Mixed bi-subject kinship verification via multi-view multi-task learning

Xiaoqian Qin\textsuperscript{a,b,c}, Xiaoyang Tan\textsuperscript{a,b,*}, Songcan Chen\textsuperscript{a,b}

\textsuperscript{a} College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China
\textsuperscript{b} Collaborative Innovation Center of Novel Software Technology and Industrialization, Nanjing 210016, China
\textsuperscript{c} Huaiyin Normal University, Huai’an 223300, China

A R T I C L E   I N F O

Article history:
Received 1 December 2015
Received in revised form 7 June 2016
Accepted 9 June 2016
Communicated by Dr Jiwen Lu
Available online 16 June 2016

Keywords:
Mixed-kinship verification
Multi-task learning
Multi-view learning

A B S T R A C T

Bi-subject kinship verification addresses the problem of verifying whether there exists some kind of kin relationship (i.e., father–son, father–daughter, mother–son and mother–daughter) between a pair of parent–child subjects based purely on their visual appearance. The task is challenging due to the involvement of two different subjects possibly with different genders and ages. In addition, collecting sufficient training samples for each type of kinship is difficult. In this work, we present a novel method to address these issues by considering each type of kin relation verification as one task and learning them at one time in the framework of multi-task learning, by sharing feature sets and useful structures among the tasks.

Particularly our contributions are three folds: first, we introduce a new type of learning problem, called mixed bi-subject kinship verification, to the topic of bi-subject kinship verification: instead of simply verifying whether some fixed kinship relationship (e.g., mother–son) can be established for a given pair of parent–child images, we try to figure out whether any type of the four kinship relations can be established according to the visual features of the image pair, with no need to know the genders of the subjects to be verified beforehand. Second, we propose a novel multi-task learning method to address this problem with two transformation matrices – one is shared amongst all the tasks and the other is unique to each task. Both matrices are simultaneously learned in a joint framework, which enables our algorithm to utilize the common knowledge of the four tasks. Third, we propose a multi-view multi-task learning (MMTL) method to perform multiple feature fusion to improve the mixed bi-subject kinship verification performance. Extensive experiments on the large scale KinFaceW kinship database demonstrate the feasibility and effectiveness of the proposed algorithm.

1. Introduction

The aim of kinship verification in the field of computer vision is to predict whether a few given subjects have certain kind of kin relation purely through their visual appearance. Applications of this include face image retrieval\textsuperscript{[6,22,37]}, face annotation\textsuperscript{[28,57]}, face images organization, face recognition\textsuperscript{[36,40]}, social media analysis\textsuperscript{[45,55]}, children adoptions\textsuperscript{[18]}, and so on.

However, the problem of kinship verification is challenging, since it involves multiple subjects possibly with different genders and ages, rather than multiple images of the same subject as in the traditional face verification. On the other hand, the researches in the field of human visual signal processing\textsuperscript{[7,9]} have provided strong evidence that facial appearance is a useful cue for genetic similarity - children look more similar to their parents than other adults of the same gender. There exists some progress in finding distinguishable cues from facial appearance\textsuperscript{[11,15,42,43,27,47,53]}. While most of them focus on bi-subject kinship verification, i.e. father–son, father–daughter, mother–son and mother–daughter, Qin et al.\textsuperscript{[32]} tackle the problem of tri-subject kinship verification, i.e., father/mother–son and father/mother–daughter. However, all these works assume that the genders of the pair of subjects be known before doing verification. In other words, they can only verify one kind of kinship at one time.

In this work, we introduce a new type of learning problem, called mixed bi-subject kinship verification. That is, instead of simply verifying whether one fixed kinship relationship (e.g., mother–son) can be established for a given pair of parent–child images, we try to figure out whether any type of the four kinship relations can be established according to the visual features of the image pair, with no need to know the genders of the subjects to be...
verified beforehand. This is not only a step toward more practical kinship verification, but also imposes interesting challenges onto this topic – without knowing the genders of subjects, the problem of bi-subject kinship verification can not be simply treated as a binary classification problem. In addition, existing works show that the gender plays a significant role in kinship verification [11,27,32,47,53]. And earlier psychological research results also indicate that the kin recognition signal is different under different gender [31].

According to the studies of human genetics, we know that some facial features, such as eye or hair color or dimples have been proposed to be specifically linked to paternity confidence [34]. Hair, skin and eye colors are highly heritable and visible traits in humans [38]. Thus, we think that different kinship relations may have some common genetic characteristics. Motivated by these researches, we propose a novel multi-task learning method to address this problem with two transformation matrices – one is shared amongst all the tasks and the other is unique to each task. Both matrices are simultaneously learned in a joint framework, which enables our algorithm to utilize the common knowledge of the four tasks. Further, compared to single task learning, the advantage of multi-task learning is usually more visible when we only have a few training examples per task. We expect that even if only a few training samples are available for each task, the use of multi-task learning can compensate discriminative information by transferring from one task to the others. To make better use of multi-feature descriptors, we propose a multi-view multi-task learning (MMTL) method to perform multiple feature fusion to improve the mixed bi-subject kinship verification performance. Extensive experiments on the large KinFaceW kinship database demonstrate the feasibility and effectiveness of the proposed algorithm. The contributions of this work are summarized as follows:

1. A new type of learning problem called mixed bi-subject kinship verification is proposed to the topic of bi-subject kinship verification. We try to figure out whether any type of the four kinship relations can be established according to the visual features of the image pair, with no need to know the genders of the subjects to be verified beforehand.

