Multiple Anthropological Fisher Kernel Framework And its Application to Kinship Verification

Ajit Puthenputhussery Qingfeng Liu Chengjun Liu
New Jersey Institute of Technology
avp38, ql69, cliu@njit.edu

Abstract

This paper presents a novel multiple anthropological Fisher kernel (MAFK) framework for kinship verification. The proposed MAFK framework, which goes beyond the Mahalanobis distance metric learning, integrates multiple anthropology inspired features and derives semantically meaningful similarities between images. The major novelty of this paper comes from the following three aspects. First, three new anthropology inspired features (AIF) are derived by extracting the AIF-SIFT, AIF-WLD and AIF-DAISY features on images that are enhanced by an anthropology inspired similarity enhancement method extended from the SIFT flow method. Second, a novel multiple anthropological Fisher kernel framework (MAFK) is proposed which combines multiple features and their metrics between images in a unified paradigm. The MAFK is optimized as a constrained, non-negative, and weighted variant of the sparse representation problem regularized by the criterion of pushing away the nearby non-kinship samples and pulling close the kinship samples. Third, a novel normalized kernel similarity measure (NKSM) is proposed by normalizing the MAFK with the fractional power transformation and L2 normalization. The feasibility of the proposed MAFK framework is assessed on two representative kinship data sets, namely the KinFaceW-I and the KinFaceW-II data sets. The experimental results show the effectiveness of the proposed method.

1. Introduction

Kinship verification has been an important topic in anthropology for many years. Pioneer work in anthropology [25], [1], [3] believe that there are some genetic related features which are inherited by children from their parents that can be used to determine the kinship relations. Recently, kinship verification from facial images is gaining increasing attention as an emerging research area in artificial intelligence [9], [34], [24], [7], [21], [37], [20], [27].

Many feature methods have been proposed for describing facial images[16], [18], [33], [4], [5], [28]. These features, which are designed specifically for distinguishing one image from others (the discriminative ability), cannot guarantee that a child image is more similar to its parent image than to other images (the inheritable ability). The major reason is that these features are designed for recognition of face image and thus cannot characterize the genetic relations between kinship images. Another reason is that the inherent similarity gap between kinship images is much larger than that in the face recognition problem, e.g. LFW [12], which means similarity between discriminative features is not sufficient for kinship verification. Lu et al. [24] proposed to apply a metric learning method on several features and proposed the MNRML method by combining different metrics on different features and subsequent work [36], [37], [22] followed their pipeline by combining features and metric learning methods sequentially. In their methods, the features and the metric learning methods are developed in different paradigms independently, which may attenuate the effect when they are combined. Besides, most metric learning methods are based on Mahalanobis distance metric, which may not achieve the best performance in some scenarios.

To address these issues, this paper proposes a novel multiple anthropological Fisher kernel (MAFK) framework for kinship verification. The proposed method derives a semantically meaningful similarity between images by combining multiple anthropology inspired features and their metrics in a unified paradigm. Specifically, three novel anthropology inspired features (AIF) are first extracted, namely the AIF-SIFT, AIF-WLD and AIF-DAISY features. The process of deriving the anthropology inspired features consists of an anthropology inspired similarity enhancement method and the extraction of opponent color SIFT [15], color WLD-SIFT and DAISY [29] descriptors based on the enhanced image. In particular, the similarity enhancement method is applied to kinship image pairs by extending the SIFT flow method [19] and generating the enhanced images by reinforcing similar facial parts. The opponent SIFT descriptor,

