Neighborhood Repulsed Metric Learning for Kinship Verification

Jiwen Lu[1], Junlin Hu[2], Xiuzhuang Zhou[3], Yuanyuan Shang[3], Yap-Peng Tan[4], Gang Wang[1,4]
[1]Advanced Digital Sciences Center, Singapore
[2]College of Information Science and Technology, Beijing Normal University, Beijing, China
[3]School of Information Engineering, Capital Normal University, Beijing, China
[4]School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore
Email: jiwen.lu@adsc.com.sg

Abstract

Kinship verification from facial images is a challenging problem in computer vision, and there is a very few attempts on tackling 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 kinship relations) 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 kinship relations) are pushed as close as possible and interclass samples lying in a neighborhood are repulsed and pulled as far as possible, simultaneously, such that more discriminative information can be exploited for verification. Moreover, we propose a multiview NRML (MNRML) method to seek a common distance metric to make better use of multiple feature descriptors to further improve the verification performance. Experimental results are presented to demonstrate the efficacy of the proposed methods.

Figure 1. Several examples of our kinship database. From top to bottom are face images with the Father-Son (FS), Father-Daughter (FD), Mother-Son (MS) and Mother-Daughter (MD) kinship relations, respectively.

1. Introduction

Facial images convey many important human characteristics, such as identity, gender, expression, age, and ethnicity. Over the past two decades, a large number of face analysis problems have been investigated in computer vision. Representative examples including face recognition [2, 15, 21, 22, 29], facial expression recognition [6], age estimation [7, 9, 14], gender classification [17] and ethnicity recognition [10]. While encouraging results have been obtained in previous studies, most existing work focus on face analysis under controlled conditions and suffer from great variations in many real-world applications where face images are captured under uncontrolled conditions.

Recently, a new face dataset called the Labeled Faces in the Wild (LFW) [11] was created for the face identification research under uncontrolled conditions. Different from most previous face datasets such as AR [16], FERET [19] and CMU PIE [20], LFW was specifically designed for advancing face recognition techniques for practical applications because facial images in this dataset were collected from the real world environments and many natural variations rather than artificial controls were included. Hence, face recognition methods/algorithms developed on this dateset are much closer to practical applications. However, due to the inevitable impact factors such as pose, expression, lighting and aging on faces, human identification through unconstrained face images remains unsolved.

In this paper, we investigate the problem of human kinship verification from facial images under uncontrolled conditions. Figure 1 shows some representative examples of kinship images[1]. This new research topic has several potential applications such as family album organization, image annotation, and missing children searching. However, little research has been systematically conducted along this direction, possibly due to lacking of such publicly available kinship databases and great challenges of this problem. To address this, we collect two new kinship face databases named KinFaceW-I (KFW-I) and KinFaceW-II (KFW-II)[2] 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 pulled as far as as possible. Since interclass samples (without 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 (without kinship relation) in a neighborhood more in learning the distance metric and expect those samples lying in a neighborhood are repulsed and pulled 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, we propose a multiview neighborhood repulsed metric learning (MNRML) method to seek a common distance metric to make better use of multiple feature descriptors to further improve the verification performance. Experimental results are presented to demonstrate the efficacy of the proposed methods.

2. Related Work

Kinship Verification: Fang *et al*. [5] was the first attempt to tackle the challenge of kinship verification from facial images by using local facial feature extraction and selection. They first localized some key parts of facial features such as kin color, gray value, histogram of gradient, and facial structure information were employed to describe facial images. Then, the k-nearest-neighbor (KNN) with Euclidean metric was applied to classify face images. More recently, Xia *et al*. [25] proposed a new transfer subspace learning method for kinship verification. Their key 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 their work: 1) they used the conventional Euclidean metric for kinship verification and such metric is not appropriate to measure the similarity of facial images because the intrinsic space that face usually lies in is a low-dimensional manifold rather than the Euclidean space; 2) their method was evaluated on comparatively small datasets (150 pairs in [5] and 180 pairs in [25], respectively), which are rather small to demonstrate the effectiveness of face analysis-based kinship verification. Hence, more robust and effective metrics and larger kinship datasets are desirable to demonstrate and improve the performance of existing kinship verification methods.