2. A novel multi-task learning method is proposed to address our proposed problem with two transformation matrices – one is shared amongst all the tasks and the other is unique to each task. Both matrices are simultaneously learned in a joint framework, which enables our algorithm to utilize the common knowledge of the four tasks.

3. A multi-view multi-task learning method is developed to perform multiple feature fusion to improve the mixed bi-subject kinship verification performance.

In what follows, we briefly review some of the related work in Section 2 and detail our proposed method in Section 3. Experimental results are given in Section 4. We conclude this paper in Section 5.

2. Related work

Existing works on kinship learning can be roughly be divided into two categories: kinship classification and verification. Fang et al. [14] tackle the former one, i.e., given a query face image, asking which family it belongs to. They do this with a minimum sparse reconstruction method. Most of the works are about kinship verification. In an early attempt, Fang et al. [15] use various features including the skin, hair and eye color, facial structure measures and local/holistic texture. Later, researchers evaluate various types of feature descriptors including DAISY [18], Gradient Orientation Pyramid (GOP [55], Gated Autoencoder [10], Self Similarity Representation (SSR) [23], prototype-based discriminative feature learning (PDFL) method [48], a spatial pyramid learning-based feature descriptor Zhou et al. [54], dense stereo matching [36] and attributes [43]. Dibeklioğlu et al. [11] propose to describe facial appearance over smile expression.

In [27], the authors show that combining several types of middle-level features is useful. For that purpose, they introduce a multi-view neighborhood repulsed metric learning method (MNRLM) by learning a distance metric. Yan et al. [47] and Hu et al. [20] extract multiple features to characterize kin face images. The Politecnico di Torino (Polito) [26] group first combines different textual features and then selects the most relevant variables to provide an effective characterization of the samples. The LIRIS-University of Lyon (LIRIS) [26] group applies the Triangular Similarity Metric Learning method to make similar vectors parallel and dissimilar vectors opposite. Recently, a scalable similarity learning method [56] is proposed to learn a diagonal bilinear similarity model by online truncated gradient learning. Another way to reduce the appearance similarity gap is to use intermediate samples which bridge the two sides with large divergence. In [35,39,42,44], such a bridge is constructed by facial images of parents at the similar ages of their children. However, it is not easy to collect such an image set in practice.

While most of the above works focus on the bi-subject kinship verification, Ghahramani et al. [16] deal with the one-versus-multi kinship verification, i.e., predicting whether a query face image has kin relation with multiple family members, by fusing similarity of each member’s facial image segments. Qin et al. [32] tackle the verification of the most basic unit that forms a family, that is tri-subject kinship verification, by incorporating the prior knowledge that a child may resemble a particular parent more than the other. Zhang et al. [53] solve the group membership prediction problem, that is predicting whether or not a collection of instances share a certain semantic property, by a probability model which consists of view-specific and view-shared random variables.

This work is different from the ones mentioned above in that we focus on the mixed bi-subject kinship verification, i.e., given a pair of query face images with genders unknown, predicting whether there is a kin relationship or not between them.

Multi-task learning (MTL) refers to the joint training of multiple problems, enforcing a common intermediate parameterization or representation. If the different problems are sufficiently related, MTL can lead to better generalization and benefit all of the tasks Evgeniou and Pontil [13,21,30,2,3,33]. To learn such common structures, some authors propose to group multiple tasks into disjoint clusters [5,12,46]. Kumar and Daume III [24] allow two tasks from different groups to overlap to some degree under the concept of latent basis tasks. Instead, Goncalves et al. [17] and Zhang and Yeung [52] propose to jointly learning the task relationship structures and the task parameters without grouping them.

In this work, we propose a new multi-task learning scheme by simultaneously learning two types of structures, one is shared amongst all the tasks and the other is unique to each task. Both are simultaneously learned in a joint framework, which enables our algorithm to utilize the common knowledge of the four tasks.

3. Mixed bi-subject kinship verification

In this section, we present our method for mixed bi-subject kinship verification. For bi-subject kinship verification problem, we mainly consider four types of kin relations, i.e., father–son (F–
S), father–daughter (F-D), mother-son (M-S) and mother-daughter (M-D). For each type of relations \( t \) \((1 \leq t \leq T = 4)\), we denote its \( N_t \) samples as \( \{(x^t_{ni}, x^t_{ni}, y^t_{ni})\}_{i=1}^{N_t} \), where \( x^t_{ni}, x^t_{ni} \in \mathbb{R}^d \) respectively denotes the \( i \)-th sample of a parent and a child and \( y^t_{ni} \in \{+1, -1\} \) indicates whether this child has a valid kinship relation with the corresponding parent. Our goal is to learn a function \( f: (x^t, x^c) \rightarrow (+1, -1) \) from the training data to check whether a kinship could be established for two previously never seen images \((x^t, x^c)\), without knowing the genders of the parent and the child.