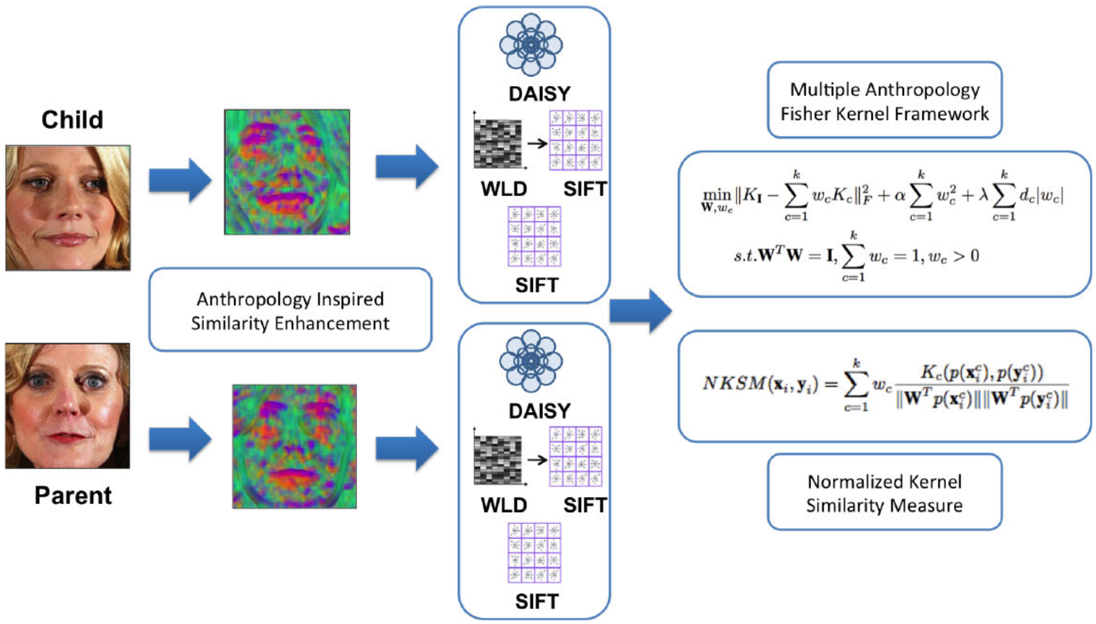

Figure 1. The framework of the proposed MAFK method.

the color WLD-SIFT descriptor and the DAISY descriptor are then extracted from the enhanced images. Second, a novel MAFK framework is derived by learning a new metric and weights for multiple features in a unified paradigm. In particular, the new metric is learned while fixing the weights by balancing the behavior of pushing away the k-nearest non-kinship samples while pulling close the kinship ones for each training pair. The weights are updated while fixing the transformation. Finally, a normalized multiple similarity measure is proposed based on the observation that the fractional power transformation is able to transform the data into a near Gaussian shape with a stable variance. This is well suited for dot product based similarity measure like the cosine similarity measure from the point of view of the Bayes decision rule for minimum error [17]. The proposed MAFK method is then evaluated on two challenging kinship databases, KinFaceW-I and KinFaceW-II data set [24] and the experimental results show the feasibility of the proposed method.

2. Related Work

Facial images convey important characteristics such as identity information, kinship information, facial expressions, gender of a person, ethnicity, emotional information, mental state of a person and so on. Among these many characteristics, kinship is believed to be one of the most dominant one since children naturally inherit genetic features from their parents [34]. The work [9] by Fang et al. shows the feasibility of applying computer vision techniques for kinship verification. Xia et al. [34] proposed a transfer subspace learning based algorithm by using the young parents set as an intermediate set to reduce the significant divergence in the appearance distributions between the facial images of parents and their children. Lu et al. [24] proposed the neighborhood repulsed metric learning (NRML) method in which the intraclass samples within a kinship relation are pulled as close as possible and interclass samples are pushed as far as possible for kinship verification.

Dehghan et al. [7] proposed to apply the generative and the discriminative gated autoencoders to learn the genetic features and metrics together for kinship verification. Yan et al. [36] proposed a multimetric learning method to combine different complementary feature descriptors for kinship verification and later [37] proposed to learn the discriminative mid-level features by constructing a reference data set instead of using hand-crafted descriptors. Lu et al. [22] presented the results of various teams on the FG 2015 Kinship Verification in the Wild challenge.

Metric learning methods have gained a lot of attention for computer vision and machine learning applications. Earlier work by Xing et al. [35] applied the semi-definite programming to learn a Mahalanobis metric. Goldberger et al. [10] proposed the neighborhood component analysis (NCA) by minimizing the cross validation error of the kNN classifier. Weinberger et al. proposed the large margin nearest neighbor (LMNN) [32] method which uses hinge loss to encourage the related neighbors to be atleast one distance unit closer than points from other classes. Davis et al. proposed the information-theoretic metric learning (ITML) [6] method to learn a class of distance functions. Hieu and Li [26] proposed the cosine similarity metric learning (CSML) method which utilizes the favorable properties of