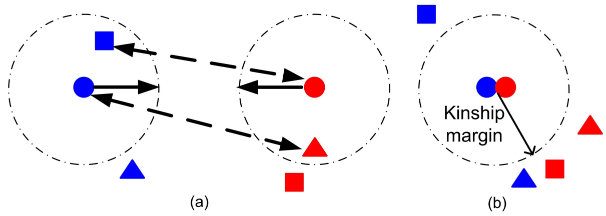


Fig. 2. Intuitive illustration of the proposed NRML method. (a) The original kinship face images in the high-dimensional feature space. (b) The expected distributions of kinship images in the learned metric space.

Metric Learning: Metric learning has received a lot of attention in recent years, and there have been some such algorithms proposed in the literature. Representative metric learning algorithms include neighborhood component analysis (NCA) [8], cosine similarity metric learning (CSML) [18], large margin nearest neighbor (LMNN) [23], and information theoretic metric learning (ITML) [4]. 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 such different contributions of the samples to learn the distance metric; 2) existing metric learning methods assume that data are drawn from a vector space and thus cannot handle multiview data directly. To address this problem, we propose a new multiview metric learning method to learn a robust metric by considering different importance of face samples and making use of multiple feature descriptors, simultaneously.

3. Proposed Methods

3.1. Basic Idea

Figure 2 shows the basic idea of our proposed NRML method. There are three pairs of kinship images in Fig-

[1]All the face images shown in this paper were collected from Internet search, which are only used for academic research (noncommercial).

[2]The difference of KFW-I and KFW-II is that each pair of kinship facial images in KFW-I was collected from different photos and that in KFW-II was collected from the same photo. More examples will be provided in Section 4.

ure 2(a), denoted by circles, squares and triangles, and the blue and red colors denotes the parent and child facial images, respectively. In the original image space, there is large difference between the parent and child images in the circle class due to some variation factors such as aging, illumination and expression. Hence, there are some other parent and child images lying in the neighborhoods of the parent and child images in the circle class, as shown in Figure 2(a), which is the main challenge in our task because there is a high chance to misclassify the images in the neighborhoods. 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 are pulled as far as as possible, as shown in Figure 2(b). As a result, the kinship margin in the learned distance metric space is much larger and more discriminative information can be exploited for kinship verification.

3.2. NRML

Let $S = \{(x_i, y_i)|i = 1, 2, \cdots, N\}$ be the training set of N pairs of kinship images, where $x_i \in R^m$ and $y_i \in R^m$ are the ith parent and child images, 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 possible, and that between x_i and y_j ($i \neq j$) are as large as possible, simultaneously, where

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

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 satisfy the symmetry, nonnegativity and triangle inequality. Hence, A must be symmetric and positive semidefinite.

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

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

where y_{it_1} represents the t_1th k-nearest neighbors of y_i and x_{it_2} denotes the t_2th k-nearest neighbors of x_i, respectively, the metric d is defined as Eq. (1). The objective function of J_1 in Eq. (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 function of J_2 in Eq. (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 Eq. (2) ensures that x_i and y_i are pushed as close as possible in the learned distance metric space because they have kinship relations.

It can be seen that the optimization criterion in Eq. (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 the best of our knowledge, there is no closed-form solution for this objective function. We solve this problem in an iterative manner inspired by recent advances in EM-based algorithms [13, 28]. 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, Eq. (1) can be rewritten as

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

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

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

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

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

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

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

Algorithm 1: NRML

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

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 $r = 1, 2, \cdots, T$, repeat
2.1. Compute H_1, H_2 and H_3, respectively.
2.2. Solve the eigenvalue problem defined in Eq. (9).
2.3. Obtain $W^r = [w_1, w_2, \cdots, w_l]$.
2.4. Update 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 $W = W^r$.