3.1. Single task bi-subject kinship verification

For single task bi-subject kinship verification, all of the four kin relations share a common transformation matrix. Let \( W \) denote the correlation between child-parent, our learning cost function takes the following form:

\[
\min_W \sum_{t=1}^{T} \sum_{i=1}^{N_t} L(y^t_{ni}, S_W(x^t_{ni}, x^c_{ni})) + \gamma R(W)
\]

(1)

where \( L(-, -) \) is the empirical loss function, \( R(W) \) is a regulariser and \( \gamma \) is a trade-off parameter balancing the effects of these two terms. In order to make \( W \) smooth we choose \( R(W) = \| W \|_F^2 \) where \( \| W \|_F \) is the Frobenius norm of matrix \( W \).

We define \( L(-, -) \equiv [a]_{+} \equiv \max\{0, a\}, a = 1 - y_i S_W(x^t_{ni}, x^c_{ni}) \). The score function \( S_W(x^t_{ni}, x^c_{ni}) \) measures the similarity between the parent and the child according to their visual appearance, which takes the following bilinear function [32],

\[
S_W(x^t, x^c) = (x^t)^T W x^c = \langle x^t, x^c \rangle_W
\]

(2)

where the transformation matrix \( W \) essentially captures the correlation between a parent and a child, to be learned from the training data. In what follows, we use the notation of STL to denote the single task learning method for mixed bi-subject kinship verification.

3.2. From single task to multi-task

Under the viewpoint of multi-task learning, each of the four types of kin relations (F-S, F-D, M-S and M-D) is treated as a separate task but are learned together. In this work, each single task is formulated as a bilinear function \( f_t(x) = \langle x^t, W_t x^c \rangle \), where \( W_t \in \mathbb{R}^{d \times d} \) is the parameter to be learnt.

Many multi-task learning methods are based on some formal definition of the notion of relatedness of the tasks. This relatedness is mostly application-dependent. For example, Hierarchical Bayes [1,4] assume that all task matrices \( W_t \) come from a particular probability distribution such as a Gaussian. This implies that all \( W_t \) are “close” to some mean function \( W_0 \). While the studies of human genetics have declared that some facial features, such as eye or hair color or dimples have been proposed to be specifically linked to paternity confidence [34]. Hair, skin and eye colors are highly inheritable and visible traits in humans [38]. Thus, we think that different kinship relations may have some common genetic characteristics. Motivated by these researches, we decide to model each task in two parts,

\[
W_t = W_0 + W_t
\]

(3)

where the transformation matrix \( W_t \) is unique to each task, while all the tasks share a common structure \( W_0 \).

To learn these two structures, we solve the following optimization problem:

\[
\min_{W_0, W_t} \sum_{t=1}^{T} \sum_{i=1}^{N_t} \max\{0, 1 - y_i(x^t_{ni})^T (W_0 + W_t x^c_{ni}) + \frac{\lambda}{2} \| W_t \|_F^2 + \frac{\eta}{2} \| W_0 \|_F^2\}
\]

(4)

where \( \lambda \) and \( \eta \) are positive regularization parameters. Intuitively, for a fixed value of \( \eta \), a large value of the ratio \( \nu = \frac{\eta}{\lambda} \) will characterize the scale of the final model. Particularly, when \( \lambda \) tends to infinity the problem reduces to solving one single-task learning problem as in Eq. (1). On the other hand, when \( \eta \) tends to infinity the problem reduces to solving the \( T \) tasks independently.

The objective function \( G(W_0, W_t, \lambda, \eta) \) is convex with respect to \( W_0 \) and \( \{W_t\} \). To obtain \( W_0 \) and \( \{W_t\} \) we optimize Eq. (4) using stochastic gradient descent method, which is commonly used to minimize risk function. The gradients of \( W_0 \) and \( \{W_t\} \) are calculated respectively as follows:

\[
\nabla W_0 = \nabla_{W_0} L(\cdot, \cdot) + \eta W_0
\]

(5)

\[
\nabla W_t = \nabla_{W_t} L(\cdot, \cdot) + \lambda W_t
\]

(5)

where \( \nabla_{W_0} L(\cdot, \cdot) \) and \( \nabla_{W_t} L(\cdot, \cdot) \) are the subgradients of \( L \) with respect to \( W_0 \) and \( W_t \) respectively, which can be computed by:

\[
\nabla_{W_0} L(\cdot, \cdot) = \left\{ \begin{array}{ll}
0 & , S_{W_t}(x^t_{ni}, x^c_{ni}) > 1 \\
-y_i x^t_{ni} x^c_{ni} & , S_{W_t}(x^t_{ni}, x^c_{ni}) < 1 \\
\mu_s(-y_i x^t_{ni} x^c_{ni}) & , S_{W_t}(x^t_{ni}, x^c_{ni}) = 1 \\
0 & , S_{W_t}(x^t_{ni}, x^c_{ni}) > 1 \\
-y_i x^t_{ni} x^c_{ni} & , S_{W_t}(x^t_{ni}, x^c_{ni}) < 1 \\
\mu_s(-y_i x^t_{ni} x^c_{ni}) & , S_{W_t}(x^t_{ni}, x^c_{ni}) = 1 \\
\end{array} \right.
\]

(5)

We first optimize the objective function with respect to \( \{W_t\} \) when \( W_0 \) is fixed, and then optimize it with respect to \( W_0 \) when \( \{W_t\} \) is fixed. This procedure is repeated until convergence. We set \( \mu = 0 \) and the initial value of \( W_0 \) and \( \{W_t\} \) to 1 which corresponds to the assumption that all the crossed positions of \((x^t, x^c)\) are unrelated.