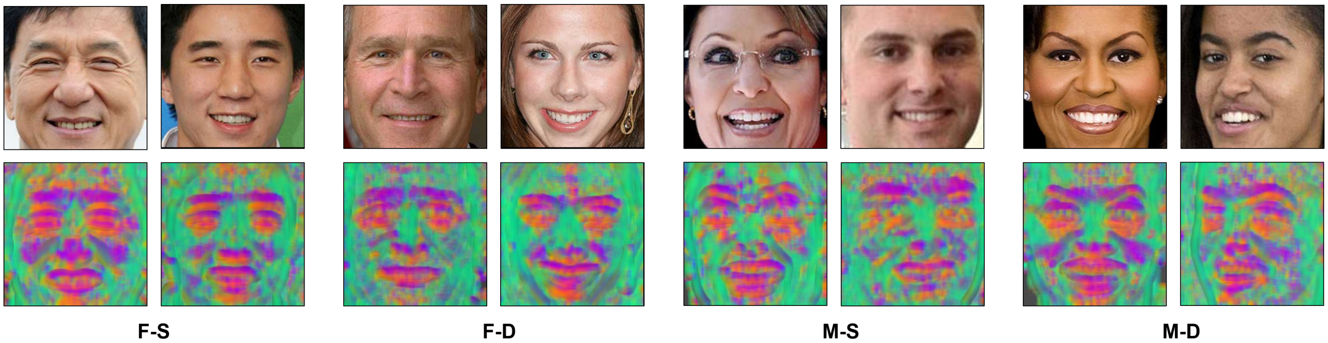

Figure 2. Visualization of SIFT images of different kinship relations using the top three principal components of SIFT descriptors extracted from the image. The purple and orange regions in the visualization highlight the inheritable genetic feature regions in the kinship images.

cosine similarity. Lu et al. [24] proposed the neighborhood repulsed metric learning (NRML) method for kinship verification which pays more attention to the neighborhood samples. Lu et al. [23] proposed the discriminative deep metric learning (DDML) method that trains a deep neurnal network for learning a discriminative set of hierarchical transformations to project the face pairs in a discriminative subspace. A large-margin multi-metric learning method was proposed by Hu et al. [11] which jointly learns global distance metrics to maximize the correlations of different feature representations of each sample. A marginalized denoising metric learning method was proposed by Wang et al. [31] to explicitly preserve the intrinsic structure of data and increase discrimination of the learned features.

3. Anthropology Inspired Feature Extraction

Naini et al. [25] analyzed the contributions of heredity and environment on external facial features. Their anthropological results [25] show that eyes, chin and parts of the forehead show higher visual resemblance between parents and their offspring and provide large feedback. From the computer vision point of view, these high resemblance in facial regions between kinship image pairs exhibit three important properties described as follows:

Given the notations that $\mathbf{p} = (x, y)$ is the grid coordinate of the image, $\mathbf{d}(\mathbf{p}) = (u(\mathbf{p}), v(\mathbf{p}))$ is the displacement vector at $\mathbf{p}$, $u(\mathbf{p})$ and $v(\mathbf{p})$ are two integers that represent the displacements along the x and y axes from the coordinates $\mathbf{p}$, respectively. Let s_1 and s_2 be the two dense SIFT descriptors to be measured, and ε represent the set of all the spatial neighborhoods.

- First, these facial regions between kinship image pairs have high visual resemblance (e.g. their eyes resemble each other), which means their local descriptors are similar, namely $\|s_1(\mathbf{p}) - s_2(\mathbf{p} + \mathbf{d}(\mathbf{p}))\|$ is small.
- Second, these facial regions should be at similar relative locations on two faces (e.g. their eyes appear at similar locations on two faces), which means there may be a small displacement between the centers of two local descriptors, namely $\|\mathbf{d}(\mathbf{p})\|$ is small.
- Third, the neighborhood regions of high resemblance facial regions tend to be similar (e.g. the neighborhood small regions around the center of eyes tend to be smoothly changed), which means $\|\mathbf{d}(\mathbf{p}) - \mathbf{d}(\mathbf{q})\|$ is small where $(\mathbf{p}, \mathbf{q}) \in \varepsilon$.

Inspired by these anthropological observations, we propose three novel anthropology inspired features to capture these high resemblance facial regions between parents and their children. First, we present a new anthropology inspired similarity enhancement (AISE) method by extending the SIFT flow [19] method from the scene alignment to kinship image pairs. The SIFT flow algorithm matches densely sampled SIFT features and finds correspondence estimated by SIFT flow. The objective function for SIFT flow [19] is defined as follows:

$$\begin{aligned} E(\mathbf{d}) = & \sum_{\mathbf{p}} (\|s_1(\mathbf{p}) - s_2(\mathbf{p} + \mathbf{d}(\mathbf{p}))\|_1) + \\ & \sum_{\mathbf{p}} \eta(\|\mathbf{d}(\mathbf{p})\|_1) + \sum_{\mathbf{p},\mathbf{q} \in \varepsilon} \theta(\|\mathbf{d}(\mathbf{p}) - \mathbf{d}(\mathbf{q})\|_1) \end{aligned} \tag{1}$$

where η is the set of neighbors and θ is the weight of the third term. The SIFT flow method satisfies the three properties of high visual resemblance facial regions between kinship pairs. It is therefore suitable to be extended to kinship image pairs for capturing the inheritable information between parents and children. The estimated SIFT flow is applied to reinforce the high visual resemblance facial regions and generate similarity enhanced images.