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 $H_3 = \frac{1}{Nk}\sum_{i=1}^{N}(x_i - y_i)(x_i - y_i)^T$.

Now, we can formulate our NRML method as

$$\max_{W} J(W) = tr[W^T(H_1 + H_2 - H_3)W] \quad (8)$$
$$\text{subject to} \quad W^T W = I.$$

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

$$(H_1 + H_2 - H_3)w = \lambda w. \quad (9)$$

Let $w_1, w_2, \cdots, w_l$ be the eigenvectors of Eq. (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 $W = [w_1, w_2, \cdots, w_l]$ can be obtained to project the original face samples x_i and y_i into low-dimensional feature feature vectors u_i and v_i, as follows:

$$u_i = W^T x_i, v_i = W^T y_i, \quad i = 1, 2, \cdots, N. \quad (10)$$

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

3.3. MNRML

Previous studies have shown that different feature descriptors could provide complementary information in characterizing facial information from different viewpoints [26, 27], 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 cannot deal with such multiview data directly. 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 $S^p = \{(x_i^p, y_i^p)|i = 1, 2, \cdots, N\}$ be 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 $p = 1, 2, \cdots, K$. 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 $(i \neq j)$ are as large as possible, simultaneously.

In order to well discover the complemental information of facial images from different views, we impose a nonnegative weighted vector $\beta = [\beta_1, \beta_2, \cdots, \beta_K]$ 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 made to learn the distance metric. Hence, we formulate MNRML as the following optimization problem

$$\max_{W,\beta} \sum_{p=1}^{K} \beta_p tr[W^T(H_1^p + H_2^p - H_3^p)W] \quad (11)$$
$$\text{subject to} \quad W^T W = I, \sum_{p=1}^{K} \beta_p = 1, \beta_p \geq 0.$$

The solution to Eq. (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

$$\max_{W,\beta} \sum_{p=1}^{K} \beta_p^q tr[W^T(H_1^p + H_2^p - H_3^p)W] \quad (12)$$
$$\text{subject to} \quad W^T W = I, \sum_{p=1}^{K} \beta_p = 1, \beta_p \geq 0.$$

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

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

$$
\begin{aligned}
L(\beta,\zeta) &= \sum_{p=1}^{K}\beta_p^q tr[W^T(H_1^p+H_2^p-H_3^p)W] \\
&\quad -\zeta(\sum_{p=1}^{K}\beta_p-1)
\end{aligned} \tag{13}
$$

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

$$
q\beta_p^{q-1}tr[W^T(H_1^p+H_2^p-H_3^p)W]-\zeta=0 \tag{14}
$$

$$
\sum_{p=1}^{K}\beta_p-1=0 \tag{15}
$$

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

$$
\beta_p=\frac{(1/tr[W^T(H_1^p+H_2^p-H_3^p)W])^{1/(q-1)}}{\sum_{p=1}^{K}(1/tr[W^T(H_1^p+H_2^p-H_3^p)W])^{1/(q-1)}} \tag{16}
$$

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

$$
\begin{aligned}
&\max_{W} tr[W^T(\sum_{p=1}^{K}(H_1^p+H_2^p-H_3^p))W] \\
&\text{subject to} \quad W^TW=I.
\end{aligned} \tag{17}
$$

And W can be obtained by solving the following eigenvalue equation

$$
\left(\sum_{p=1}^{K}(H_1^p+H_2^p-H_3^p)\right)w=\lambda w. \tag{18}
$$

The proposed MNRML algorithm is summarized in **Algorithm 2**.