3.3. Mixed kinship verification

When the transformation matrices \( W_0 \) and \( \{W_t\} \) are obtained from the training samples with genders known, it is very straightforward to make a decision for given query images with the corresponding \( \{W_t\} \) for \( t \)-th task as in multi-task learning. However, in the mixed kinship verification, the genders of the query images are unknown, which means that we don’t know which task to perform in the first place.

To address this, in our experiments the verification function \( f(x^t, x^c) \) is modeled as the linear combination of four pieces of evidence, i.e., the confidence of \((x^t, x^c)\) under each task, respectively

\[
f(x^t, x^c) = \sum_{t=1}^{T} \beta_t S_{W_t}(x^t, x^c) + b
\]

(6)

where the combination coefficients \( \beta_t \) are four scalars and \( b \) is the similarity threshold term. To learn these parameters, we maximize the conditional likelihood through a sigmoid function, i.e., \( p(y = 1|x^t, x^c) = \sigma(f(x^t, x^c)) \), with \( L_2 \) regularization added where sigmoid function \( \sigma \) is defined to be \( \sigma(x) = \frac{1}{1+e^{-x}} \).

3.4. Multi-view multi-task learning for mixed kinship verification

Previous face verification/recognition, hand detection and palm vein recognition studies have shown that multiple feature
descriptors could provide complementary information from different viewpoints [29,49–51], and hence it is desirable to utilize multiple feature information for our mixed kinship verification task. However, multiple feature descriptors generally have multiple modalities. Our MTL method is designed for single feature representation and can not be used to deal with multi features directly. A nature solution for MTL with multi-feature is to directly concatenate different features into a new feature vector and then apply MTL for verification learning. However, such an operation is not desirable and it sacrifices the diversity of different descriptors. To address this problem, we propose a new multi-view MTL (MMTL) method to seek multiple transformation matrices to perform multiple features fusion.

Assume there are $V$ views of feature representations, and $(x_{i}^{p}, x_{i}^{c})_{i=1}^{N}$ denotes the $v$-th view set of $N$ pairs of kinship images from the $t$-th task, where $x_{i}^{p} \in R^{d}$ and $x_{i}^{c} \in R^{d}$ are the $i$-th parent and child images from the $v$-th view, respectively, and $v = 1, 2, ..., V, t = 1, 2, ..., T$. The aim of MMTL is to learn multiple transformation matrices $W_{t}^{p}, W_{t}^{c}, ..., W_{t}^{v}$ and $(W_{1}, W_{2}, ... W_{v})$ so that the property of MTL in all feature spaces can be jointly exploited.

In order to achieve this goal, we impose a nonnegative weighted vector $\beta = [\beta_{1}, \beta_{2}, ..., \beta_{V}]$ for each feature and formulate MMTL as follows:

$$\min_{W_{t}^{p}, W_{t}^{c}, \beta_{1}, ..., \beta_{V}} \sum_{v=1}^{V} \beta_{v} G_{v}(W_{t}^{p}, W_{t}^{c}, x^{p}, x^{c})$$

subject to $\sum_{v=1}^{V} \beta_{v} = 1$, $\beta_{v} \geq 0$ \hspace{1cm} (7)

where $G_{v}(W_{t}^{p}, W_{t}^{c}, x^{p}, x^{c})$ is the object function of MTL for the $v$-th feature, which is computed according to Eq. (4).

The trivial solution of Eq. (7) is $\beta_{v} = 1$, which corresponds to the minimum $G_{v}(W_{t}^{p}, W_{t}^{c}, x^{p}, x^{c})$ and $\beta_{v} = 0$ otherwise. This means that only the best view is selected by our method, such that the complementary information of different features has not been exploited. To address this problem, we modify $\beta_{v}$ to be $\beta_{v}^{\tau} (u > 0)$, and the new objective function for MMTL is defined as:

$$\min_{W_{t}^{p}, W_{t}^{c}, \beta_{1}, ..., \beta_{V}} \sum_{v=1}^{V} \beta_{v}^{\tau} G_{v}(W_{t}^{p}, W_{t}^{c}, x^{p}, x^{c})$$

subject to $\sum_{v=1}^{V} \beta_{v}^{\tau} = 1$, $\beta_{v}^{\tau} \geq 0$ \hspace{1cm} (8)

To our best knowledge, there is no closed-form solution to Eq. (8) and there are $V_{v}(T + 1)$ matrices to be optimized simultaneously, we solve it in an iterative manner.