To visualize the effectiveness of our method, the top three principal components of the SIFT descriptors of the image are mapped to the principal components of the RGB space, as shown in figure 2. The purple and the orange regions in the visualization highlight the high visual resemblance regions in the kinship images. It can be discovered that these regions focus on eyes, mouth, chin and parts of the forehead. Therefore our proposed AISE method derives interesting phenomena that are consistent to the anthropology results in [25]. Other interesting patterns can also be deduced for different relations from figure 2. It can be observed that the father-son and mother-daughter relation show large visual correspondence in different parts of facial regions leading to the deduction that individuals of the same gender in kinship relations share higher visual resemblance. It can also be seen that mother-daughter relation has higher genetic responses compared to father-daughter relation confirming the observation that mothers resemble their daughters more as in [1].

The SIFT, WLD and DAISY descriptors are then extracted from the similarity enhanced images derived by our anthropology inspired similarity enhancement method. Therefore we name these three anthropology inspired features as AIF-SIFT, AIF-WLD and AIF-DAISY. In particular, the AIF-SIFT feature is computed in the opponent color space [15] of the enhanced image. We then derive densely sampled SIFT features from the image encoded by the Weber local descriptors (WLD) and the process is repeated separately for the three components of the image resulting in color AIF-WLD feature. To improve the robustness against photometric and geometric transformations of the enhanced image, dense AIF-DAISY descriptors are also computed.

4. Multiple Anthropological Fisher Kernel Framework

The complementary nature of the discriminative and generative approaches leads to the generative score space. One example is the Fisher score [13], which has been widely applied for visual classification problems such as face recognition [28], object recognition [14]. In this section, we extend the Fisher score from the classification problem to the metric learning problem. Particularly, let $\mathbf{X}_i = \{\mathbf{l}_t, t = 1, 2, ..., T\}$ be the set of T local descriptors extracted from an image of the $i-$th pair. Similarly, we define $\mathbf{Y}_i$ for the other image of the $i-$th pair. Let $p(\mathbf{X}|\boldsymbol{\lambda})$ be the probability density function of generating $\mathbf{X}_i$ or $\mathbf{Y}_i$ with a set of parameters $\boldsymbol{\lambda}$, then the Fisher score is defined as follows:

$$\mathbf{F}(\mathbf{X}_i) = \frac{1}{T} \bigtriangledown_{\boldsymbol{\lambda}} \log[p(\mathbf{X}_i|\boldsymbol{\lambda})] \tag{2}$$

As a matter of fact, the Fisher score is the gradient vector of the log-likelihood that describes the contribution of the parameters to the generation process. It therefore provides information about the generative perspective of the features. Based on the Fisher score, a score space based similarity measure, namely the Fisher kernel [13], is derived as

$$K_F(\mathbf{X}_i, \mathbf{Y}_i) = (\mathbf{F}(\mathbf{X}_i))^T \mathbf{I}^{-1} \mathbf{F}(\mathbf{Y}_i) \tag{3}$$

where $\mathbf{I}$ is the Fisher information matrix. The conventional Fisher kernel provides a natural similarity measure between images by considering the underlying probability distribution. However, three major issues inherent to the conventional Fisher kernel are still waiting for solutions. First, the conventional Fisher kernel fails to take into account the label information. Second, the Fisher information matrix $\mathbf{I}$ is difficult to obtain and approximation techniques are not sufficient to guarantee performance. Third, it only measures the similarity of a single aspect between images, which depends on the type of the local image descriptors.

Therefore, this paper presents a novel multiple anthropological Fisher kernel framework (MAFK) to address these three issues. The MAFK learns a new distance metric that captures the pairwise information, and the weights of multiple distance metrics that exploits information from different features. Specifically, the score space based multiple distance metric is defined as follows with the weights $w_c (c = 1, 2, ..., k)$:

$$\begin{aligned} D(\mathbf{X}_i, \mathbf{Y}_i) &= \sum_{c=1}^{k} w_c D_c(\mathbf{X}_i^c, \mathbf{Y}_i^c) \\ &= \sum_{c=1}^{k} w_c (\mathbf{p}_i^c)^T \mathbf{M} (\mathbf{c}_i^c) \\ &= \sum_{c=1}^{k} w_c (\mathbf{p}_i^c)^T \mathbf{W}\mathbf{W}^T (\mathbf{c}_i^c) \\ &= \sum_{c=1}^{k} w_c (\mathbf{x}_i^c)^T (\mathbf{y}_i^c) \end{aligned} \tag{4}$$

