Neighborhood Repulsed Metric Learning for Kinship Verification

Jiwen Lu, *Member, IEEE*, Xiuzhuang Zhou, *Member, IEEE*, Yap-Pen Tan, *Senior Member, IEEE*, Yuanyuan Shang, *Member, IEEE*, and Jie Zhou, *Senior Member, IEEE*

Abstract—Kinship verification from facial images is an interesting and challenging problem in computer vision, and there are very limited attempts on tackle this problem in the literature. In this paper, we propose a new neighborhood repulsed metric learning (NRML) method for kinship verification. Motivated by the fact that interclass samples (without a kinship relation) with higher similarity usually lie in a neighborhood and are more easily misclassified than those with lower similarity, we aim to learn a distance metric under which the intraclass samples (with a kinship relation) are pulled as close as possible and interclass samples lying in a neighborhood are repulsed and pushed away as far as possible, simultaneously, such that more discriminative information can be exploited for verification. To make better use of multiple feature descriptors to extract complementary information, we further propose a multiview NRML (MNRML) method to seek a common distance metric to perform multiple feature fusion to improve the kinship verification performance. Experimental results are presented to demonstrate the efficacy of our proposed methods. Finally, we also test human ability in kinship verification from facial images and our experimental results show that our methods are comparable to that of human observers.

Index Terms—Face and gesture recognition, kinship verification, metric learning, multiview learning, biometrics

✦

1 INTRODUCTION

FACIAL images convey many important human characteristics, such as identity, gender, expression, age, ethnicity and so on. Over the past two decades, a large number of face analysis problems have been investigated in the computer vision and pattern recognition community. Representative examples include face recognition [5], [8], [10], [20], [22], [27], [28], [35], [37], [40], [41], [42], [43], [54], [55], [56], [58], [64], [65], [66], [67], [68], facial expression recognition [11], [18], [70], facial age estimation [19], [21], [24], [25], [31], [39], gender classification [44], [45], and ethnicity recognition [26], [47].

In this paper, we investigate a new face analysis problem: kinship verification from facial images. To the best of our knowledge, there are very limited attempts on tackle this problem in the literature. Given each pair of face images, our objective is to determine whether there is a kinship relation between these two people. We define kinship as a relationship between two people who are biologically related with overlapping genes. Specifically,

we examine in this paper four different types of kinship relations: father-son (F-S), father-daughter (F-D), mother-son (M-S), and mother-daughter (M-D) kinship relations. This new research topic has several potential applications such as family album organization, image annotation, social media analysis, and missing children/parents search. However, limited research has been conducted along this direction, possibly due to lacking of such publicly available kinship databases and inherent challenges of this problem. To this end, we construct two new kinship databases named KinFaceW-I and KinFaceW-II[1] from Internet search under uncontrolled conditions. Then, we learn a robust distance metric under which facial images with kinship relations are projected as close as possible and those without kinship relations are pushed away as far as possible, simultaneously. Since interclass samples (without a kinship relation) with higher similarity usually lie in a neighborhood and are more easily misclassified than those with lower similarity, we emphasize the interclass samples in a neighborhood more in learning the distance metric and expect those samples lying in a neighborhood are repulsed and pushed away as far as possible, simultaneously, such that more discriminative information can be exploited for verification. Inspired by the fact that multiple feature descriptors could provide complementary information in characterizing facial information from different viewpoints to extract more discriminative information, we propose a multiview neighborhood repulsed metric learning (MNRML) method by learning a common distance metric, under which multiple feature descriptors can be effectively and well combined to further improve the verification performance.

- J. Lu is with the Advanced Digital Sciences Center, 1 Fusionopolis Way, #08-10, Connexis North Tower, Singapore 138632. E-mail: jiwen.lu@adsc.com.sg.
- X. Zhou and Y. Shang are with the College of Information Engineering, Capital Normal University, No. 56, West Ring 3rd Road North, Haidian District, Beijing 100048, China. E-mail: zxz@xeehoo.com, syy@bao.ac.cn.
- Y.-P. Tan is with the School of Electrical and Electronics Engineering, Nanyang Technological University, Block S1, Nanyang Avenue, Singapore 639798. E-mail: eyptan@ntu.edu.sg.
- J. Zhou is with the Department of Automation, Tsinghua University, Room 404, Main Building, Beijing 100084, China. E-mail: jzhou@tsinghua.edu.cn.

1. The difference of KinFaceW-I and KinFaceW-II is that each pair of kinship facial images in KinFaceW-I was acquired from different photos and that in KinFaceW-II was extracted from the same photo. Some face examples with different relations will be provided in Section 3.
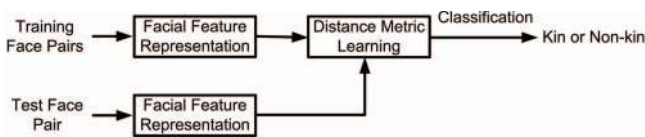
Fig. 1. Framework of our proposed kinship verification approach via facial image analysis. Given a set of training face images, we first extract features for each face image and learn a distance metric to map these feature representations into a low-dimensional feature subspace, under which the kinship relation of face samples can be better discriminated. For each test face pair, we also extract features of each face image and map these features to the learned low-dimensional feature subspace. Finally, a classifier is used to verify whether there is a kinship relationship or not between the test face pair.

Experimental results are presented to demonstrate the feasibility of verifying human kinship via facial image analysis and the efficacy of our proposed methods. Fig. 1 illustrates the basic framework of our proposed approach.

The contributions of this paper are summarized as follows:

- We have proposed a novel NRML method for kinship verification. The proposed metric can pull intraclass samples (with a kinship relation) to be as close as possible and push interclass samples (without a kinship relation) in a neighborhood to be as far as possible, simultaneously, such that more discriminative information can be exploited for verification.
- We have proposed a new multiview NRML (MNRML) method to seek a common distance metric to perform multiple feature fusion by making better use of multiple face feature descriptors to extract complementary information to improve the kinship verification performance.
- We have constructed two new kinship databases, named KinFaceW-I and KinFaceW-II, from the Internet search, where face images were captured under uncontrolled conditions. To the best of our knowledge, they are the largest face data sets for kinship verification for practical applications because facial images in our data sets were collected from real-world environments. We have made these two databases and labels publicly available online.[2]
- We have conducted a number of kinship verification experiments to demonstrate the efficacy of our proposed methods. Moreover, we have also tested human ability in kinship verification from facial images and our experimental results show that our methods are comparable to that of human observers.

This is an extended version of our conference paper [36]. The following describes the extensions in this paper from its conference version:

- We have investigated the performance of our proposed NRML and MNRML methods versus different classifiers. Experimental results have shown that the verification accuracies of our proposed methods are not sensitive to the selection of the classifier.
- We have conducted experiments on two additional, publicly available kinship databases and compared

the performance of our proposed methods with that of their algorithms. Experimental results have shown that our proposed methods significantly outperform their methods in terms of the correct verification rate, which further shows the efficacy of our proposed methods.

- We have included more experimental results in this paper such as the ROC curve comparisons, computational costs of different methods, and parameter analysis of NRML and MNRML, to further show the efficacy of our proposed methods.
- We have compared our MNRML method with several existing multiview learning methods in our kinship verification experiments. Experimental results show that our MNRML can achieve comparable performance with most existing multiview learning methods.
- We have conducted age-invariant face verification experiments on the FG-NET and MORPH face data sets to further demonstrate the effectiveness of our proposed methods.

The remainder of this paper is organized as follows: Section 2 briefly reviews related work. Section 3 presents our kinship data sets. Section 4 details the proposed methods. Section 5 provides the experimental results, and Section 6 concludes the paper.

2   RELATED WORK

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

2.1   Kinship Verification

Recently, human perception of kinship verification has been investigated in the psychology community [2], [13], [14], [16], [29], [30], and one key finding is observed: Humans have the capability to recognize kinship based on face images even if these images are from unfamiliar faces. Motivated by this finding, researchers in computer vision are interested in developing computational approaches to verify human kinship relations, and there have been a limited number of attempts to address this problem in recent years. To our best knowledge, Fang et al. [17] made the first attempt to tackle the challenge of kinship verification from facial images by using local facial feature extraction and selection. They first localized several face parts and extracted some features such as skin color, gray value, histogram of gradient, and facial structure information to describe facial images. Then, the k-nearest-neighbor (KNN) and support vector machine (SVM) classifiers were applied to classify face images. More recently, Xia et al. [60], [61] proposed a transfer subspace learning approach for kinship verification. Their basic idea is to utilize an intermediate young parent facial image set to reduce the divergence between the children and old parent images based on the assumption that the children and young parents possess more facial resemblance in facial appearances.

While encouraging results were obtained, there are still two shortcomings among the existing kinship verification works: 1) The conventional euclidean metric was usually utilized for kinship verification and such metric is not

2. Available at: https://sites.google.com/site/elujiwen/kinfacew.

appropriate to measure the similarity of facial images because the intrinsic space that face usually lies in is a low-dimensional manifold rather than a euclidean space; 2) existing methods were only evaluated on relatively small data sets (150 and 200 pairs in [17] and [61], respectively), which may not be sufficient to demonstrate the effectiveness of face analysis-based kinship verification. Hence, more robust and effective distance metrics and larger kinship data sets are desirable to be adopted to demonstrate and improve the performance of existing kinship verification methods.

2.2 Metric Learning

Metric learning has received a lot of attention in computer vision and machine learning in recent years, and there have been a number of metric learning algorithms in the literature. Existing metric learning methods can be mainly divided into two categories: unsupervised and supervised. Unsupervised methods aim to learn a low-dimensional manifold where the geometrical information of the samples is preserved. Representatives of such algorithms include principal component analysis (PCA) [54], locally linear embedding (LLE) [50], multidimensional scaling (MDS) [52], and Laplacian eigenmaps (LE) [6]. Supervised methods aim to seek an appropriate distance metric for classification tasks. Generally, an optimization objective function is formulated based on some supervised information of the training samples to learn the distance metric. The difference among these methods lies mainly in the objective functions, which are designed for their specific tasks. Typical supervised metric learning methods include linear discriminant analysis (LDA) [5], neighborhood component analysis (NCA) [23], cosine similarity metric learning (CSML) [46], large margin nearest neighbor (LMNN) [57], and information theoretic metric learning (ITML) [15].