First, we fix $\beta$ and update $W$. When $\beta$ is fixed, the optimization problem in Eq. (8) can be rewritten as:

$$\min_{W_{t}^{p}, W_{t}^{c}, \beta_{1}, ..., \beta_{V}} \sum_{v=1}^{V} \beta_{v}^{\tau} G_{v}(W_{t}^{p}, W_{t}^{c}, x^{p}, x^{c})$$

subject to $\sum_{v=1}^{V} \beta_{v}^{\tau} = 1$, $\beta_{v}^{\tau} \geq 0$ \hspace{1cm} (9)

We sequentially optimize $(W_{t}^{p}, W_{t}^{c})$ with the fixed $(W_{t}^{p}, W_{t}^{c})$, $(W_{t}^{p}, W_{t}^{c})$, ..., $(W_{t}^{p}, W_{t}^{c})$, $(W_{t}^{p}, W_{t}^{c})$, ..., $(W_{t}^{p}, W_{t}^{c})$.

Then, Eq. (9) can be rewritten as:

$$\min_{W_{t}^{p}, W_{t}^{c}, \beta_{1}, ..., \beta_{V}} \sum_{v=1}^{V} \beta_{v}^{\tau} G_{v}(W_{t}^{p}, W_{t}^{c}, x^{p}, x^{c})$$

subject to $\sum_{v=1}^{V} \beta_{v}^{\tau} = 1$, $\beta_{v}^{\tau} \geq 0$ \hspace{1cm} (10)

Now, $(W_{t}^{p}, W_{t}^{c})$ can be solved as that we deal with Eq. (4).

Then, we update $\beta$ with the fixed $W$. We construct the following Lagrange function:

$$T(\beta, \xi) = \sum_{v=1}^{V} \beta_{v}^{\tau} G_{v}(W_{t}^{p}, W_{t}^{c}, x^{p}, x^{c}) - \xi (\sum_{v=1}^{V} \beta_{v}^{\tau} - 1)$$

$$\min_{\beta_{1}, ..., \beta_{V}} T(\beta, \xi)$$

subject to $\sum_{v=1}^{V} \beta_{v}^{\tau} = 1$, $\beta_{v}^{\tau} \geq 0$ \hspace{1cm} (11)

Let $\frac{\partial T(\beta, \xi)}{\partial \beta_{v}^{\tau}} = 0$ and $\frac{\partial T(\beta, \xi)}{\partial \xi} = 0$, we have

$$\sum_{v=1}^{V} \beta_{v}^{\tau} G_{v}(W_{t}^{p}, W_{t}^{c}, x^{p}, x^{c}) - \xi (\sum_{v=1}^{V} \beta_{v}^{\tau} - 1) = 0$$

$$\sum_{v=1}^{V} \beta_{v}^{\tau} - 1 = 0$$

Combining Eqs. (12) and (13), we can obtain $\beta_{v}$ as follows:

$$\beta_{v} = \frac{1}{\sum_{v=1}^{V} (1/|G_{v}(W_{t}^{p}, W_{t}^{c}, x^{p}, x^{c})|^{1/(r-1)})}$$

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

Algorithm 1. MMTL

Input:
Training images: $(x_{i}^{p}, x_{i}^{c})_{i=1}^{N}$ be the $v$-th view set of $N_{v}$ pairs of kinship images from the $t$-th task;
Parameters: regularization term $\lambda^{u}$, $\eta^{v}$, $u$, iteration number $Nu$, and convergence error $r$

Output:
Symmetric transformation matrix $(W_{t}^{p}, W_{t}^{c}, ..., W_{t}^{v})$ and weights $\beta_{1}, \beta_{2}, ..., \beta_{V}$.

1: Step1: Initialization:
2: Set $\beta_{v} = 1/N_{v}$, $W_{t}^{p} = I$, $W_{t}^{c} = I$, $t = 1$, $r = 1$, $..., T$, $v = 1, 2, ..., V$.
3: Calculate the loss $L^{v}(\cdot, \cdot)$
4: Step2: Local optimization:
5: For $r = 1, 2, ..., Nu$, repeat
6: Compute $(W_{t}^{p}, W_{t}^{c})$ in Eq. (10);
7: Obtain $\beta_{v}$ by using Eq. (14); 8: Calculate the loss $L^{v}(\cdot, \cdot)$;
9: if $r > 2$ and $L^{v} - L^{v-1} < r$, go to Step3;
10: Step3: Output transformation matrices and weighting vector:
11: output $(W_{t}^{p}, W_{t}^{c}, ..., W_{t}^{v})$, $(W_{t}^{p}, W_{t}^{c}, ..., W_{t}^{v})$ and weights $\beta_{1}, \beta_{2}, ..., \beta_{V}$.

4. Experimental results and analysis

To validate the effectiveness of the proposed approach we conducted experiments on the KinFaceW-I [27] and KinFaceW-II [27] database.