where $\mathbf{p}_i^c = \mathbf{F}(\mathbf{X}_i)$, $\mathbf{c}_i^c = \mathbf{F}(\mathbf{Y}_i)$, $\mathbf{x}_i^c = \mathbf{W}^T\mathbf{p}_i^c$ and $\mathbf{y}_i^c = \mathbf{W}^T\mathbf{c}_i^c$ $(i = 1, 2, ..., m)$. It is easy to see that matrix $\mathbf{M} = \mathbf{W}\mathbf{W}^T$ is symmetric and positive definite. To keep the notation simple, we use $D(\mathbf{x}_i, \mathbf{y}_i)$ instead of $D(\mathbf{X}_i, \mathbf{Y}_i)$ in the remaining parts of the paper, where $\mathbf{x}_i$ is the vector that represents the image derived from the feature descriptors in $\mathbf{X}_i$. The introduction of $\mathbf{W}$ alleviates the assumptions on the Fisher information matrix since $\mathbf{W}$ can be learned from the training data and contains sufficient information for recognizing kinship relations.

The derivation of $\mathbf{W}$ and w_c consists of two iterative procedures. Let $\mathbf{D} = \{(\mathbf{x}_i^c, \mathbf{y}_i^c) | \mathbf{x}_i^c, \mathbf{y}_i^c \in \mathbb{R}^{n\times 1} (i = 1, 2, ..., m, c = 1, 2, ..., k)\}$ where k is the number of the type of features and m is the number of training samples. The main purpose of the transformation $\mathbf{W}$ and weights w_c is to push away the nearby non-kinship samples as far as possible while pulling the kinship relation samples as close as possible, and approximate the ideal similarity matrix. In other words, the distance between $\mathbf{x}_i^c$ and $\mathbf{y}_i^c$ should be as small as possible if $\mathbf{x}_i^c$ and $\mathbf{y}_i^c$ have kinship relations and the distance should be as large as possible otherwise. Therefore, the objective function for the MAFK method can be formulated as follows.

$$\begin{aligned} &\min_{\mathbf{W}, w_c} \|D_{\mathbf{I}} - \sum_{c=1}^{k} w_c D_c\|_F^2 + \alpha \sum_{c=1}^{k} w_c^2 + \lambda \sum_{c=1}^{k} d_c |w_c| \\ &\quad s.t. \mathbf{W}^T\mathbf{W} = \mathbf{I}, \sum_{c=1}^{k} w_c = 1, w_c > 0 \end{aligned} \tag{5}$$

In the above objective function, the first and second term show the reconstruction criterion and the regularization for the weights of different metrics. The third term represents the criterion of pushing away the nearby non-kinship samples as far as possible while pulling the kinship samples as close as possible. d_c is defined as follows:

$$\begin{aligned} d_c &= \sum_{i=1}^{m} 2 * D_c(\mathbf{x}_i^c, \mathbf{y}_i^c) - D_c(\mathbf{x}_i^c, (\mathbf{y}_i^c)^*) - D_c((\mathbf{x}_i^c)^*, \mathbf{y}_i^c) \\ &= \mathrm{Tr}\,(\mathbf{W}^T(2\mathbf{M}_1^c - \mathbf{M}_2^c - \mathbf{M}_3^c)\mathbf{W}) \end{aligned} \tag{6}$$

where $\mathbf{M}_1^c = \sum_{i=1}^{m} \mathbf{p}_i^c(\mathbf{c}_i^c)^T$, $(\mathbf{x}_i^c)^*$ is the nearest neighbor of $\mathbf{x}_i^c$, $(\mathbf{y}_i^c)^*$ is the nearest neighbor of $\mathbf{y}_i^c$, $D_c \in \mathbb{R}^{m\times m}$ is the similarity matrix for the c-th feature ($c = 1, 2, ..., k$) and $D_{\mathbf{I}} \in \mathbb{R}^{m\times m}$ is the ideal similarity matrix which is derived by multiplying the scaled label vector (0.5 for scaling in our experiment) with its transpose. Note that $\mathbf{M}_1^c$ is not symmetric and we make it symmetric by using $\mathbf{M}_1^c = (\mathbf{M}_1^c + (\mathbf{M}_1^c)^T)/2$ without influencing the value of d_c. $\mathbf{M}_2^c$ and $\mathbf{M}_3^c$ are computed in a similar way.