While metric learning methods have achieved reasonably good performance in many visual analysis applications, there are still two shortcomings among most existing methods: 1) Some training samples are more informative in learning the distance metric than others, and most existing metric learning methods consider them equally and ignore potentially different contributions of the samples to learn the distance metric; 2) most existing metric learning methods only learn a distance metric from single view data and cannot handle multiview data directly. Previous research has shown that different feature descriptors could provide complementary information in characterizing facial information from different viewpoints [62], [63], and hence it is desirable to utilize multiple feature information for our kinship verification task. However, multiple feature descriptors generally have multiple modalities and existing metric learning methods cannot work well for such multiview data directly. To address this shortcoming, we propose a new multiview metric learning method to learn a common and robust metric to measure the similarity of facial images represented by multiple feature descriptors, simultaneously.

3 DATA SETS

There are many potential applications for kinship verification. One representative example is family album organization and missing parent/child search, where we

Fig. 2. Several image examples of our KinFaceW-I database. From top to bottom are the father-son (F-S), father-daughter (F-D), mother-son (M-S), and mother-daughter (M-D) kinship relations, and the neighboring two images in each row are with kinship relation, respectively.

usually need to determine the kinship relation from two face photos taken at different times. Another important application is social media analysis, such as understanding the relationships of people in a photo. For this application, there are usually many face images in a photo and we need to determine the kinship relation from two face photos in the same photo.

To advance the kinship verification research for different practical applications and show the efficacy of our proposed methods, we collected two kinship face data sets, named KinFaceW-I and KinFaceW-II, from the Internet through an online search, where some public figure face images, as well as their parents' or children's face images. The difference of KinFaceW-I and KinFaceW-II is that each pair of kinship facial images in KinFaceW-I was acquired from different photos and that in KinFaceW-II was obtained from the same photo. Hence, experimental results on the KinFaceW-I and KinFaceW-II data sets show different potentials in the real applications. We impose no restriction in terms of pose, lighting, background, expression, age, ethnicity, and partial occlusion on the images used for training and testing. Some examples from these two data sets are shown in Figs. 2 and 3, respectively.[3]

There are four kinship relations in both the KinFaceW-I and KinFaceW-II data sets: father-son (F-S), father-daughter (F-D), mother-son (M-S), and mother-daughter (M-D). In the KinFaceW-I data set, there are 156, 134, 116, and 127 pairs of kinship images for these four relations. For the KinFaceW-II data set, each relationship contains 250 pairs of kinship images. Fig. 4 shows the ethnicity distributions of these two data sets.

4 PROPOSED METHODS

4.1 Basic Idea

Fig. 5 shows the basic idea of our proposed NRML method. There are two sample sets in Fig. 5a, where samples in the left (blue color) denote face images of parents, and those in the right (red color) denote face images of children, respectively. These samples are denoted by circles, squares and triangles, respectively. Let the two circles in this figure

Fig. 3. Several image examples of our KinFaceW-II database. From top to bottom are the father-son (F-S), father-daughter (F-D), mother-son (M-S), and mother-daughter (M-D) kinship relations, and the neighboring two images in each row are with kinship relation, respectively.

denote a pair of face samples with a kinship relation, where the blue circle and red circle represent face images of the parent and child, respectively. In the original face image space, there is usually a large difference between the parent and child images in the circle class due to some variation factors such as aging, illumination, pose and expression. Hence, there are some other parent and children images lying in the neighborhood of the parent and child images in the circle class, respectively, as shown in Fig. 5a. From the classification viewpoint, neighboring samples in the parent and child image sets are more easily misclassified than those are not in the neighborhood because there is a high chance to misclassify the images in the neighborhood. To address this challenge, we aim to learn a distance metric under which facial images with kinship relations are projected as close as possible and those without kinship relations in the neighborhoods are pulled as far as possible, as shown in Fig. 5b. In other words, the similarities of the (blue) triangle samples and the (red) circle and that of the (blue) circle and the (red) triangles should be decreased such that the kinship margin in the learned distance metric is much larger and more discriminative information can be exploited for kinship verification.

4.2 NRML

Let $S = \{(x_i, y_i) | i = 1, 2, \ldots, N\}$ be the training set of N pairs of kinship images, where x_i and y_i are two m-dimensional column vectors, of the ith parent and child image pair, respectively. The aim of NRML is to seek a good metric d such that the distance between x_i and y_j ($i = j$) is as small as
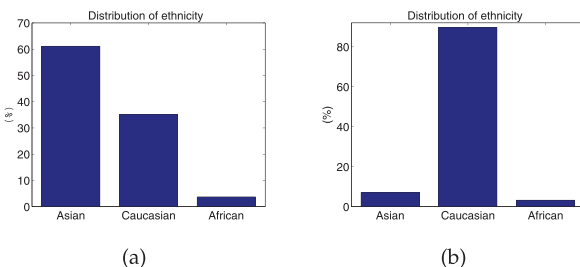

Fig. 4. The ethnicity distribution of the (a) KinFaceW-I and (b) KinFaceW-II data sets, respectively.
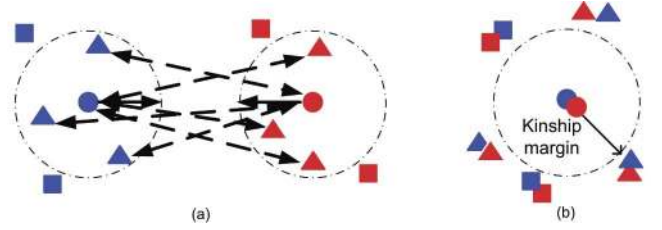

Fig. 5. Intuitive illustration of our proposed NRML method. (a) The original face images with/without kinship relations in the high-dimensional feature space. The samples in the left denote face images of parents, and those in the right denote face images of children, respectively. Given one pair of face images with kinship relation (denoted as circles), the triangles and squares denote face samples in the neighborhood and non-neighborhood, respectively. We aim to learn a distance metric such that facial images with kinship relations are projected as close as possible and those without kinship relations in the neighborhoods are pushed away as far as possible. (b) The expected distributions of face images in the learned metric space, where the similarity of the circle pair (with a kinship relation) is increased and those of the circle and triangle pairs are decreased, respectively.

possible, and that between x_i and $y_j (i \neq j)$ is as large as possible, simultaneously, where

$$d(x_i, y_j) = \sqrt{(x_i - y_j)^T A(x_i - y_j)}, \qquad (1)$$

where A is an $m \times m$ square matrix, and $1 \leq i, j \leq N$. Since d is a metric, $d(x_i, y_j)$ should have the properties of nonnegativity, symmetry, and triangle inequality. Hence, A is symmetric and positive semidefinite.

As discussed above, we formulate the proposed NRML as the following optimization problem:

$$
\begin{aligned}
\max_A J(A) &= J_1(A) + J_2(A) - J_3(A) \\
&= \frac{1}{Nk} \sum_{i=1}^{N} \sum_{t_1=1}^{k} d^2(x_i, y_{it_1}) + \frac{1}{Nk} \sum_{i=1}^{N} \sum_{t_2=1}^{k} d^2(x_{it_2}, y_i) \\
&\quad - \frac{1}{N} \sum_{i=1}^{N} d^2(x_i, y_i) \\
&= \frac{1}{Nk} \sum_{i=1}^{N} \sum_{t_1=1}^{k} (x_i - y_{it_1})^T A(x_i - y_{it_1}) \\
&\quad + \frac{1}{Nk} \sum_{i=1}^{N} \sum_{t_2=1}^{k} (x_{it_2} - y_i)^T A(x_{it_2} - y_i) \\
&\quad - \frac{1}{N} \sum_{i=1}^{N} (x_i - y_i)^T A(x_i - y_i),
\end{aligned} \qquad (2)
$$

where y_{it_1} represents the t_1th k-nearest neighbor of y_i and x_{it_2} denotes the t_2th k-nearest neighbor of x_i, respectively, the metric d is defined as (1). The aim of J_1 in (2) is to ensure that if y_{it_1} and y_i are close, then they should be separated as far as possible with x_i in the learned distance metric space. Similarly, the objective of J_2 in (2) is to ensure that if x_{it_2} and x_i are close, they should be separated as far as possible with y_i in the learned distance metric space. On the other hand, J_3 in (2) ensures that x_i and y_i are pulled as close as possible in the learned distance metric space because they have kinship relations.

It can be seen that the optimization criterion in (2) poses a chicken-and-egg problem because the distance metric d needs to be known for computing the k-nearest neighbors of

x_i and y_i. To our best knowledge, there is no closed-form solution for such an optimization problem. Alternatively, we solve this problem in an iterative manner. The basic idea is to first use the euclidean metric to search the k-nearest neighbors of x_i and y_i, and solve d sequentially.

Since A is symmetric and positive semidefinite, we can seek a nonsquare matrix W of size $m \times l$, where $l \leq m$, such that

$$A = WW^T. \tag{3}$$

Then, (1) can be rewritten as

$$
\begin{aligned}
d(x_i, y_j) &= \sqrt{(x_i - y_j)^T A(x_i - y_j)} \\
&= \sqrt{(x_i - y_j)^T WW^T(x_i - y_j)} \\
&= \sqrt{(u_i - v_j)^T(u_i - v_j)},
\end{aligned} \tag{4}
$$

where $u_i = W^T x_i$ and $v_j = W^T y_j$.

Combining (2) and (4), we simplify $J_1(A)$ to the following form:

$$
\begin{aligned}
J_1(A) &= \frac{1}{Nk} \sum_{i=1}^{N} \sum_{t_1=1}^{k} (x_i - y_{it_1})^T WW^T(x_i - y_{it_1}) \\
&= tr\left(W^T \frac{1}{Nk} \sum_{i=1}^{N} \sum_{t_1=1}^{k} (x_i - y_{it_1})(x_i - y_{it_1})^T W \right) \\
&= tr(W^T H_1 W),
\end{aligned} \tag{5}
$$

where $H_1 \triangleq \frac{1}{Nk} \sum_{i=1}^{N} \sum_{t_1=1}^{k} (x_i - y_{it_1})(x_i - y_{it_1})^T$.

Similarly, $J_2(A)$ and $J_3(A)$ can be simplified as

$$
\begin{aligned}
J_2(A) &= tr\left(W^T \frac{1}{Nk} \sum_{i=1}^{N} \sum_{t_2=1}^{k} (x_{it_2} - y_i)(x_{it_2} - y_i)^T W \right) \\
&= tr(W^T H_2 W),
\end{aligned} \tag{6}
$$

$$
\begin{aligned}
J_3(A) &= tr\left(W^T \frac{1}{N} \sum_{i=1}^{N} (x_i - y_i)(x_i - y_i)^T W \right) \\
&= tr(W^T H_3 W),
\end{aligned} \tag{7}
$$

where $H_2 \triangleq \frac{1}{Nk} \sum_{i=1}^{N} \sum_{t_2=1}^{k} (x_{it_2} - y_i)(x_{it_2} - y_i)^T$ and $H_3 \triangleq \frac{1}{N} \sum_{i=1}^{N} (x_i - y_i)(x_i - y_i)^T$.