4.1. Database and experimental settings

The KinFaceW-I database [27] consists of 156 FS (Father–Son), 134 FD (Father–Daughter), 116 MS (Mother–Son) and 127 MD (Mother–Daughter) pairs, while the KinFaceW-II [27] contains 250 pairs of these bi-subject kin relations each. The major difference between KinFaceW-I and KinFaceW-II lies in that each pair of faces in KinFaceW-I comes from the same photo while from different photos in KinFaceW-II. Fig. 1 gives some examples of this database.

For each face image, we extracted three different features: (1) SIFT: We sampled SIFT descriptors on each 16 × 16 patch with a grid spacing of 8 pixels. Then, we concatenated these SIFT descriptors to form a 6272-dimensional feature vector. (2) LBP: We divided each image into 4 × 4 non-overlapping blocks, where the size of each block is 16 × 16. Then, we extracted a 256-dimensional uniform pattern LBP feature for each block and concatenated them to form a 4096-dimensional feature vector. (3) HOG: We divided
each image into \(8 \times 8\) non-overlapping blocks with the size of \(2 \times 2\) pixels. Then, we extracted a 9-dimensional HOG feature for each block and concatenated them to form a 576-dimensional vector. For these three features, we apply PCA to reduce each feature into 300 dimensions to remove some noise components.

We evaluated our proposed methods with the image-restricted setting and followed the evaluation protocol as proposed in Lu et al. [26] which includes pre-specified training/testing split and independently for 5-fold cross validation. For face images in each fold, we consider all pairs of face images with the four kinship relations. Hence, the positive samples are the total true pairs of face images and the negative samples are the total false pairs of images from the four kin relations.

Note that for the parameters \((\lambda, \eta, u, Nu\) and \(\tau)\) of our method, we use 5-cross validation to set the value from a range of \((0.001, 0.01)\) for \(\lambda\), \((1 \times 10^{-4}; 10; 1 \times 10^{-3}; 10; 1 \times 10^{-2}; 10; 1 \times 10^{1})\) for \(u\) and we set \(Nu=200\), \(\tau = 0.1\).

\[\text{4.2. Results and analysis} \]

\[\text{4.2.1. Baseline results} \]

We design some baselines. One is concatenating the feature vectors of two visual entities, learning a linear SVM for each kin relation, then aggregating these meta-decisions through linear SVM for the final verification judgement. We denote this method as ‘concatenated + SVM’. Generally, the nonlinear SVM might get better results than the linear one, so we use a Gaussian kernel SVM for each kin relation and denote this method as ‘concatenated + kernel SVM’. The third one is ‘STL’ as in Eq. (1).

\[
\begin{array}{|c|c|c|}
\hline
\text{Method} & \text{Features} & \text{KinFaceW-I} & \text{KinFaceW-II} \\
\hline
\text{STL} & \text{SIFT} & 72.7 \pm 0.2597 & 74.3 \pm 0.3543 \\
\text{CMTL (proposed)} & \text{LBP} & 51.5 \pm 0.2729 & 52.3 \pm 0.5196 \\
\text{Concatenated + SVM} & \text{HOG} & 72.7 \pm 0.4213 & 74.3 \pm 0.3543 \\
\text{Concatenated + kernel SVM} & \text{SIFT} & 51.5 \pm 0.2773 & 53.4 \pm 0.3194 \\
\text{CMTL (proposed)} & \text{LBP} & 52.8 \pm 0.6684 & 53.2 \pm 0.9839 \\
\text{STL} & \text{HOG} & 55.7 \pm 0.2694 & 59.7 \pm 0.3070 \\
\text{CMTL (proposed)} & \text{SIFT} & 61.3 \pm 0.5422 & 63.1 \pm 0.4563 \\
\text{STL} & \text{LBP} & 65.7 \pm 0.4343 & 67.7 \pm 0.6319 \\
\text{CMTL (proposed)} & \text{HOG} & 66.0 \pm 0.3651 & 68.8 \pm 0.4213 \\
\text{STL} & \text{SIFT} & 67.2 \pm 0.3144 & 71.1 \pm 0.4281 \\
\text{CMTL (proposed)} & \text{SIFT} & 70.1 \pm 0.4522 & 72.6 \pm 0.7012 \\
\hline
\end{array}
\]

\[\text{Table 1 Correct verification rates (%) for different methods on the KinFaceW-I and KinFaceW-II database.}\]

that MTL utilizes the common knowledge of different kin relations as supplementary information to facilitate decision making while STL has not effectively utilized this information.

To further demonstrate the effect of our MTL method, we visualize the similarity distribution over 4000 positive and 4000 negative pairs from the KinFaceW-II database for ‘STL’ and MTL methods in Fig. 2. One can see that, the similarity of kinship and non-kinship pairs overlaps heavily in the learned space through ‘STL’ method in Fig. 2(a). However, these two similarity distribution histograms become more separate in the learned space through MTL method in Fig. 2(b), so that more discriminative information is exploited.

\[\text{4.2.2. Proposed vs. different feature fusion strategies} \]

We compare our method with three other different feature fusion strategies: (1) Single MTL (MTL): we learn a single model by using our MTL method with each single feature representation. (2) Concatenated MTL (CMTL): we first concatenate different features into a longer feature vector and then learn a single model by using our MTL method with the augmented feature representation. (3) Equal weight MTL (EMTL): we learn the model for each feature representation by our MTL method and then use the equal weight to compute the similarity of two face images.

\[\text{Table 2 summarizes the results. One can see that CMTL strategy obtains the worst result among the three multi-view strategies.}\]
MTL method does not assume any grouping structure, and tasks are allowed to overlap with each other through sharing one common matrix. This is a more realistic assumption since we can have tasks in our pool that are not related enough still share some information that can be exploited for better learning.