Now the problem becomes a constrained, non-negative, and weighted variant of the sparse representation problem. The term $\sum_{c=1}^{k} d_c|w_c|$ that corresponds to the criterion of pushing away the nearby non-kinship samples and pulling close the kinship samples behaves as a regularization for the multiple metric learning problem. The objective function in equation 5 is optimized using an iterative procedure. Specifically, given the fixed w_c, we approximately update $\mathbf{W}$ by discarding the reconstruction criterion and optimizing the following objective function:

$$\begin{aligned} &\max_{\mathbf{W}} \mathrm{Tr}(\mathbf{W}^T \sum_{c=1}^{k} w_c(\mathbf{M}_2^c + \mathbf{M}_3^c - 2\mathbf{M}_1^c)\mathbf{W}) \\ &s.t. \mathbf{W}^T\mathbf{W} = \mathbf{I} \end{aligned} \tag{7}$$

This can be done by deriving the eigenvectors of the matrix $\sum_{c=1}^{k} w_c(\mathbf{M}_2^c + \mathbf{M}_3^c - 2\mathbf{M}_1^c)$.

Then given the $\mathbf{W}$, we optimize the following problem to derive w_c:

$$\begin{aligned} &\min_{w_c} \|D_{\mathbf{I}} - \sum_{c=1}^{k} w_c D_c\|_F^2 + \alpha \sum_{c=1}^{k} w_c^2 + \lambda \sum_{c=1}^{k} d_c |w_c| \\ &\quad s.t. \sum_{c=1}^{k} w_c = 1, w_c > 0 \end{aligned} \tag{8}$$

We apply the FISTA algorithm [2] to optimize the objective function defined in equation 8. The structure of the FISTA algorithm remains the same but the proximal operator is different since our method is a constrained, nonnegative, and weighted variation of the sparse representation problem. We thus replace the original soft thresholding operator in the FISTA algorithm with an efficient projection operator [8] considering the non-negative constraint. In order to further improve the efficiency of the optimization process, we transform the objective function defined in equation 8 into a quadratic programming problem by using the fact that $\lambda \sum_{c=1}^{k} d_c|w_c| = \lambda \sum_{c=1}^{k} d_c w_c$ since $w_c > 0$.

After the MAFK is derived, a novel normalized multiple similarity measure (NMSM) is further proposed, where the MAFK is normalized as follows

$$NMSM(\mathbf{x}_i, \mathbf{y}_i) = \sum_{c=1}^{k} w_c \frac{D_c(g(\mathbf{x}_i^c), g(\mathbf{y}_i^c))}{\|\mathbf{W}^T g(\mathbf{x}_i^c)\| \|\mathbf{W}^T g(\mathbf{y}_i^c)\|} \tag{9}$$

where $g(\mathbf{x})$ is the power transformation defined as $g(\mathbf{x}) = sign(\mathbf{x})|\mathbf{x}|^{\beta}$, β ($0 < \beta < 1$) is the power parameter, and both the power and the sign operations are element-wise.

The proposed NMSM takes advantage of normalization through fractional power transformation and the L_2 normalization. The fractional power transformation is able to transform from the data into a near Gaussian shape with a stable variance [14], [30]. With the help of the L_2 normalization, it can be proved that the NMSM is proportional to a weighted linear combination of the whitened cosine similarity measure for each feature. This shows its theoretical roots to the Bayes decision rule for minimum error [17] under some conditions such as the multivariate Gaussian distribution assumption, therefore, provides theoretical guarantee to achieve better performance.

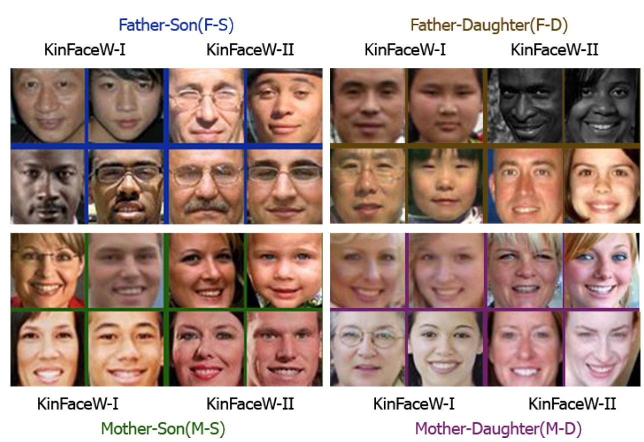

Figure 3. Example images from the KinFaceW-I and KinFaceW-II data set

Table 1. Comparison between the MAFK and other methods on the KinFaceW-I data set

