits discriminative information while the original low-level feature cannot.

2) MPDFL achieves better performance than LDFL, which indicates that combining multiple local-level descriptors to learn mid-level features is better than using a single one because multiple features can provide complementary information for feature learning.

3) PDFL achieves comparable performance and achieves better performance than existing kinship verification methods. The reason is that most existing kinship verification methods used low-level hand-crafted features for face representation, which is not discriminative enough to characterize the kin relation of face images.

4) Both PDFL and MPDFL achieve better kinship verification performance than human observers, which further shows the potentials of our computational face based kinship verification models for practical applications.

### V. Conclusion

In this paper, we presented two discriminative mid-level feature learning methods called PDFL and MPDFL for kinship verification via facial images. Experimental results on four publicly available face kinship datasets have shown that our proposed methods consistently outperform both...
the existing face descriptors and state-of-the-art kinship verification methods.

For future work, we are interested in applying the proposed feature learning approach to other computer vision applications such as object recognition and visual tracking to further show its effectiveness.

REFERENCES


Haibin Yan received the B.Eng. and M.Eng. degrees from the Xi’an University of Technology, Xi’an, China, and the Ph.D. degree from the National University of Singapore, Singapore, in 2004, 2007, and 2013, all in mechanical engineering. She is currently a Research Fellow with the Department of Mechanical Engineering, National University of Singapore, Singapore. Her current research interests include human-robotic interaction, social robotics, industrial robot, and computer vision.

Jiwen Lu (S’10–M’11) received the B.Eng. degree in mechanical engineering and the M.Eng. degree in electrical engineering from the Xi’an University of Technology, Xi’an, China, and the Ph.D. degree in electrical engineering from Nanyang Technological University, Singapore, in 2003, 2006, and 2011, respectively. He is currently a Research Scientist with the Advanced Digital Sciences Center, Singapore. His current research interests include computer vision, pattern recognition, machine learning, and biometrics. He has authored/co-authored over 90 scientific papers in the above areas, where over 20 papers were published in the IEEE transactions journals and top-tier computer vision conferences such as IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), IEEE Transactions on Image Processing (TIP), IEEE Transactions on Information Forensics and Security (TIFS), IEEE Transactions on Circuits and Systems for Video Technology (TCSVT), IEEE Transactions on Systems, Man, and Cybernetics (TSMC), IEEE International Conference on Computer Vision (ICCV), IEEE Conference on Computer Vision and Pattern Recognition (CVPR), and European Conference on Computer Vision (ECCV).

Dr. Lu was the recipient of the First-Prize National Scholarship and the National Outstanding Student Award from the Ministry of Education of China in 2002 and 2003, the Best Student Paper Award from the PREMIA of Singapore in 2012, the Top 10% Best Paper Award from IEEE International Workshop on Multimedia Signal Processing 2014, and the Best Paper Award Nominations from the IEEE International Conference on Multimedia and Expo (ICME) 2011 and 2013, respectively. He served as a TPC/Reviewer for over 20 international conferences such as ICCV, CVPR, ECCV, International Conference on Pattern Recognition, International Joint Conference on Biometrics (IJC), IEEE International Conference on Acoustics, Speech, and Signal Processing, and ICME, and a Reviewer for over 40 international journals such as TPAMI, TIP, TIFS, TCSVT, IEEE Transactions on Multimedia, IEEE Transactions on Neural Networks and Learning Systems, TSMC, and IEEE Transactions on Geoscience and Remote Sensing. He organizes several workshops/competitions at some international conferences such as ICME 2014, Asian Conference on Computer Vision (ACCV) 2014, IJCB 2014, and IEEE International Conference on Automatic Face and Gesture Recognition 2015. He gave tutorials at ACCV 2014, ICME 2014, and IJCB 2014. He serves as an Area Chair for the ICME 2015, and the 2015 IAPR/IEEE International Conference on Biometrics.

Xiuzhuang Zhou (M’11) received the B.Sci. degree in atmosphere physics from the Chengdu University of Information Technology, Chengdu, China, and the M.Eng. and Ph.D. degrees from the School of Computer Science, Beijing Institute of Technology, Beijing, China, in 1996, 2005, and 2011, respectively. He is currently an Associate Professor with the College of Information Engineering, Capital Normal University, Beijing. His current research interests include computer vision, pattern recognition, and machine learning.