Now, we can formulate our NRML method as

$$
\begin{aligned}
\max_{W} J(W) &= tr[W^T(H_1 + H_2 - H_3)W] \\
\text{subject to} \quad & W^T W = I,
\end{aligned} \tag{8}
$$

where $W^T W = I$ is a constraint to restrict the scale of W such that the optimization problem with respect to W is well posed. Then, W can be obtained by solving the following eigenvalue problem

$$(H_1 + H_2 - H_3)w = \lambda w. \tag{9}$$

Let $w_1, w_2, \ldots, w_l$ be the eigenvectors of (9) corresponding to the l largest eigenvalues ordered according to $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_l$. An $m \times l$ transformation matrix $W = [w_1, w_2, \ldots, w_l]$ can be obtained to project the original face samples x_i and y_i into low-dimensional feature vectors u_i and v_i, as follows:

$$u_i = W^T x_i, v_i = W^T y_i, \quad i = 1, 2, \ldots, N. \tag{10}$$

Having obtained W, we can recalculate the k-nearest neighbors of x_i and y_i by using (1), respectively, and update W by re-solving the eigenvalue equation in (9). The proposed NRML algorithm is summarized in Algorithm 1.

Algorithm 1. NRML.

Input: Training images: $S = \{(x_i, y_i)|i = 1, 2, \ldots, N\}$,
Parameters: neighborhood size k, iteration number T, and convergence error ε (set as 0.0001).

Output: Distance metric W.

Step 1 (Initialization):
Search the k-nearest neighbors for each x_i and y_i by using the conventional euclidean metric.

Step 2 (Local optimization):
For $r = 1, 2, \ldots, T$, repeat
2.1. Compute H_1, H_2 and H_3, respectively.
2.2. Solve the eigenvalue problem in (9).
2.3. Obtain $W^r = [w_1, w_2, \ldots, w_l]$.
2.4. Update the k-nearest neighbors of x_i and y_i by W^r.
2.5. If $r > 2$ and $|W^r - W^{r-1}| < \varepsilon$, go to Step 3.

Step 3 (Output distance metric):
Output distance metric $W = W^r$.

4.3 MNRML

Previous studies have shown that different feature descriptors could provide complementary information in characterizing facial information from different viewpoints [7], [12], [38], [51], [69], [71], [73], and hence it is desirable to utilize multiple feature information for our kinship verification task. However, multiple feature descriptors generally have multiple modalities and existing metric learning methods cannot be used for multiview data directly. A nature solution for multiview metric learning is to directly concatenate different features together as a new vector and then apply existing metric learning methods for feature extraction. However, such operation is not physically meaningful because different features have different statistical characteristics and such concatenation ignores the diversity of different feature representations, which cannot efficiently exploit the complementary information of different features. To address this problem, we propose a new multiview NRML (MNRML) method to learn a common distance metric for measuring multiple feature representations of facial images for kinship verification.

Assume there are K views of feature representations, and $S^p = \{(x_i^p, y_i^p)|i = 1, 2, \ldots, N\}$ is the feature representation of the pth view set of N pairs of kinship images, where $x_i^p \in R^m$ and $y_i^p \in R^m$ are the ith parent and child images from the pth view, respectively, where $p = 1, 2, \ldots, K$. The aim of MNRML is to seek a common metric d such that the distance between x_i^p and y_j^p ($i = j$) is as small as possible, and that between x_i^p and y_j^p ($i \neq j$) is as large as possible, simultaneously.

To discover the complemental information of facial images from different views, we impose a nonnegative weighted vector $\beta = [\beta_1, \beta_2, \ldots, \beta_K]$ on the objective function of NRML of each view. Generally, the larger β_p is, the more contribution of the feature representations from the

pth view is made to learn the distance metric. Hence, we formulate MNRML as the following optimization problem:

$$\max_{W,\beta} \sum_{p=1}^{K} \beta_p tr\left[W^T\left(H_1^p + H_2^p - H_3^p\right)W\right]$$

(11)

$$\text{subject to} \quad W^T W = I, \sum_{p=1}^{K} \beta_p = 1, \beta_p \geq 0.$$

The solution to (11) is $\beta_p = 1$ corresponding to the maximal $tr[W^T(H_1^p + H_2^p - H_3^p)W]$ over different views, and $\beta_p = 0$ otherwise, which means only the best view is selected by our method, such that the complementary information of facial features from different views has not been exploited. To address this, we modify β_p to be β_p^q, where $q > 1$, and the new objective function is defined as

$$\max_{W,\beta} \sum_{p=1}^{K} \beta_p^q tr\left[W^T\left(H_1^p + H_2^p - H_3^p\right)W\right]$$

(12)

$$\text{subject to} \quad W^T W = I, \sum_{p=1}^{K} \beta_p = 1, \beta_p \geq 0.$$

To the best of our knowledge, there is no closed-form solution to (12) since it is nonlinearly constrained nonconvex optimization problem. Similarly to NRML, we also solve it in an iterative manner.

First, we fix W and update β. We construct a Lagrange function

$$L(\beta,\zeta) = \sum_{p=1}^{K} \beta_p^q tr\left[W^T\left(H_1^p + H_2^p - H_3^p\right)W\right]$$
$$- \zeta\left(\sum_{p=1}^{K} \beta_p - 1\right).$$

(13)

Letting $\frac{\partial L(\beta,\zeta)}{\partial \beta_p} = 0$ and $\frac{\partial L(\beta,\zeta)}{\partial \zeta} = 0$, we have

$$q\beta_p^{q-1} tr\left[W^T\left(H_1^p + H_2^p - H_3^p\right)W\right] - \zeta = 0,$$

(14)

$$\sum_{p=1}^{K} \beta_p - 1 = 0.$$

(15)

Combining (14) and (15), we can obtain β_p as follows:

$$\beta_p = \frac{\left(1/tr\left[W^T\left(H_1^p + H_2^p - H_3^p\right)W\right]\right)^{1/(q-1)}}{\sum_{p=1}^{K}\left(1/tr\left[W^T\left(H_1^p + H_2^p - H_3^p\right)W\right]\right)^{1/(q-1)}}.$$

(16)

Then, we update W by using the new β. When β is fixed, (12) is equivalent to

$$\max_{W} tr\left[W^T\left(\sum_{p=1}^{K} \beta_p^q\left(H_1^p + H_2^p - H_3^p\right)\right)W\right]$$

(17)

$$\text{subject to} \quad W^T W = I.$$

And W can be obtained by solving the following eigenvalue equation:

$$\left(\sum_{p=1}^{K} \beta_p^q\left(H_1^p + H_2^p - H_3^p\right)\right)w = \lambda w.$$

(18)

Fig. 6. Aligned and cropped image examples of our KinFaceW-I database. From top to bottom are the father-son (F-S), father-daughter (F-D), mother-son (M-S), and mother-daughter (M-D) kinship relations, and the neighboring two images in each row are with the kinship relation, respectively.

The proposed MNRML algorithm is summarized in Algorithm 2.

Algorithm 2. MNRML.
  Input: Training images: $S^p = \{(x_i^p, y_i^p)|i = 1, 2, \ldots, N\}$
        be the pth view set of N pairs of kinship images,
        Parameters: neighborhood size k, iteration number
        T, tuning parameter q, and convergence error ε
        (set as 0.0001).
  Output: Distance metric W.
  Step 1 (Initialization):
      1.1. Set $\beta = [1/K, 1/K, \ldots, 1/K]$;
      1.2. Obtain W^0 by solving (18).
  Step 2 (Local optimization):
      For $r = 1, 2, \ldots, T$, repeat
      2.1. Compute β by using (16).
      2.2. Obtain W^r by solving (18).
      2.3. If $r > 2$ and $|W^r - W^{r-1}| < \varepsilon$, go to Step 3.
  Step 3 (Output distance metric):
      Output distance metric $W = W^r$.

5 EXPERIMENTS

We have developed an experimental protocol for kinship verification on our data sets and provided a benchmark baseline for other researchers to compare their methods and algorithms with our baseline results. Moveover, we have also evaluated our proposed NRML and MNRML methods by conducting a number of kinship verification experiments on our data sets. The following details experimental settings and results.

5.1 Experimental Settings

5.1.1 Data Preparation

In our experiments, we manually label the coordinates of the eyes position of each face image, and cropped and aligned facial region into 64×64 according to the template used in [53], such that the nonfacial regions such as the background and hairs were removed and only facial region was used for kinship verification. If those color images, we converted them into grayscale images. For each cropped image, histogram equalization was applied to mitigate the illumination issue. Figs. 6 and 7 show some cropped and aligned images of our KinFaceW-I and KinFaceW-II data sets, respectively.

Fig. 7. Aligned and cropped image examples of our KinFaceW-II database. From top to bottom are the father-son (F-S), father-daughter (F-D), mother-son (M-S), and mother-daughter (M-D) kinship relations, and the neighboring two images in each row are with the kinship relation, respectively.

5.1.2 Feature Representation

We have experimented with the following four feature descriptors for our kinship verification task:

- LBP [1]: Each face image was divided into 4×4 nonoverlapped blocks and the size of each block is 16×16. For each block, we extracted a 256D histogram feature rather than the 59D uniform pattern to describe each image block because we found such parameter setting achieved better performance than that used in [1]. Therefore, there are 16 256D features and these features were concatenated into a 4,096-dimensional vector for feature representation.

- LE [9]: For each pixel in the face image, its neighboring pixels in a ring-based pattern were sampled to generate a low-level feature vector. Specifically, we sampled $r * 8$ pixels at equal intervals on the ring of radius r, and r was empirically set as 0, 1, and 2, respectively, such that $25 \, (1 + 8 + 16)$ neighboring pixels were sampled to construct a feature vector for each pixel. Then, we normalized the sampled feature vector into unit length such that it can be more robust to illumination-invariant. For all these feature vectors in the training set, we performed K-means to quantize them into M discrete types, and each face image will be encoded as an M-dimensional feature vector. To make better use of the spatial information, we applied a spatial pyramid approach for feature representation [72]. First, we constructed a sequence of grids at resolution $0, \ldots, l, \ldots, L$, such that the grid at level l has 2^l cells along each dimension, where $0 \le l \le L$. Then, we extracted the LE feature in each cell and concatenated them into a long feature vector. In our experiments, M and L are empirically set to be 200 and 2, respectively, such that the final LE feature vector is 4,200D.