Methods	F-S	F-D	M-S	M-D	Mean
CSML [26]	61.10	58.10	60.90	70.00	62.50
NCA [10]	62.10	57.10	61.90	69.00	62.30
LMNN [32]	63.10	58.10	62.90	70.00	63.30
NRML [24]	64.10	59.10	63.90	71.00	64.30
MNRML [24]	72.50	66.50	66.20	72.00	69.90
ITML [6]	75.30	64.30	69.30	76.00	71.20
DMML [36]	74.50	69.50	69.50	75.50	72.25
MPDFL [37]	73.50	67.50	66.10	73.10	70.10
GGA [7]	70.50	70.00	67.20	74.30	70.50
DGA [7]	76.40	72.50	71.90	77.30	74.50
Polito [22]	85.30	**85.80**	**87.50**	86.70	86.30
LIRIS [22]	83.04	80.63	82.30	84.98	82.74
NUAA [22]	86.25	80.64	81.03	83.93	82.96
DDMML [23]	86.40	79.10	81.40	87.00	83.50
MAFK	**88.15**	85.22	82.41	**90.95**	**86.68**

Table 2. Comparison between the MAFK and other methods on the KinFaceW-II data set

Methods	F-S	F-D	M-S	M-D	Mean
CSML [26]	71.80	68.10	73.80	74.00	71.90
NCA [10]	73.80	70.10	74.80	75.00	73.50
LMNN [32]	74.80	71.10	75.80	76.00	74.50
NRML [24]	76.80	73.10	76.80	77.00	75.70
MNRML [24]	76.90	74.30	77.40	77.60	76.50
ITML [6]	69.10	67.00	65.60	68.30	67.50
DMML [36]	78.50	76.50	78.50	79.50	78.25
MPDFL [37]	77.30	74.70	77.80	78.00	77.00
GGA [7]	81.80	74.30	80.50	80.80	79.40
DGA [7]	83.90	76.70	83.40	84.80	82.20
Polito [22]	84.00	82.20	84.80	81.20	83.10
LIRIS [22]	89.40	83.60	86.20	85.00	86.05
NUAA [22]	84.40	81.60	82.80	81.60	82.50
DDMML [23]	87.40	83.80	83.20	83.00	84.30
MAFK	**91.40**	**87.20**	**90.80**	**89.80**	**89.80**

5. Experiments

This section evaluates the effectiveness of our proposed method on two challenging kinship databases: the KinFaceW-I data set and the KinFaceW-II data set [24]. These two data sets contain images for four kinship relations, namely, father-son (F-S), father-daughter (F-D), mother-son (M-S), and mother-daughter (M-D). The KinFaceW-I data set has 156, 134, 116, and 127 image pairs for each relation respectively, whereas, the KinFaceW-II data set has 250 image pairs for each kinship relation. Example images are shown in figure 3. The parameters for the MAFK method are selected based on a grid search with cross validation approach. In our experiments, we follow the experimental protocol as defined in [24], [22] to have a fair comparison with other methods.

5.1. Implementation Details

The AISE method is first applied to derive the similarity enhanced images. Second, we derive the AIF-DAISY feature and the AIF-WLD feature on the similarity enhanced images. The dense color SIFT feature is derived with a step size of 1 and five scale patch sizes as 2, 4, 6, 8, 10. Then, the dimensionality of the opponent color SIFT feature is further reduced to 64 by PCA. The spatial information [28] is also added to the SIFT feature increasing the dimension to 66. The AIF-WLD and the AIF-DAISY features are computed similarly. For the AIF-DAISY feature, the dimensionality is reduced from 200 to 66 by PCA. Afterwards, a Gaussian mixture model with 256 components is estimated for the Fisher score computation. Then the score space based multiple metric learning is learned from the data with the parameters $\alpha = 1$ and $\lambda = 0.1$ for both the KinFaceW-I data set and the KinFaceW-II data set. The normalized multiple similarity measure with $\beta = 0.5$ is applied. Finally a two class support vector machine is used to determine the kinship relations between images.

5.2. Comparison with popular learning methods

The experimental results in table 1 and table 2 show that our method is able to achieve better performance compared to other learning methods. Note that our method can improve upon other learning methods that use multiple features, such as MNRML [24], DMML [36], DDMML [23], Polito [22], LIRIS [22] and MPDFL[37]. The MNRML [24] method uses multiple facial feature representations in a common learned distance metric so that complementary information can be used to improve the verification performance. The DMML [36] method simultaneously learns multiple distance metrics using local features such as LBP, SPLE and SIFT to extract different and complementary information from each face image. The DDMML [23] method jointly learns multiple neural networks so that the correlation of different features of each sample is maximized. The Polito [22] group proposed a multi-perspective approach to kinship verification by applying a multi-step feature selection process on three features namely LPQ, TBLBP and WLD. Similarly, the LIRIS [22] group applied the triangular similarity metric learning method (TSML) on four different face descriptors LBP, HOG, OCLBP and FV. The multiview prototype-based discriminative feature learning (MPDFL) method learns a common coefficient matrix using multiple low-level descriptors such as LBP, SPLE and SIFT for mid-level feature representation.