3.4. Computational Complexity

We now briefly analyze the computational complexity of the NRML and MNRML methods, which involves T iterations. For NRML, each iteration calculate 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 MNRML method, each iteration involves calculate β and solving a standard eigenvalue equation. The time complexity of implement these two parts in each iteration is $O((K+m)N^2)$ and $O(m^3)$. Hence, the computational complexity of our proposed NRML is $O((K+m)N^2T)+O(m^3T)$.

Algorithm 2: MNRML

Input: Training images:
$S^p=\{(x_i^p,y_i^p)|i=1,2,\cdots,N\}$ 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.

Output: Distance metric W.

Step 1 (Initialization):
- **1.1**. Set $\beta=[1/K,1/K,\cdots,1/K]$;
- **1.2**. Obtain W^0 by solving Eq. (18).

Step 2 (Local optimization):
For $r=1,2,\cdots,T$, repeat
- **2.1**. Compute β by using Eq. (16).
- **2.2**. Obtain W^r by solving Eq. (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 $W=W^r$.

---

4. Experiments

We have evaluated the proposed NRML and MNRML methods by conducting a number of kinship verification experiments on our two datasets. The following describes the details of the experiments and results.

4.1. Data Sets

To advance the kinship verification research and show the efficacy of our proposed methods, we collected two kinship face datasets from the internet through an online search for images of public figures or celebrities and their parents or children, named KFW-I and KFW-II. The difference of KFW-I and KFW-II is that each pair of kinship facial images in KFW-I was collected from different photos and that in KFW-II was collected from the same photo. We pose no restrictions in terms of pose, lighting, background, expression, age, ethnicity and partial occlusion on the images used for training and testing. Some examples from the KFW-I dataset are shown in Figure 1, and Figure 3 shows some samples in the KFW-II dataset.

There are four kinship relations in both the KFW-I and KFW-II datasets: Father-Son (FS), Father-Daughter (FD), Mother-Son (MS) and or Mother-Daughter (MD). In the KFW-I dataset, there are 134, 156, 127 and 116 pairs of kinship images for these four relations. For the KFW-II dataset, each relation contains 250 pairs of kinship images. Figure 4 shows the ethnicity distributions of our datasets.

4.2. Experimental Settings

In our experiments, the images were converted to grayscale and normalized to 64×64 pixels according to the manually labeled eyes positions. We adopted the 5-fold cross-validation strategy for experiments. For each of the four

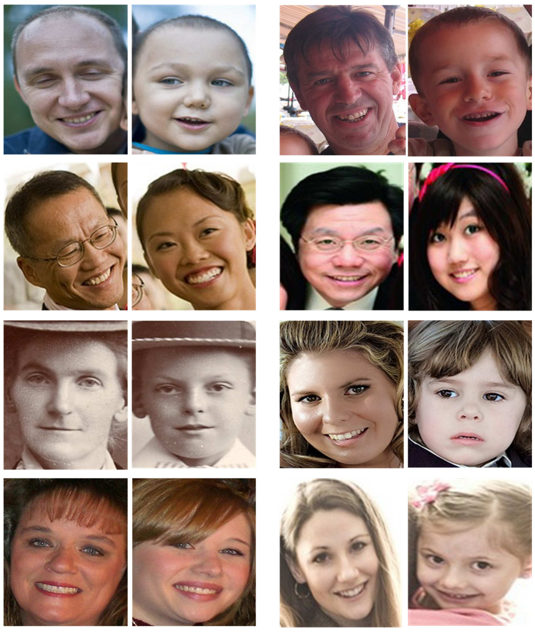
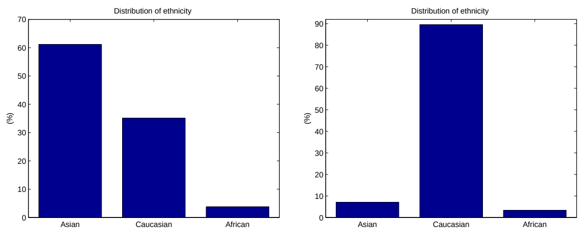


Figure 3. Several examples of the KFW-II database. From top to bottom are the Father-Son (FS), Father-Daughter (FD), Mother-Son (MS) and Mother-Daughter (MD) kinship relations, respectively.