- SIFT [34]: For each face image, we divided each face image into several overlapping patches and extracted local features from each patch. In our experiments, the patch size was set as 16×16 and the overlapping radius is 8. Hence, there are 49 blocks for each face image and the whole image was represented by a 6,272D feature vector.

- Three-patch LBP (TPLBP) [59]: Similarly to LBP, each face image was also divided into 4×4 nonoverlapped

blocks and the size of each block is 16×16. For each pixel in each block, we considered a 3×3 patch centered on the pixel and eight additional patches distributed uniformly in a ring of radius around it. In our implementation, the radius was set as 2. For each block, we extracted a 256D histogram feature and concatenated features from all blocks into a 4,096D vector for face representation.

5.1.3 Classifier

Since our kinship verification is a binary classification problem and support vector machine (SVM) has demonstrated excellent performance for such tasks, we here apply SVM for classification. In our experiments, the RBF kernel was used for similarity measure of each pair of samples because we also found this kernel yields higher verification accuracy than other kernels. To tune the gamma parameter of this kernel, we applied the fourfold cross validation strategy on the training set to seek the optimal parameter. Specifically, we divided samples in the training set into 4 folds and each fold has nearly the same number of face pairs with kinship relation. We used three folds to train the SVM classifier and the remaining one to tune the parameter of SVM. This parameter was empirically set as 1.0 in our experiments.

5.1.4 Experimental Protocol

Generally, there are two types of protocols for verification tasks: closed-set and open-set [4]. In our experiments, we designed an open-set verification protocol because we expect our system can verify where there is a kinship relation for a new face pair without redesigning the verification system. We conducted fivefold cross validation experiments on our kinship data sets. Specifically, each subset of the KinFaceW-I and KinFaceW-II data sets was equally and sequentially divided into five folds such that each fold contains nearly the same number of face pairs with kinship relation. Table 1 lists the face number index for the five folds of these two data sets. For face images in each fold, we consider all pairs of face images with kinship relation as positive samples, and those without kinship relation as negative samples. Hence, the positive samples are the true pairs of face images (one from the parent and the other from the child), and the negative samples are false pairs of face images (one from the parent and the other from the child's image who is not his/her true child of the parent). Generally, the number of positive samples is much smaller than that of the negative samples. In our experiments, each parent face image was randomly combined

TABLE 1
Face Number Index Range of Different Folds
of the KinFaceW-I and KinFaceW-II Data Sets

| Fold | KinFaceW-I | | | | KinFaceW-II |
	F-D	F-S	M-D	M-S	All subset
1	[1, 27]	[1,31]	[1, 25]	[1,23]	[1, 50]
2	[28,54]	[32,64]	[26,50]	[24,46]	[51,100]
3	[55,81]	[65,96]	[51,75]	[47,69]	[101,150]
4	[82,108]	[97,124]	[76,101]	[70,102]	[151,200]
5	[109,134]	[125,156]	[102,127]	[93,116]	[201,250]

TABLE 2
Classification Accuracy (Percent) of Different
Feature Representation Methods on Different
Subsets of the KinFaceW-I Data Set

Feature	F-S	F-D	M-S	M-D	Mean
LBP	62.7	60.2	54.4	61.4	59.7
LE	66.1	59.1	58.9	68.0	63.0
SIFT	65.5	59.0	55.5	55.4	58.8
TPLBP	56.3	60.5	56.0	62.2	58.7

TABLE 3
Classification Accuracy (Percent) of Different Feature
Representation Methods on Different Subsets
of the KinFaceW-II Data Set

Feature	F-S	F-D	M-S	M-D	Mean
LBP	64.0	63.5	62.8	63.0	63.3
LE	69.8	66.1	72.8	72.0	69.9
SIFT	60.0	56.9	54.8	55.4	56.8
TPLBP	64.4	60.6	60.8	62.9	62.2

with a child image who is not his/her true child of the parent to construct a negative pair. Moreover, each pair of parent and child images appeared once in the negative samples. Therefore, the number of positive pairs and negative pairs are the same to learn the SVM classifier.

5.2 Results and Analysis

5.2.1 Baseline Results

For each test face pair, we decide whether there is a kinship relation by using the learned SVM classifier. The classification rate is defined as N_c/N_t, where N_c is the number of correct classification testing face pairs and N_t is the number of total testing face pairs. Tables 2 and 3 tabulate the classification rate of different feature representations methods on our kinship databases. As can be seen from these two tables, the best feature representation for kinship verification on our data sets is the LE feature and it outperforms the other three feature representation methods with the lowest gains in classification accuracy of 0.6 percent on the F-S subset, 0.1 percent on the F-D subset, 2.9 percent on the M-S subset, 5.8 percent on the M-D subset, and 3.3 percent on the mean accuracy of the KinFaceW-I data set, 5.5 percent on the F-S subset, 2.6 percent on the F-D subset, 10.0 percent on the M-S subset, 9.0 percent on the M-D subset, and 6.6 percent on the mean accuracy of the KinFaceW-II data set, respectively.

To better visualize the difference of different feature representation methods, the receiver operating characteristic (ROC) curves of different methods are shown in Fig. 8, where Figs. 8a and 8b plot the ROC curves of the results on the KinFaceW-I and KinFaceW-II data sets, respectively. We can see from this figure that the LE feature can yield the best performance in terms of the ROC curve.

5.2.2 Comparisons with Existing Metric Learning Algorithms

We have compared our methods with three other metric learning-based face verification algorithms which could also address the kinship verification problem, including CSML [46], NCA [23], and (LMNN) [57]. The neighborhood size k was empirically set as 5 for all five metric learning
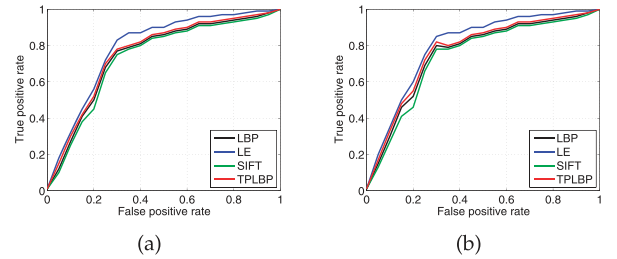

Fig. 8. The ROC curves of different methods obtained on the (a) KinFaceW-I and (b) KinFaceW-II data sets, respectively.

TABLE 4
Classification Accuracy (Percent) of Different Methods
on Different Subsets of the KinFaceW-I Data Set

Method	Feature	F-S	F-D	M-S	M-D	Mean
CSML	LBP	63.7	61.2	55.4	62.4	60.7
	LE	61.1	58.1	60.9	70.0	62.5
	SIFT	66.5	60.0	60.0	56.4	59.8
	TPLBP	57.3	61.5	63.2	57.0	59.7
NCA	LBP	61.7	62.2	56.4	62.4	60.7
	LE	62.1	57.1	61.9	69.0	62.3
	SIFT	67.5	61.0	61.0	57.4	60.8
	TPLBP	56.3	60.5	62.2	56.0	58.7
LMNN	LBP	62.7	63.2	57.4	63.4	61.7
	LE	63.1	58.1	62.9	70.0	63.3
	SIFT	69.5	63.0	63.0	59.4	62.8
	TPLBP	57.3	61.5	65.2	57.0	59.7
NRML	LBP	64.7	65.2	59.4	65.4	63.7
	LE	64.1	59.1	63.9	71.0	64.3
	SIFT	70.5	64.0	64.0	60.4	63.8
	TPLBP	59.3	63.5	65.2	60.0	62.9
MNRML	All	**72.5**	**66.5**	**66.2**	**72.0**	**69.9**

TABLE 5
Classification Accuracy (Percent) of Different Methods
on Different Subsets of the KinFaceW-II Data Set

Method	Feature	F-S	F-D	M-S	M-D	Mean
CSML	LBP	66.0	65.5	64.8	65.0	65.3
	LE	71.8	68.1	73.8	74.0	71.9
	SIFT	62.0	58.9	56.8	57.4	58.8
	TPLBP	66.4	62.6	62.8	64.9	64.2
NCA	LBP	67.0	66.5	65.8	66.0	66.3
	LE	73.8	70.1	74.8	75.0	73.5
	SIFT	63.0	59.9	58.8	59.4	60.4
	TPLBP	67.4	63.6	63.8	66.9	66.5
LMNN	LBP	68.0	68.5	68.8	67.0	68.2
	LE	74.8	71.1	75.8	76.0	74.5
	SIFT	65.0	57.9	58.8	59.4	60.4
	TPLBP	68.4	65.6	65.8	67.9	68.1
NRML	LBP	69.0	69.5	69.8	69.0	69.5
	LE	76.8	73.1	76.8	77.0	75.7
	SIFT	68.0	60.9	60.8	61.4	62.8
	TPLBP	70.4	67.6	67.8	69.9	70.1
MNRML	All	**76.9**	**74.3**	**77.4**	**77.6**	**76.5**

methods, and the feature dimensions of our proposed NRML and MNRML methods are empirically as 30 and 40, respectively.

Tables 4 and 5 tabulate the verification rate of different metric learning methods with different features on our kinship databases. The best recognition accuracy of different methods was selected for a fair comparison. As seen from these two tables, our proposed NRML (MNRML) methods outperform the other three compared methods with the lowest gains in accuracy of 1.0 percent (3.0 percent) on the F-S subset, 2.0 percent (3.3 percent) on the F-D subset, 2.0 percent (3.0 percent) on the M-S subset, 1.0 percent
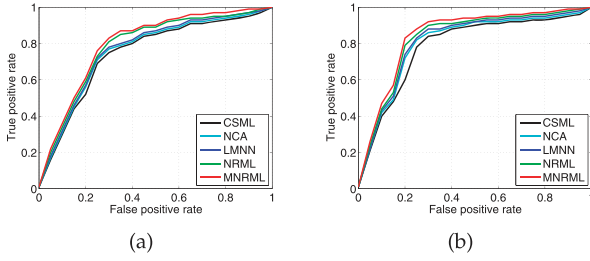
Fig. 9. The ROC curves of different methods obtained on the (a) KinFaceW-I and (b) KinFaceW-II data sets, respectively.