4.2.4. Proposed vs. state-of-the-art

To the best of our knowledge, there are very few works that tackle the mixed bi-subject kinship verification problem, and it is very difficult to find an existing method directly comparable to ours.

We have compared our methods with three existing state-of-the-art separate kinship verification models, i.e., neighborhood repulsed metric learning (NRML) [27], Polito [26] and LIRIS [26].

Furthermore, considering that the similarity modeling is related to metric learning, we also include two classical metric learning algorithms, i.e., Information-theoretic metric learning (ITML) [8] and large margin nearest neighbor classification (LMNN) [41] as the base models.

Although they deal with a different problem, the image set based face verification bears some similarities to the problem of kinship verification from the respect of methodology, i.e., both involve similarity matching between multiple faces. Hence in this work, we also adopt one of the best performers on the YouTube Face database, i.e., DDML (Discriminative Deep Metric Learning) [19], to score the pairwise similarity between a child and his/her parent. Particularly, we train a deep metric learning network with three layers using our own implementation, with the threshold $r$, the learning rate $\mu$ and regularization parameter $\lambda$ set to be $3, 10^{-3}, 10^{-2}$, respectively.

For all these methods we learn a transformation matrix for each feature, our MMTL can obtain the best result.

The reason may be that directly concatenate different features into a new longer feature vector may sacrifice the diversity of different descriptors. EMTL strategy can further improve the performance over MTL method with SIFT descriptor which indicates that different feature descriptors can provide complementary information. With the learned nonnegative weighted vector for each feature, our MMTL can obtain the best result.

4.2.3. Proposed vs. existing multi-task learning

We carry out empirical comparisons with following two competing subspace regularized multi-task learning approaches: (1) No-group MTL [25]: all tasks are assumed to be related, and the task parameters are assumed to lie in a low dimensional subspace. This is done by penalizing the $L_{q,1}$-norm of weight matrix. (2) GO-MTL [24]: a recently proposed approach that assumes each task parameter vector is a linear combination of a finite number of underlying basis tasks.

Table 3 summarizes the results. One can see that all the multi-task learning approaches are able to outperform single task learning (STL) method, indicating that common information relevant to prediction can be shared among these tasks and learning them jointly can result in better generalization performance than independently learning each task.

By incorporating over-lapping group structure, GO-MTL method improves the performance about 2.4% and 1.0% on the KinFaceW-I and KinFaceW-II database which indicates that the GO-MTL method is effective. MTL method can improve the performance about 1.1% and 1.3% on the KinFaceW-I and KinFaceW-II database. The reason is that

<table>
<thead>
<tr>
<th>Method</th>
<th>KinFaceW-I</th>
<th>KinFaceW-II</th>
</tr>
</thead>
<tbody>
<tr>
<td>STL</td>
<td>67.2 ± 0.3144</td>
<td>71.1 ± 0.4281</td>
</tr>
<tr>
<td>No-group MTL [25]</td>
<td>69.2 ± 0.3749</td>
<td>72.0 ± 0.3683</td>
</tr>
<tr>
<td>GO-MTL [24]</td>
<td>71.6 ± 0.3479</td>
<td>73.0 ± 0.3215</td>
</tr>
<tr>
<td>MTL (proposed)</td>
<td>72.7 ± 0.2597</td>
<td>74.3 ± 0.3543</td>
</tr>
</tbody>
</table>

MTL method does not assume any grouping structure, and tasks are allowed to overlap with each other through sharing one common matrix. This is a more realistic assumption since we can have tasks in our pool that are not related enough still share some information that can be exploited for better learning.

For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.
Table 4 Correct verification rates (%) for different methods on the KinFaceW-I and KinFaceW-II database.

<table>
<thead>
<tr>
<th>Method</th>
<th>KinFaceW-I</th>
<th>KinFaceW-II</th>
</tr>
</thead>
<tbody>
<tr>
<td>DDML [19]</td>
<td>64.4 ± 0.5596</td>
<td>66.3 ± 0.4000</td>
</tr>
<tr>
<td>ITML [8]</td>
<td>64.2 ± 0.3188</td>
<td>67.0 ± 0.5165</td>
</tr>
<tr>
<td>LMNN [41]</td>
<td>67.4 ± 0.3856</td>
<td>67.9 ± 0.0975</td>
</tr>
<tr>
<td>NRML [27]</td>
<td>62.7 ± 0.3872</td>
<td>67.6 ± 0.4222</td>
</tr>
<tr>
<td>Polito [28]</td>
<td>71.1 ± 0.8273</td>
<td>71.8 ± 0.7365</td>
</tr>
<tr>
<td>LIRIS [26]</td>
<td>72.1 ± 0.2823</td>
<td>72.5 ± 0.5635</td>
</tr>
<tr>
<td>MTL (proposed)</td>
<td>72.7 ± 0.2597</td>
<td>74.3 ± 0.3543</td>
</tr>
<tr>
<td>MMTL (proposed)</td>
<td>73.7 ± 0.3800</td>
<td>77.2 ± 0.5751</td>
</tr>
</tbody>
</table>

Table 5 Correct verification rates(%) for different methods with a gender detector as a pre-processing on the KinFaceW-I and KinFaceW-II database.