(a) (b)

Figure 4. The ethnicity distributions of (a) KFW-I and (b) KFW-II datasets, respectively.

subset, we construct all pairs of positive (true) and negative negative (false) samples for experiments. The positive samples are the true pairs and the negative samples are each parent with the selected child from the children images who is not his/her true child.

We have experimented with several feature sets for face analysis in recent work: Local Binary Patterns (LBP) [1], LEarning-based (LE) [3], SIFT [12], and Three-Patch LBP (TPLBP) [24]. For the LBP feature, we used 256 bins rather than bins to describe each face image because we found such parameter setting achieved better performance than that used in [1]. For the LE method, we followed the parameter setting in [3] and used 200 bins to encode a histogram feature for each image. For the SIFT feature, we densely sampled and computed the SIFT descriptors of 16×16 patches over a grid with spacing of 8 pixels. For the TPLBP feature, we followed the parameter setting in [24] and used 256 bins to encode a histogram feature for each image. For details on these feature descriptors, we refer the readers to [1, 3, 12, 24].

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 yield higher verification accuracy than other kernels.

We have compared our method with three other metric learning-based face verification algorithms which could also address the kinship verification problem, including C-SML [18], NCA [8], and (LMNN) [23]. For our proposed NRML and MNRML methods, the neighborhood size k was empirically set to be 5.

4.3. Results and Analysis

Table 1 & 2 tabulate the verification rate of different methods with different features on our kinship databases. The best recognition accuracy of each method was recorded for a fair comparison. As can be seen from these two tables, the proposed NRML (MNRML) methods outperforms the other three compared methods with the lowest gains in accuracy of 1.0% (3.0%) on the F-S subset, 2.0% (3.3%) on the F-D subset, 2.0% (3.0%), 1.0% (2.0% on the M-S subset, and 1.0% (6.6% on the mean accuracy of the KFW-I dataset, 2.0% (3.1%) on the F-S subset, 2.0% (3.2%) on the F-D subset, 1.0% (1.6%), 1.0% (1.6%) on the M-S subset, and 1.2% (2.0%) on the mean accuracy of the KFW-II dataset, respectively.

We have made four observations from the results listed in Tables 1 & 2:

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

2. 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 for our kinship verification task.

3. 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 experimental results which also demonstrated that the LE methods can beat most conventional feature representations in face verification [3]. That is because the LE method has better utilized the local patch information of face images and such local

Table 1. Verification accuracy (%) of different methods on different subsets of the KFW-I dataset.

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	63.2	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 2. Verification accuracy (%) of different methods on different subsets of the KFW-II dataset.

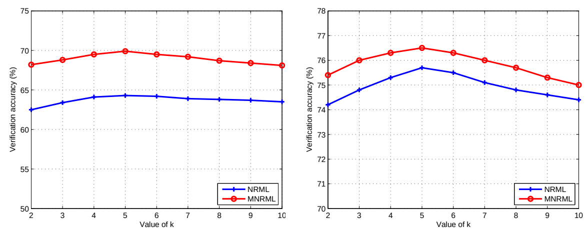
Method	Feature	F-S	F-D	M-S	M-D	Mean
CSML	LBP	66.0	65.5	64.8	65.0	65.3
	LE	71.8	68.1	73.8	74.0	71.9
	SIFT	62.0	58.9	56.8	57.4	58.8
	TPLBP	66.4	62.6	62.8	64.9	64.2
NCA	LBP	67.0	66.5	65.8	66.0	66.3
	LE	73.8	70.1	74.8	75.0	73.5
	SIFT	63.0	59.9	58.8	59.4	60.4
	TPLBP	67.4	63.6	63.8	66.9	66.5
LMNN	LBP	68.0	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**

patches may provide more discriminative information for kinship relation discovery.