(2.0 percent) on the M-D subset, and 1.0 percent (6.6 percent) on the mean accuracy of the KinFaceW-I data set, 2.0 percent (3.1 percent) on the F-S subset, 2.0 percent (3.2 percent) on the F-D subset, 1.0 percent (1.6 percent) on the M-S subset, 1.0 percent (1.6a percent) on the M-D subset, and 1.2 percent (2.0 percent) on the mean accuracy of the KinFaceW-II data set, respectively.

We make four observations from the results listed in Tables 4 and 5:

- NRML consistently outperforms the other compared methods on all experiments on our data sets, which implies that learning a distance metric by considering and exploring the differences of different interclass samples can provide better discriminative information for verification.

- MNRML can improve the verification performance of NRML. The reason is MNRML can make use of multiple facial feature representations in a common learned distance metric such that some complementary information can be utilized for our kinship verification task.

- LE is the best feature representation among all used feature descriptors and it has achieved the best performance, which is consistent with some previous face verification results [9]. That is because the LE method has better utilized the local patch information of face images and such local patches may provide more discriminative information for kinship relation discovery.

- The results obtained on the KinFaceW-II data set are generally higher than those obtained on the KinFaceW-I data set, which indicates that kinship verification on the KinFaceW-I data set is more difficult than that on the KinFaceW-II data set. The reason is that face images in the KinFaceW-II data set are collected from the same photo and the kinship images have the same collection conditions, which could reduce some challenges caused by the illumination and aging variations in the KinFaceW-I data set.

- It is interesting to see that the performance of the same kinship relation from different kinship data sets are usually different. For example, we can see that the most difficult kinship verification task is the M-S for KinFace-I dataset and the F-D for the KinFace-II data set, respectively. We consider that this inconsistent observation may stem from the biased data sets. While there are hundreds of face pairs for each kinship relation in our data sets, the size of these data sets are still not enough to well model and discover the kinship relation. Hence, it is

still desirable to collect larger kinship data sets to further evaluate and enhance the generalization ability of existing kinship verification methods.

To further visualize the difference between our proposed methods and the other compared methods, the ROC curves of different methods are plotted in Fig. 9, where Figs. 9a and 9b plot the ROC curves of the results on the KinFaceW-I and KinFaceW-II data sets, respectively. Note that besides MNRML, the other four methods adopt the LE feature because it achieves the best verification accuracy among all four feature representation methods. We can see from this figure that the ROC curve of our methods is significantly higher than the other compared methods.

5.2.3 Comparisons with Different Classifiers

We compared the performance of our proposed NRML and MNRML methods versus different classifiers in kinship verification. Besides SVM, another two widely used classifiers are employed: the nearest neighbor (NN) and k-nearest neighbors (KNN) classifiers. For the KNN classifier, the number of k was empirically set as 11. Tables 6 and 7 tabulate the mean verification accuracy (percent) of our NRML and MNRML methods with different classifiers on the KinFaceW-I and KinFaceW-II data sets, respectively. We can see that the performance of our proposed methods are not sensitive to the selection of classifier and different classifiers can obtain comparable performance, which further demonstrate the robustness of our proposed methods in practical applications.

5.2.4 Comparisons with Existing Multiview Learning Methods

We compare our MNRML method with two existing multiview learning methods in our kinship verification task. The first is the multiview spectral embedding (MSE) method [62] which extends the conventional spectral embedding methods to handle multiview data; the second is the multiple kernel learning method (MKL) [3] where multiple feature representations are constructed as multiple kernels to describe data complementarily. Table 8 shows the

TABLE 6
Classification Accuracy (Percent) of Different Methods
on the KinFaceW-I Data Set

Method	Feature	SVM	NN	KNN
NRML	LBP	63.7	61.5	63.4
	LE	64.3	62.4	64.3
	SIFT	63.8	62.7	63.5
	TPLBP	62.9	61.8	62.5
MNRML	All	69.9	67.5	68.5

TABLE 7
Classification Accuracy (Percent) of Different Methods
on the KinFaceW-II Data Set

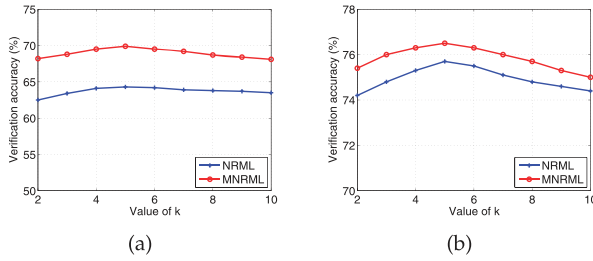| Method	Feature	SVM	NN	KNN
NRML	LBP	69.5	67.8	69.0
	LE	75.7	73.2	75.2
	SIFT	62.8	60.2	61.9
	TPLBP	70.1	68.7	69.4
MNRML	All	76.5	74.1	76.2

Fig. 10. Mean verification accuracy of NRML and MNRML versus different values of parameter k on the (a) KinFaceW-I and (b) KinFaceW-II data sets, respectively.
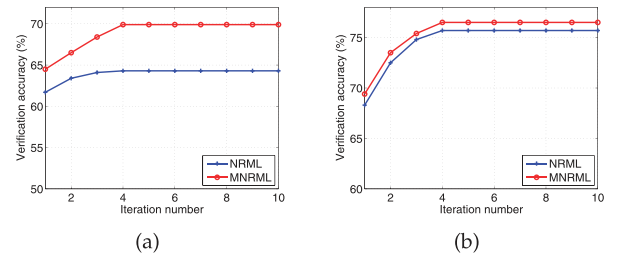

Fig. 11. Classification accuracy of NRML and MNRML versus different number of iterations on the (a) KinFaceW-I and (b) KinFaceW-II data sets, respectively.

mean verification accuracy of different multiview learning methods on the KinFaceW-I and KinFaceW-II data sets. We can see from this table that our MNRML can obtain comparable or even better performance than the other methods. That is because we emphasize the neighborhood samples in the learned distance metric while others have not effectively utilized this information.

5.2.5 Experiments on Race-Balanced Kinship Data Set

To further demonstrate that our methods are not primarily making decisions based on race, we selected the faces in the nonkinship face pairs such that all face pairs have the same race. Specifically, we selected 150 East Asian faces pairs and another 150 Caucasian faces pairs from the KinFaceW-I data set. Each selected face pairs are with the kinship relation, and hence there are 300 positive pairs in the selected subset. For the same race group, each parent face image was randomly combined with a child image who is not his/her true child of the parent to construct a negative pair. Hence, we have 300 positive face pairs and 300 negative pairs in total and all face pairs have the same race in the selected subset. We conducted kinship verification on such subset of the KinFaceW-I data set. We followed the fivefold experimental setting as discussed in Section 5.1. Table 9 shows the mean classification accuracy of our methods with different feature representations on this subset. As can be seen from this table, the results of our methods are not primarily making decisions based on race.

5.2.6 Parameter Analysis

We first investigate the effect of the parameter k of our proposed NRML and MNRML methods in this section. Fig. 10 shows the mean classification accuracy of NRML and MNRML versus different number of k of our NRML and MNRML, where Figs. 10a and 10b are the results obtained on the KinFaceW-I and KinFaceW-II databases, respectively. Here, the way we tuned the parameter k is similar to that of the parameter of the SVM classifier on the training set. We can see that our proposed NRML and MNRML can obtain the best classification performance when k is set as 5 for both NRML and MNRML. We can also observe that NRML and MNRML demonstrate a stable recognition performance versus varying neighborhood sizes. Hence, it is easy to select an appropriate neighborhood size for NRML and MNRML to obtain good performance in real applications.

Since NRML and MNRML are iterative algorithms, we also evaluated their performance with different number of iterations. Fig. 11 shows the mean verification accuracy of NRML and MNRML versus different number of iterations, where Figs. 11a and 11b are the results obtained on the

KinFaceW-I and KinFaceW-II databases, respectively. We can see that our proposed NRML and MNRML can converge to a local optimal peak in a few iterations.

Finally, we investigate the effect of the feature dimension of l of our proposed NRML and MNRML methods. Fig. 12 plots the mean classification accuracy of NRML and MNRML versus different number of feature dimension of NRML and MNRML, where Figs. 12a and 12b are the results obtained on the KinFaceW-I and KinFaceW-II databases, respectively. We can see that our proposed NRML and MNRML can obtain stable verification performance when the feature dimension is larger than 20 for NRML and 25 for MNRML, respectively.

5.2.7 Computational Complexity

We now briefly analyze the computational complexity of the NRML and MNRML methods, which involves T iterations. For NRML, each iteration calculates three matrices H_1, H_2, and H_3, and solves a standard eigenvalue equation. The time complexity of computing these two parts in each iteration is $O(Nk)$ and $O(m^3)$. Hence, the computational complexity of our proposed NRML is $O(NkT) + O(m^3T)$.

For the proposed MNRML method, each iteration involves calculating β and solving a standard eigenvalue equation. The time complexity of implementing these two parts in each iteration is $O((K + m)N^2)$ and $O(m^3)$. Hence, the computational complexity of our proposed MNRML is $O((K + m)N^2T) + O(m^3T)$.

We also list the computational times of the proposed NRML and MNRML methods and compare them with other three metric learning methods including CSML, NCA, and LMNN. Our hardware configuration is comprised of a 2.4 GHz CPU and a 6 GB RAM. Table 10 shows the time spent on the training and the testing (recognition) phases by these subspace methods, where the Matlab software, the
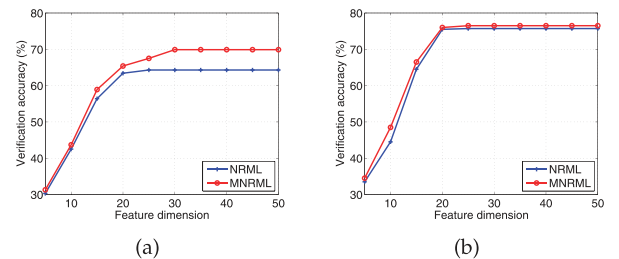

Fig. 12. Classification accuracy of NRML and MNRML versus different feature dimensions on the (a) KinFaceW-I and (b) KinFaceW-II data sets, respectively.

KinFaceW-I database and the nearest neighbor classifier were used.

From Table 10, we can see that the computational complexity of the proposed NRML and MNRML methods for training is generally larger than other metric learning methods. In practical applications, however, training is usually an offline process and only recognition needs to be performed in real time. Thus, the recognition time is usually more our concern than the training time. As shown in Table 10, the recognition times of the proposed methods are comparable to other metric learning methods. Hence, the computational complexity will not limit the real world applications of our proposed methods.