<table>
<thead>
<tr>
<th>Method</th>
<th>KinFaceW-I</th>
<th>KinFaceW-II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concatenated + SVM</td>
<td>51.4 ± 0.5184</td>
<td>51.8 ± 0.3571</td>
</tr>
<tr>
<td>NRML Lu et al. [27]</td>
<td>61.2 ± 0.7236</td>
<td>67.5 ± 0.5123</td>
</tr>
<tr>
<td>Polito Lu et al. [26]</td>
<td>69.8 ± 0.5334</td>
<td>70.4 ± 0.6934</td>
</tr>
<tr>
<td>LIRIS Lu et al. [26]</td>
<td>71.8 ± 0.3236</td>
<td>71.8 ± 0.4430</td>
</tr>
<tr>
<td>DDML Hu et al. [19]</td>
<td>63.3 ± 0.2917</td>
<td>66.9 ± 0.4314</td>
</tr>
<tr>
<td>ITML Davis et al. [8]</td>
<td>65.9 ± 0.2815</td>
<td>66.0 ± 0.4393</td>
</tr>
<tr>
<td>LMNN Weinberger et al. [41]</td>
<td>66.6 ± 0.4209</td>
<td>67.0 ± 0.2031</td>
</tr>
<tr>
<td>MTL (proposed)</td>
<td>72.7 ± 0.2597</td>
<td>74.3 ± 0.3543</td>
</tr>
</tbody>
</table>

SVM for the final verification judgement. And we tune the parameters of all the compared methods to reach the best possible performance.

Table 4 summarizes the results. One can see that MTL consistently outperforms the other compared methods. This is partly due to the fact that the compared methods fail to model the relation among the two visual entities with mixed kin relationship. By contrast, ‘MTL’ utilizes the common knowledge of different kin relations as supplementary information to facilitate decision making and achieves better verification performance.

MMTL can improve the verification performance of MTL. The reason is MMTL can make use of multiple feature representations such that some complementary information can be utilized for our mixed kinship verification task.

4.2.5. Proposed vs. gender detector pre-processing

For mixed kinship verification, one can predict the genders of the given query images through a stand-alone gender detector and then make a decision for the image pair with the corresponding bi-subject kinship verification model. Motivated by this, we compared our MTL method with all the compared methods with a gender detector as a pre-processing with our own implementation.

Table 5 summarizes the results. It can be seen that MTL can obtain comparable or even better performance than the other compared methods. This is because we emphasize two transformation matrices—one is shared amongst all the tasks and the other is unique to each task while others have not effectively utilized this information. Another reason is that although the gender detector can achieve high performance, one must notice that only when the given query two images are correctly predicted, the pair of images can be predicted through the corresponding kin relation verification model, otherwise, the given query two images may be decided through a mismatched kin relation verification model which could not exploit this kin relation during training process.

4.2.6. Parameter analysis

We investigate the effect of the ratio $\nu = \frac{\lambda}{\eta}$, where we let $\lambda$ fix and $\eta$ vary to see the effects. Fig. 3 shows the mean verification accuracy of MTL as a function of the value of $\nu$. One can see that when $\nu$ is large we get a relatively low performance, indicating that the four kin relations should not be processed without distinction. Moreover, for KinFaceW-I the best result is obtained when $\nu = 1 \times 10^{-3}$ while for KinFaceW-II $\nu = 1 \times 10^{-1}$, hence it appears that the KinFaceW-II is close to being from a single task.

5. Conclusions

In this work, we make the first attempt to investigate the mixed bi-subject kinship verification problem. Instead of predicting whether there is a kin relation between two given entities with genders known, we try to answer this question without knowing such information. We then propose a new multi-task learning scheme to address this problem. The key idea of our method is to decompose each single task model into two parts—one is shared amongst all the tasks and the other corresponds to the task-specific structure. The model can be efficiently solved using the stochastic gradient descent method. Extensive experiments on the large scale KinFaceW kinship database show that the proposed methods significantly improve the performance of the traditional kinship verification methods. How to explore more discriminative features and combine them with our proposed MTL and MMTL methods to further improve the verification performance appears to be another interesting direction of future work.

Acknowledgments

We thank the anonymous reviewers for your in-depth comments, suggestions, and corrections, which have greatly improved the manuscript. This work is partially supported by National Science Foundation of China (61373060, 61472186), Qing Lan Project and the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (13KJD520002).

References


[38] X. Zhou, H. Yan, Y. Shang, Kinship verification from facial images by scalable similarity fusion, Neurocomputing (2016).