4. The results obtained on the KFW-II dataset are generally higher than those obtained on the KFW-I dataset, which indicates that kinship verification on the KFW-I dataset is more difficult that that on the KFW-II dataset. The reason is that face images in the KFW-II dataset are collected from the same photo and the kinship images have the same collection conditions, which could reduce some challenges of the some factors such as illumination and aging variations in the KFW-I dataset.

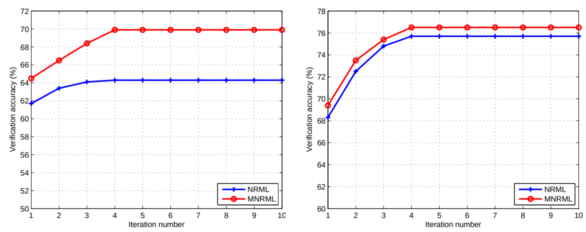
Then, we investigate the effect of the parameter k of our proposed NRML and MNRML methods. Figure 5 shows the mean kinship verification accuracies of NRML (the LE feature is applied) and MNRML (all the four features are applied) versus different values of k, where Figure 5(a) and 5(b) are the results obtained on the KFW-I and KFW-II databases, respectively. We can observe from this figure that NRML and MNRML demonstrates a stable recognition performance with versus varying neighborhood sizes. Hence it is easy to select an appropriate neighborhood sizes for NRML and MNRML to obtain good performance in real applications.

Since NRML and MNRML are iterative algorithms, we also evaluate their performance with different number of iterations. Figure 6 shows the mean verification accuracy of DMMA versus different number of iterations, where Figure 6(a) and 6(b) are the results obtained on the KFW-I and KFW-II databases, respectively. We can see from this figure that our proposed NRML and MNRML can converge to a local optimal peak in several iterations.



(a) (b)

Figure 5. Mean verification accuracy of NRML and MNRML versus different values of parameter k on the (a) KFW-I and (b) KFW-II datasets, respectively.



(a) (b)

Figure 6. Mean verification accuracy of NRML and MNRML versus different number of iterations on the (a) KFW-I and (b) KFW-II datasets, respectively.

As an important baseline, the human ability in kinship verification from facial images was also tested. From each of the above four subsets in the KFW-I and KFW-II datasets, we randomly selected 100 pairs of face samples, 50 are positive and the other 50 are negative, and presented them to 10 human observers (5 males and 5 females) with age of 20 to 30 years old. None of them received training on the task

Table 3. The classification accuracy (%) of human ability on kinship verification on different kinship subsets of the KFW-I dataset.

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 4. The classification accuracy (%) of human ability on kinship verification on different kinship subsets of the KFW-II dataset.

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

before the experiment. There are two stages in the experiment. The difference is that, in the first stage (HumanA), only the cropped face regions are shown, while, in the second stage (HumanB), the whole original color images are shown. HumanA intends to test kinship verification purely based on face, while HumanB intends to test kinship verification based on multiple cues including face, hair, skin color, and background. Note that the information provided in HumanA is the same as that provided to the algorithms. Tables 3 & 4 show the classification accuracy of human ability on kinship verification on the KFW-I and KFW-II datasets, respectively. We can observe that our proposed NRML method can obtain better performance than HumanA, and performs slightly worse than HumanB on the KFW-I dataset, 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 KFW-II dataset.

5. 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 in the wild on the largest kinship data sets. Experimental results have shown that the performance of our proposed methods are not only significantly better than that of the state-of-the-art metric learning algorithms for kinship verification, but also comparable to that of the human observers. 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, A. Hadid, et al. Face description with local binary patterns: application to face recognition. *PAMI*, 28(12):2037–2041, 2006.

[2] P. N. Belhumeur, J. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *PAMI*, 19(7):711–720, 1997.