5.2.8 Comparisons with Human Observers in Kinship Verification

We also tested human ability in kinship verification from facial images. From each of the above four subsets in the KinFaceW-I and KinFaceW-II data sets, we randomly selected 100 pairs of face samples, 50 are positive and the other 50 are negative, and presented them to 10 human observers (five males and five females) aged 20 to 30 years old. These human observers have not received any training on how to verify kinship from facial images before the experiment. There are two parts in this experiment. For the first part, only the cropped face regions are shown to human observers (HumanA). For the second part, the whole original color face images are presented to human observers. Hence, HumanA aims to test kinship verification ability only from face part in the image, and HumanB intends to test the ability from multiple cues in the images such as face region, skin color, hair, and background. Therefore, the information provided in HumanA is the same as that provided to the algorithms tested in this study. Tables 11 and 12 show the accuracy of human ability on kinship verification on the KinFaceW-I and KinFaceW-II data sets, respectively. We can observe that our proposed NRML method can obtain better performance than HumanA, and perform slightly worse than HumanB on the KinFaceW-I data set, which further indicates that some other cues such as hair, skin color, and background also contribute to kinship verification. Moreover, our methods can achieve higher verification accuracies than both HumanA and HumanB on the KinFaceW-II data set.

5.2.9 Experiments with Other Data Sets

To further show the efficacy of our proposed methods, we conducted kinship verification experiments on two other data sets: Cornell KinFace [17] and UB KinFace [60], and compared our results with the performance of their methods.

There are 150 pairs of face images in the Cornell KinFace data set. Each pair contains one parent image and one child image, respectively. Face images in this data set are nearly frontal and with neutral facial expressions. Among all 150 pairs, 40 percent are of father-son, 22 percent are of father-daughter, 13 percent are of mother-son, and the remainder are mother-daughter relations, respectively.

For the UB KinFace data set, there are 600 face images of 400 different people, corresponding to 200 different groups. For each group, there are three images, corresponding to the child, young parent, and old parent, respectively. The young parent and old parent are the same person who is the parent of the child in this group, which are face photos of the parent when he/she was young and old, respectively. All images in this database are images of public figures (celebrities in entertainments, sports, and politicians) and downloaded from the Internet search. Since there are three images for each group, which can construct two kinship pairs. We constructed two subsets from the UB KinFace data set: Set 1 (200 old parent and child image pairs) and Set 2 (200 young parent and child image pairs). Since there are large imbalances of the four kinship relations of the UB Kinface database (nearly 80 percent of them are father-son relations), we have not considered separate kinship relation verification on this data set.

Similarly to previous experimental settings, we also extracted the LBP, LE, SIFT, TPLBP features for kinship verification. We also adopted the fivefold cross-validation strategy for experiments. Tables 13 and 14 tabulate the verification rate of our proposed methods on the Cornell KinFace and UB Kinface databases, respectively. We can observe from these two tables that our proposed methods notably outperform Xia's method [60], and is comparable to Fang's method [17] in kinship verification experiments.

5.2.10 Self-Kinship Verification

Finally, we investigated the problem of self-kinship verification by using two widely used face aging databases, namely FG-NET [31] and MORPH [49] data sets. Given a pair of face images, the objective of self-kinship verification is to determine whether they are from the same person or not. Hence, self-kinship verification is equivalent to the age-invariant face verification problem.

The FG-NET face database has been widely used in many age-related face analysis tasks such as facial age estimation and age-invariant face recognition. This data set contains 1,002 images from 82 subjects, and each subject has around 12 images, captured at different ages. The MORPH (version 2) data set is the largest face aging data set which is publicly available. It contains about 78,000 images of over 13,000 subjects, which were captured at different ages. In our study, we selected a subset of 26,000 images from 13,000 subjects and each subject has two images. These two images were selected such that the age gap of them is largest for the same person.

To evaluate the verification performance of our methods, each face image was automatically preprocessed:

1. rotate each image and align it vertically;
2. scale each image such that the eye positions of each face are the same;
3. crop face region and remove the background and the hair region; and
4. resize each image into 64×64. Fig. 13 shows several aligned and cropped example images from five different people from these two databases, respectively.

For the FG-NET database, we generated all 5,808 intrapersonal and 12,000 extrapersonal pairs, respectively. For the intrapersonal pairs, each pair of intraclass images was selected. For the interpersonal pairs, 12,000 pairs were randomly selected and generated from different subjects. For the MORPH data set, we generated all 13,000 intrapersonal pairs by collecting all image pairs from the same subjects. Fifteen thousand extrapersonal pairs are randomly selected from images of different subjects. In our
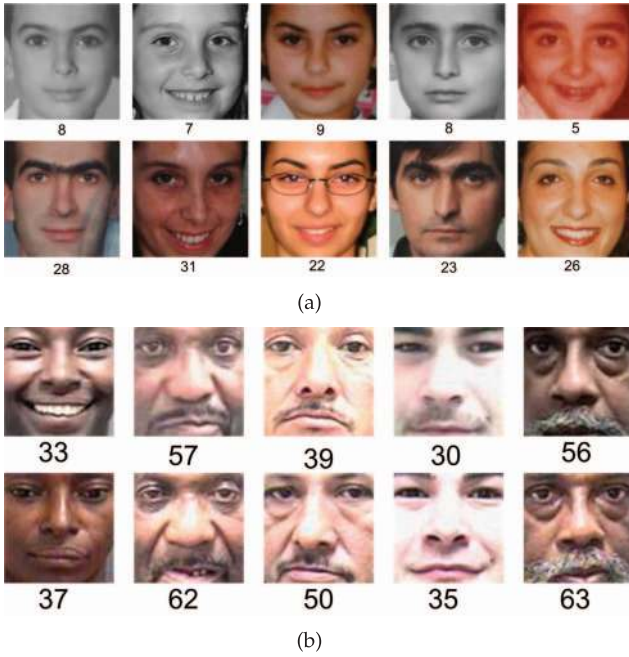
Fig. 13. Examples of face images from different people in the (a) FG-NET and (b) MORPH databases with different age values, where each column contains face images of the same person captured at different age values and the number below each image is the age value of the person.

TABLE 8
Classification Accuracy (Percent) of
Different Multiview Learning Methods on the
KinFaceW-I and KinFaceW-II Data Sets

Method	KinFaceW-I	KinFaceW-II
MSE	68.4	74.3
MKL	67.5	73.4
MNRML	69.9	76.5

TABLE 9
Classification Accuracy (Percent) of Our Methods on the
Race-Balanced Subset of the KinFaceW-I Data Set

Method	Feature	Mean
NRML	LBP	64.5
	LE	66.0
	SIFT	67.5
	TPLBP	64.5
MNRML	All	**69.0**

experiments, threefold cross validation was used, where images from the same subject never appear in both training and testing set in each fold. Therefore, there are 1,936 intrapersonal pairs and 4,000 extrapersonal pairs in each fold of the FG-NET data set, and about 4,366 intrapersonal pairs and 5,000 extrapersonal pairs in each fold of the MORPH data set, respectively. For each face image, we extracted the LBP, LE, SIFT, and TPLBP features, respectively. The SVM classifier was used for verification. We compared our methods with the following three state-of-the-art age-invariant face verification methods:

- GOP+SVM [33]: Each face image was represented by the gradient orientation pyramid (GOP) feature and the SVM classifier was used for verification.

TABLE 10
CPU Times (in Seconds) Used by
Different Methods on the KinFaceW-I Database

Method	Training	Recognition
CSML	12.50	1.24
NCA	14.50	1.46
LMNN	10.80	2.35
NRML	20.15	1.05
MNRML	25.50	1.05

TABLE 11
Accuracy (Percent) of Human Ability on Kinship Verification
on Different Kinship Subsets of the KinFaceW-I Data Set

Method	F-S	F-D	M-S	M-D	Mean
HumanA	61.00	58.00	66.00	70.00	63.75
HumanB	67.00	65.00	75.00	77.00	71.00

TABLE 12
Accuracy (Percent) of Human Ability on Kinship Verification
on Different Kinship Subsets of the KinFaceW-II Data Set

Method	F-S	F-D	M-S	M-D	Mean
HumanA	61.00	61.00	69.00	73.00	66.75
HumanB	70.00	68.00	78.00	80.00	74.00

TABLE 13
Classification Accuracy (Percent) of Different Methods
on the Cornell KinFace Data Set

Method	Feature	F-S	F-D	M-S	M-D	Mean
Fang's method	geological	72.9	54.6	73.8	61.3	65.7
NRML	LBP	67.1	63.8	75.0	60.0	66.5
	LE	72.7	66.8	75.4	63.2	69.5
	SIFT	64.5	67.3	68.4	61.8	65.5
	TPLBP	66.8	62.8	72.6	60.3	65.6
MNRML	All	**74.5**	**68.8**	**77.2**	**65.8**	**71.6**

- Bayesian eigenfaces [48]: Face images were represented by the eigenspace features and the Bayesian classifier was combined to model the intrapersonal and interpersonal face differences. The feature dimension of the eigenspace was selected as 200.
- Bagging LDA [32]: SIFT is used for feature representation and the bagging LDA method is used for verification. This method was originally designed for age-invariant face recognition, and we considered our face verification as a binary class face recognition problem. 20 classifiers were bagged for the final classification.

Table 15 tabulates the equal error rates (EER) of different methods on the FG-NET and MORPH data sets. We can see from this table that our proposed methods notably outperform the other compared age-invariant face verification methods, which can further demonstrate the effectiveness of our proposed methods.

5.3 Discussions

We make the following five key observations from experimental results listed in Tables 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, and 15, and Figs. 8, 9, 10, 11, and 12:

- LE is the best feature descriptor for kinship verification from facial images. Different from other hand crafted feature representation methods such as LBP, SIFT, and TPLBP, the LE feature is directly learned from training samples and hence it is more

TABLE 14
Classification Accuracy (Percent) of Different Methods
on the UB KinFace Data Set

Method	Feature	Set 1	Set 2
Xia's method	global intensity	55.0	N.A.
	local intensity	55.3	N.A.
	global Gabor	56.1	N.A.
	local Gabor	60.0	N.A.
NRML	LBP	63.4	61.2
	LE	61.9	61.3
	SIFT	62.5	62.8
	TPLBP	65.4	65.8
MNRML	All	**67.3**	**66.8**

TABLE 15
EERs of Different Age-Invariant Face Verification Methods
on the FG-NET and MORPH Data Sets

Method	FG-NET	MORPH
GOP+SVM [33]	29.5	10.5
Bayesian Eigenfaces [48]	32.5	9.7
Bagging LDA [32]	28.5	10.2
NRML	24.5	8.6
MNRML	22.5	7.5

data-adaptive and higher verification accuracy can be achieved.

- Our proposed NRML and MNRML methods outperform the other compared metric learning methods on our kinship verification tasks. That is because our methods emphasize the neighborhood samples in learning the distance metric while others have not effectively utilized this information such that more discriminative information can be exploited. Moreover, our proposed methods are not sensitive to the parameters and it is easy to select appropriate parameters to obtain good performance in real applications.

- Verifying human kinship relation in the same photo can obtain higher accuracy than in different photos. That is because face images collected from the same photo can reduce some challenges caused by the illumination and aging variations.

- Our proposed NRML and MNRML methods can obtain comparable kinship verification performance to that of human observers, which can further demonstrate the feasibility of verifying human kinship via facial image analysis and the efficacy of our proposed methods for practical applications.

- Our proposed NRML and MNRML methods notably outperform the other compared age-invariant face verification methods, which can further demonstrate their effectiveness in other face analysis applications.

6 CONCLUSION AND FUTURE WORK

We have proposed a neighborhood repulsed metric learning (NRML) method for kinship verification via facial image analysis. To the best of our knowledge, this paper is the first attempt to investigate kinship verification on the largest kinship data sets. Experimental results have shown that our proposed methods are not only significantly better than state-of-the-art metric learning methods, but also comparable to human observers in kinship verification

from facial images. How to explore more discriminative features and combine them with our proposed NRML and MNRML methods to further improve the verification performance appears to be another interesting direction of future work.

REFERENCES

[1] T. Ahonen et al., "Face Description with Local Binary Patterns: Application to Face Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 28, no. 12, pp. 2037-2041, Dec. 2006.
[2] A. Alvergne, R. Oda, C. Faurie, A. Matsumoto-Oda, V. Durand, and M. Raymond, "Cross-Cultural Perceptions of Facial Resemblance between Kin," *J. Vision,* vol. 9, no. 6, 2009.
[3] F. Bach, G. Lanckriet, and M. Jordan, "Multiple Kernel Learning, Conic Duality, and the SMO Algorithm," *Proc. Int'l Conf. Machine Learning,* pp. 1-8, 2004.
[4] E. Bailly-Bailliére et al., "The BANCA Database and Evaluation Protocol," *Proc. Int'l Conf. Audio-and Video-Based Biometric Person Authentication,* pp. 1057-1057, 2003.
[5] P.N. Belhumeur, J. Hespanha, and D.J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 7, pp. 711-720, July 1997.
[6] M. Belkin and P. Niyogi, "Laplacian Eigenmaps for Dimensionality Reduction and Data Representation," *Neural Computation,* vol. 15, no. 6, pp. 1373-1396, 2003.
[7] S. Bickel and T. Scheffer, "Multi-View Clustering," *Proc. IEEE Int'l Conf. Data Mining,* pp. 19-26, 2004.
[8] R. Brunelli and T. Poggio, "Face Recognition: Features versus Templates," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 15, no. 10, pp. 1042-1052, Oct. 2002.
[9] Z. Cao, Q. Yin, X. Tang, and J. Sun, "Face Recognition with Learning-Based Descriptor," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 2707-2714, 2010.
[10] H. Chen, H. Chang, and T. Liu, "Local Discriminant Embedding and Its Variants," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 846-853, 2005.
[11] I. Cohen, N. Sebe, A. Garg, L. Chen, and T. Huang, "Facial Expression Recognition from Video Sequences: Temporal and Static Modeling," *Computer Vision and Image Understanding,* vol. 91, no. 1, pp. 160-187, 2003.
[12] T. Cootes, G. Edwards, and C. Taylor, "Active Appearance Models," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 23, no. 6, pp. 681-685, June 2001.
[13] M. Dal Martello and L. Maloney, "Where Are Kin Recognition Signals in the Human Face?" *J. Vision,* vol. 6, no. 12, 2006.
[14] M. Dal Martello and L. Maloney, "Lateralization of Kin Recognition Signals in the Human Face," *J. Vision,* vol. 10, no. 8, 2010.
[15] J. Davis, B. Kulis, P. Jain, S. Sra, and I. Dhillon, "Information-Theoretic Metric Learning," *Proc. Int'l Conf. Machine Learning,* pp. 209-216, 2007.
[16] L. DeBruine, F. Smith, B. Jones, S. Craig Roberts, M. Petrie, and T. Spector, "Kin Recognition Signals in Adult Faces," *Vision Research,* vol. 49, no. 1, pp. 38-43, 2009.

[17] R. Fang, K. Tang, N. Snavely, and T. Chen, "Towards Computational Models of Kinship Verification," *Proc. IEEE Int'l Conf. Image Processing,* pp. 1577-1580, 2010.

[18] B. Fasel and J. Luettin, "Automatic Facial Expression Analysis: A Survey," *Pattern Recognition,* vol. 36, no. 1, pp. 259-275, 2003.

[19] Y. Fu and T. Huang, "Human Age Estimation with Regression on Discriminative Aging manifold," *IEEE Trans. Multimedia,* vol. 10, no. 4, pp. 578-584, June 2008.

[20] Y. Fu, S. Yan, and T.S. Huang, "Classification and Feature Extraction by Simplexization," *IEEE Trans. Information Forensics and Security,* vol. 3, no. 1, pp. 91-100, Mar. 2008.

[21] X. Geng, Z. Zhou, and K. Smith-Miles, "Automatic Age Estimation Based on Facial Aging Patterns," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 29, no. 12, pp. 2234-2240, Dec. 2007.

[22] A.S. Georghiades, P.N. Belhumeur, and D.J. Kriegman, "From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 23, no. 6, pp. 643-660, June 2001.

[23] J. Goldberger, S. Roweis, G. Hinton, and R. Salakhutdinov, "Neighborhood Component Analysis," *Proc. Conf. Advances in Neural Information Processing Systems,* pp. 2539-2544, 2004.

[24] G. Guo, Y. Fu, C. Dyer, and T. Huang, "Image-Based Human Age Estimation by Manifold Learning and Locally Adjusted Robust Regression," *IEEE Trans. Image Processing,* vol. 17, no. 7, pp. 1178-1188, July 2008.

[25] G. Guo, G. Mu, Y. Fu, and T. Huang, "Human Age Estimation Using Bio-Inspired Features," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 112-119, 2009.

[26] S. Gutta, J. Huang, P. Jonathon, and H. Wechsler, "Mixture of Experts for Classification of Gender, Ethnic Origin, and Pose of Human Faces," *IEEE Trans. Neural Networks,* vol. 11, no. 4, pp. 948-960, July 2002.

[27] X. He, S. Yan, Y. Hu, P. Niyogi, and H.J. Zhang, "Face Recognition Using Laplacianfaces," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 27, no. 3, pp. 328-340, Mar. 2005.

[28] H. Hu, "Orthogonal Neighborhood Preserving Discriminant Analysis for Face Recognition," *Pattern Recognition,* vol. 41, no. 6, pp. 2045-2054, 2008.

[29] G. Kaminski, S. Dridi, C. Graff, and E. Gentaz, "Human Ability to Detect Kinship in Strangers' Faces: Effects of the Degree of Relatedness," *Proc. Royal Soc. B: Biological Sciences,* vol. 276, no. 1670, pp. 3193-3200, 2009.

[30] G. Kaminski, F. Ravary, C. Graff, and E. Gentaz, "Firstborns' Disadvantage in Kinship Detection," *Psychological Science,* vol. 21, no. 12, pp. 1746-1750, 2010.

[31] A. Lanitis, "Evaluating the Performance of Face-Aging Algorithms," *Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition,* pp. 1-6, 2008.

[32] Z. Li, U. Park, and A. Jain, "A Discriminative Model for Age Invariant Face Recognition," *IEEE Trans. Information Forensics and Security,* vol. 6, no. 3, pp. 1028-1037, Sept. 2011.

[33] H. Ling, S. Soatto, N. Ramanathan, and D. Jacobs, "Face Verification across Age Progression Using Discriminative Methods," *IEEE Trans. Information Forensics and Security,* vol. 5, no. 1, pp. 82-91, Mar. 2010.

[34] D. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Int'l J. Computer Vision,* vol. 60, no. 2, pp. 91-110, 2004.

[35] H. Lu, K. Plataniotis, and A. Venetsanopoulos, "MPCA: Multilinear Principal Component Analysis of Tensor Objects," *IEEE Trans. Neural Networks,* vol. 19, no. 1, pp. 18-39, Jan. 2008.

[36] J. Lu, J. Hu, X. Zhou, Y. Shang, Y.-P. Tan, and G. Wang, "Neighborhood Repulsed Metric Learning for Kinship Verification," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 2594-2601, 2012.

[37] J. Lu and Y. Tan, "A Doubly Weighted Approach for Appearance-Based Subspace Learning Methods," *IEEE Trans. Information Forensics and Security,* vol. 5, no. 1, pp. 71-81, Mar. 2010.

[38] J. Lu and Y. Tan, "Fusing Shape and Texture Information for Facial Age Estimation," *Proc. IEEE Int'l Conf. Acoustics, Speech, and Signal Processing,* pp. 1477-1480, 2011.

[39] J. Lu and Y.-P. Tan, "Ordinary Preserving Manifold Analysis for Human Age and Head Pose Estimation." *IEEE Tans. Human-Machine Systems,* vol. 43, no. 2, pp. 249-258, 2013.

[40] J. Lu and Y.-P. Tan, "Regularized Locality Preserving Projections and Its Extensions for Face Recognition," *IEEE Trans. Systems, Man, and Cybernetics: Part B, Cybernetics,* vol. 40, no. 2, pp. 958-963, June 2010.

[41] J. Lu and Y.-P. Tan, "Uncorrelated Discriminant Nearest Feature Line Analysis for Face Recognition," *IEEE Signal Processing Letters,* vol. 17, no. 2, pp. 185-188, Feb. 2010.

[42] J. Lu, Y.-P. Tan, and G. Wang, "Discriminative Multi-Manifold Analysis for Face Recognition from a Single Training Sample per Person," *Proc. IEEE Int'l Conf. Computer Vision,* pp. 1943-1950, 2011.