[3] Z. Cao, Q. Yin, X. Tang, and J. Sun. Face recognition with learning-based descriptor. In *CVPR*, pages 2707–2714, 2010.

[4] J. Davis, B. Kulis, P. Jain, S. Sra, and I. Dhillon. Information-theoretic metric learning. In *ICML*, pages 209–216, 2007.

[5] R. Fang, K. Tang, N. Snavely, and T. Chen. Towards computational models of kinship verification. In *ICIP*, pages 1577–1580, 2010.

[6] B. Fasel and J. Luettin. Automatic facial expression analysis: a survey. *Pattern Recognition*, 36(1):259–275, 2003.

[7] Y. Fu, S. Yan, and T. S. Huang. Classification and feature extraction by simplexization. *IEEE Transactions on Information Forensics and Security*, 3(1):91–100.

[8] J. Goldberger, S. Roweis, G. Hinton, and R. Salakhutdinov. Neighborhood component analysis. In *NIPS*, pages 2539–2544, 2004.

[9] G. Guo, Y. Fu, C. Dyer, and T. Huang. Image-based human age estimation by manifold learning and locally adjusted robust regression. *IEEE Trans. on Image Processing*, 17(7):1178–1188, 2008.

[10] 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. on Neural Networks*, 11(4):948–960, 2002.

[11] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled faces in the wild: a database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, October 2007.

[12] D. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.

[13] H. Lu, K. Plataniotis, and A. Venetsanopoulos. MPCA: Multilinear principal component analysis of tensor objects. *IEEE Transactions on Neural Networks*, 19(1):18–39, 2008.

[14] J. Lu and Y. Tan. Ordinary preserving manifold analysis for human age and head pose estimation. *IEEE Transactions on Systems, Man, Cybernetics, Part C: Applications and Reviews*, 2012 (in press).

[15] J. Lu, Y. Tan, and G. Wang. Discriminative multi-manifold analysis for face recognition from a single training sample per person. *PAMI*, 2012 (in press).

[16] A. M. Martinez and R. Benavente. The AR face database. Technical report, CVC Technical Report, 1998.

[17] B. Moghaddam and M. Yang. Learning gender with support faces. *PAMI*, 24(5):707–711, 2002.

[18] H. Nguyen and L. Bai. Cosine similarity metric learning for face verification. *ACCV*, pages 709–720, 2011.

[19] P. Phillips, H. Moon, S. Rizvi, and P. Rauss. The FERET evaluation methodology for face-recognition algorithms. *PAMI*, 22(10):1090–1104, 2000.

[20] T. Sim, S. Baker, and M. Bsat. The cmu pose, illumination, and expression (pie) database. In *FGR*, pages 46–51, 2002.

[21] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[22] X. Wang and X. Tang. A unified framework for subspace face recognition. *PAMI*, 26(9):1222–1228, 2004.

[23] K. Weinberger, J. Blitzer, and L. Saul. Distance metric learning for large margin nearest neighbor classification. In *NIPS*, 2005.

[24] L. Wolf, T. Hassner, and Y. Taigman. Descriptor based methods in the wild. In *ECCVW*, 2008.

[25] S. Xia, M. Shao, and Y. Fu. Kinship verification through transfer learning. In *IJCAI*, pages 2539–2544, 2011.

[26] T. Xia, D. Tao, T. Mei, and Y. Zhang. Multiview spectral embedding. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 40(6):1438–1446, 2010.

[27] B. Xie, Y. Mu, D. Tao, and K. Huang. m-sne: Multiview stochastic neighbor embedding. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 41(4):1088–1096, 2011.

[28] S. Yan, J. Liu, X. Tang, and T. Huang. A parameter-free framework for general supervised subspace learning. *IEEE Transactions on Information Forensics and Security*, 2(1):69–76, 2007.

[29] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: a literature survey. *ACM Computing Surveys*, 35(4):399–458, 2003.