[43] J. Lu, Y.-P. Tan, and G. Wang, "Discriminative Multimanifold Analysis for Face Recognition from a Single Training Sample per Person," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 35, no. 1, pp. 39-51, Jan. 2013.

[44] B. Moghaddam and M. Yang, "Gender Classification with Support Vector Machines," *Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition,* pp. 306-311, 2000.

[45] B. Moghaddam and M. Yang, "Learning Gender with Support Faces," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 24, no. 5, pp. 707-711, May 2002.

[46] H. Nguyen and L. Bai, "Cosine Similarity Metric Learning for Face Verification," *Proc. Asian Conf. Computer Vision,* pp. 709-720, 2011.

[47] Y. Ou, X. Wu, H. Qian, and Y. Xu, "A Real Time Race Classification System," *Proc. IEEE Int'l Conf. Information Acquisition,* pp. 1-6, 2005.

[48] N. Ramanathan and R. Chellappa, "Face Verification across Age Progression," *IEEE Trans. Image Processing,* vol. 15, no. 11, pp. 3349-3361, Nov. 2006.

[49] K. Ricanek and T. Tesafaye, "MORPH: A Longitudinal Image Database of Normal Adult Age-Progression," *Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition,* pp. 341-345, 2006.

[50] S. Roweis and L. Saul, "Nonlinear Dimensionality Reduction by Locally Linear Embedding," *Science,* vol. 290, no. 5500, pp. 2323-2326, 2000.

[51] S. Ruping and T. Scheffer, "Learning with Multiple Views," *Proc. IEEE Int'l Conf. Machine Learning Workshop Learning with Multiple Views,* 2005.

[52] S. Schiffman, M. Reynolds, F. Young, and J. Carroll, *Introduction to Multidimensional Scaling: Theory, Methods, and Applications.* Academic Press New York, 1981.

[53] S. Shan, Y. Chang, W. Gao, B. Cao, and P. Yang, "Curse of Mis-Alignment in Face Recognition: Problem and a Novel Mis-Alignment Learning Solution," *Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition,* pp. 314-320, 2004.

[54] M. Turk and A. Pentland, "Eigenfaces for Recognition," *J. Cognitive Neuroscience,* vol. 3, no. 1, pp. 71-86, 1991.

[55] X. Wang and X. Tang, "A Unified Framework for Subspace Face Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 26, no. 9, pp. 1222-1228, Sept. 2004.

[56] X. Wang and X. Tang, "Random Sampling for Subspace Face Recognition," *Int'l J. Computer Vision,* vol. 70, no. 1, pp. 91-104, 2006.

[57] K. Weinberger, J. Blitzer, and L. Saul, "Distance Metric Learning for Large Margin Nearest Neighbor Classification," *Proc. Conf. Advances in Neural Information Processing Systems,* 2005.

[58] L. Wiskott, J.-M. Fellous, N. Kuiger, and C. von der Malsburg, "Face Recognition by Elastic Bunch Graph Matching," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 7, pp. 775-779, July 1997.

[59] L. Wolf, T. Hassner, and Y. Taigman, "Descriptor Based Methods in the Wild," *Proc. European Conf. Computer Vision Workshop,* 2008.

[60] S. Xia, M. Shao, and Y. Fu, "Kinship Verification through Transfer Learning," *Int'l Joint Conf. Artificial Intelligence,* pp. 2539-2544, 2011.

[61] S. Xia, M. Shao, J. Luo, and Y. Fu, "Understanding Kin Relationships in a Photo," *IEEE Trans. Multimedia,* vol. 14, no. 4, pp. 1046-1056, Aug. 2012.

[62] T. Xia, D. Tao, T. Mei, and Y. Zhang, "Multiview Spectral Embedding," *IEEE Trans. Systems, Man, and Cybernetics, Part B: Cybernetics,* vol. 40, no. 6, pp. 1438-1446, Dec. 2010.

[63] B. Xie, Y. Mu, D. Tao, and K. Huang, "M-SNE: Multiview Stochastic Neighbor Embedding," *IEEE Trans. Systems, Man, and Cybernetics, Part B: Cybernetics,* vol. 41, no. 4, pp. 1088-1096, Aug. 2011.

[64] S. Yan, D. Xu, Q. Yang, L. Zhang, X. Tang, and H. Zhang, "Multilinear Discriminant Analysis for Face Recognition," *IEEE Trans. Image Processing,* vol. 16, no. 1, pp. 212-220, Jan. 2007.

[65] S. Yan, D. Xu, B. Zhang, H. Zhang, Q. Yang, and S. Lin, "Graph Embedding and Extensions: A General Framework for Dimensionality Reduction," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 29, no. 1, pp. 40-51, Jan. 2007.

[66] M.H. Yang, "Kernel Eigenfaces vs. Kernel Fisherfaces: Face Recognition Using Kernel Methods," *Proc. IEEE Int'l Conf. Face and Gesture Recognition,* pp. 215-220, 2002.

[67] W. Yu, X. Teng, and C. Liu, "Face Recognition Using Discriminant Locality Preserving Projections," *Image and Vision Computing,* vol. 24, no. 3, pp. 239-248, 2006.

[68] 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.

[69] Z. Zhao and H. Liu, "Multi-Source Feature Selection via Geometry-Dependent Covariance Analysis," *J. Machine Learning Research,* vol. 4, pp. 36-47, 2008.

[70] W. Zheng, X. Zhou, C. Zou, and L. Zhao, "Facial Expression Recognition Using Kernel Canonical Correlation Analysis (KCCA)," *IEEE Trans. Neural Networks,* vol. 17, no. 1, pp. 233-238, Jan. 2006.

[71] D. Zhou and C. Burges, "Spectral Clustering and Transductive Learning with Multiple Views," *Proc. Int'l Conf. Machine Learning,* pp. 1159-1166, 2007.

[72] X. Zhou, J. Hu, Y. Lu, J. Shang, and Y. Guan, "Kinship Verification from Facial Images under Uncontrolled Conditions," *Proc. ACM Int'l Conf. Multimedia,* pp. 953-956, 2011.

[73] A. Zien and C. Ong, "Multiclass Multiple Kernel Learning," *Proc. Int'l Conf. Machine Learning,* pp. 1191-1198, 2007.

Jiwen Lu received the BEng degree in mechanical engineering and the MEng degree in electrical engineering, both from the Xi'an University of Technology, China, in 2003 and 2006, respectively, and the PhD degree in electrical engineering from the Nanyang Technological University, Singapore, in 2011. He is currently a research scientist at the Advanced Digital Sciences Center, Singapore. His research interests include computer vision, pattern recognition, machine learning, and biometrics. He has authored/coauthored more than 70 scientific papers in peer-reviewed journals and conferences including some top venues such as the *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *IEEE Transactions on Image Processing*, *IEEE Transactions on Circuits and Systems for Video Technology*, *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*, *IEEE Transactions on Information Forensics and Security*, ICCV, CVPR, and ACM MM. He received the First-Prize National Scholarship and the National Outstanding Student awarded by the Ministry of Education of China, in 2002 and 2003, and the Best Student Paper Award from PREMIA of Singapore in 2012, respectively. He is a member of the IEEE.

Xiuzhuang Zhou received the BSci degree from the Department of Atmosphere Physics, Chengdu University of Information Technology, China, in 1996, and the MEng and PhD degrees from the School of Computer Science, Beijing Institute of Technology, China, in 2005 and 2011, respectively. He is currently an assistant professor in the College of Information Engineering, Capital Normal University, Beijing, China. His research interests include computer vision, pattern recognition, and machine learning. He has authored several scientific papers in peer-reviewed journals and conferences, including some top venues such as the *IEEE Transactions on Image Processing*, *IEEE Transactions on Information Forensics and Security*, CVPR, and ACM MM. He is a member of the IEEE.

Yap-Peng Tan received the BS degree from National Taiwan University, Taipei, in 1993, and the MA and PhD degrees from Princeton University, New Jersey, in 1995 and 1997, respectively, all in electrical engineering. From 1997 to 1999, he was with Intel Corporation, Chandler, Arizona, and Sharp Laboratories of America, Camas, Washington. In November 1999, he joined the Nanyang Technological University of Singapore, where he is currently an associate professor and associate chair (academic) in the School of Electrical and Electronic Engineering. His current research interests include image and video processing, content-based multimedia analysis, computer vision, and pattern recognition. He is the principal inventor or co-inventor on 15 US patents in the areas of image and video processing. He was the recipient of an IBM Graduate Fellowship from the IBM T.J. Watson Research Center, Yorktown Heights, New York, from 1995 to 1997. He currently serves as the secretary of the Visual Signal Processing and Communications Technical Committee of the IEEE Circuits and Systems Society, a member of the Multimedia Signal Processing Technical Committee of the IEEE Signal Processing Society, and a voting member of the ICME Steering Committee. He is an editorial board member of the *EURASIP Journal on Advances in Signal Processing* and *EURASIP Journal on Image and Video Processing*, and an associate editor of the *Journal of Signal Processing Systems*. He was also a guest editor for special issues of several journals, including the *IEEE Transactions on Multimedia*. He was the general cochair of the 2010 IEEE International Conference on Multimedia and Expo (ICME 2010). He is a senior member of the IEEE.

Yuanyuan Shang received the PhD degree from National Astronomical Observatories, Chinese Academy of Sciences in 2005. She is currently a professor and vice dean (research) of the College of Information Engineering, Capital Normal University, Beijing, China. Her research interests include digital imaging sensor, image processing, and computer vision. She has authored more than 40 scientific papers in peer-reviewed journals and conferences, including some top venues such as the *IEEE Transactions on Information Forensics and Security* and ACM MM. She is a member of the IEEE.

Jie Zhou received the BS and MS degrees from the Department of Mathematics, Nankai University, Tianjin, China, in 1990 and 1992, respectively, and the PhD degree from the Institute of Pattern Recognition and Artificial Intelligence, Huazhong University of Science and Technology, Wuhan, China, in 1995. From 1995 to 1997, he was a postdoctoral fellow with the Department of Automation, Tsinghua University, Beijing, China. Currently, he is a full professor with the Department of Automation, Tsinghua University. His research interests include pattern recognition, computer vision, and data mining. In recent years, he has authored more than 100 papers in peer-reviewed international journals and conferences such as the *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *IEEE Transactions on Image Processing*, and CVPR. He is acting as an associate editor for several journals, including the *International Journal of Robotics and Automation* (*IJRA*) and *Acta Automatica Sinica*. He received the Outstanding Youth Fund of NSFC in 2012. He is a senior member of the IEEE.