Table 3. Evaluation of the effectiveness of the anthropology inspired features (AIF-SIFT, AIF-WLD and AIF-DAISY) on the KinFaceW-I and the KinFaceW-II data sets.

KinFaceW-I	F-S	F-D	M-S	M-D	Mean
SIFT	73.41	69.02	66.40	79.56	72.09
WLD	73.35	65.69	70.69	71.70	70.36
DAISY	71.79	65.68	66.34	75.96	69.94
AIF-SIFT	75.61	72.75	75.04	85.87	77.32
AIF-WLD	85.27	78.40	79.40	84.18	82.22
AIF-DAISY	80.75	81.40	77.60	84.18	80.98
MAFK	**88.15**	**85.22**	**82.41**	**90.95**	**86.68**
KinFaceW-II	F-S	F-D	M-S	M-D	Mean
SIFT	80.40	70.20	79.80	80.00	77.60
WLD	68.80	62.00	63.20	65.00	64.75
DAISY	76.40	69.80	71.00	70.60	71.95
AIF-SIFT	88.20	82.00	87.80	85.20	85.80
AIF-WLD	75.40	71.60	73.00	77.00	74.25
AIF-DAISY	87.80	85.00	89.20	86.00	87.00
MAFK	**91.40**	**87.20**	**90.80**	**89.80**	**89.80**

The second observation is that our method often achieves better results on F-S and M-D kinship relations than F-D and M-S kinship relations, which is consistent to the anthropological results [1]. The reason is that the similarity variation between images of different gender is larger than the images of the same gender, and our proposed MAFK method is able to capture such a variation by learning the new transformation and the weights of multiple features.

The third observation is that our proposed MAFK method achieves more improvement on the KinFaceW-II data set compared to the KinFaceW-I data set. The reason is due to the availability of more training samples in the KinFaceW-II data set.

To better visualize the difference between our proposed MAFK method and other learning methods, the ROC curves of different methods on the KinFaceW-I and KinFaceW-II data sets are shown in figures 4 and 5. It can be seen that the ROC curves of the proposed method are higher compared to other multi-metric learning methods.

5.3. Evaluation of the Anthropology Inspired Features

This section assesses the effectiveness of the anthropology inspired features (AIF). Note that three anthropology inspired features (AIF-SIFT, AIF-WLD and AIF-DAISY features) are evaluated separately first (simply assign the weight of the feature set to 1, and others to 0). Similarly, without applying the AISE method for deriving similarity enhanced images, we do the same for SIFT, WLD and DAISY features separately to obtain the results. The experimental results in table 3 show that the performance of the anthropology inspired features (AIF-SIFT, AIF-WLD and AIF-DAISY features) derived from the enhanced images using the AISE method significantly improve the performance of SIFT, WLD and DAISY features without applying the AISE method.

6. Conclusion

This paper presents a novel multiple anthropological Fisher kernel framework (MAFK) for kinship verification. First, three new anthropology inspired features are extracted, namely the AIF-SIFT, AIF-WLD and AIF-DAISY features. Second, a multiple anthropological Fisher kernel framework is proposed to combine multiple features and their metrics between images in a unified paradigm by iteratively learning a new transformation and the weights. Third, a novel normalized multiple similarity measure is presented for effective normalization. Experimental results show that the proposed method is able to achieve better results compared to other popular methods for kinship verification.

Acknowledgements

This work is partially supported by the NSF award 1647170.

References

[1] A. Alvergne, C. Faurie, and M. Raymond. Differential facial resemblance of young children to their parents: who do children look like more? *Evolution and Human Behavior*, 28(2):135 – 144, 2007.

[2] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sciences*, 2(1):183–202, 2009.

[3] P. Bressan and M. Dal Martello. Talis pater, talis filius: Perceived resemblance and the belief in genetic relatedness. *Psychological Science*, 13(3):213–218, 2002.

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

[5] D. Chen, X. Cao, F. Wen, and J. Sun. Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification. In *CVPR*, pages 3025–3032. IEEE, 2013.

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

[7] A. Dehghan, E. G. Ortiz, R. Villegas, and M. Shah. Who do i look like? determining parent-offspring resemblance via gated autoencoders. In *CVPR*, pages 1757–1764. IEEE, 2014.

[8] J. Duchi, S. Shalev-Shwartz, Y. Singer, and T. Chandra. Efficient projections onto the l1-ball for learning in high dimensions. In *ICML*, pages 272–279, 2008.

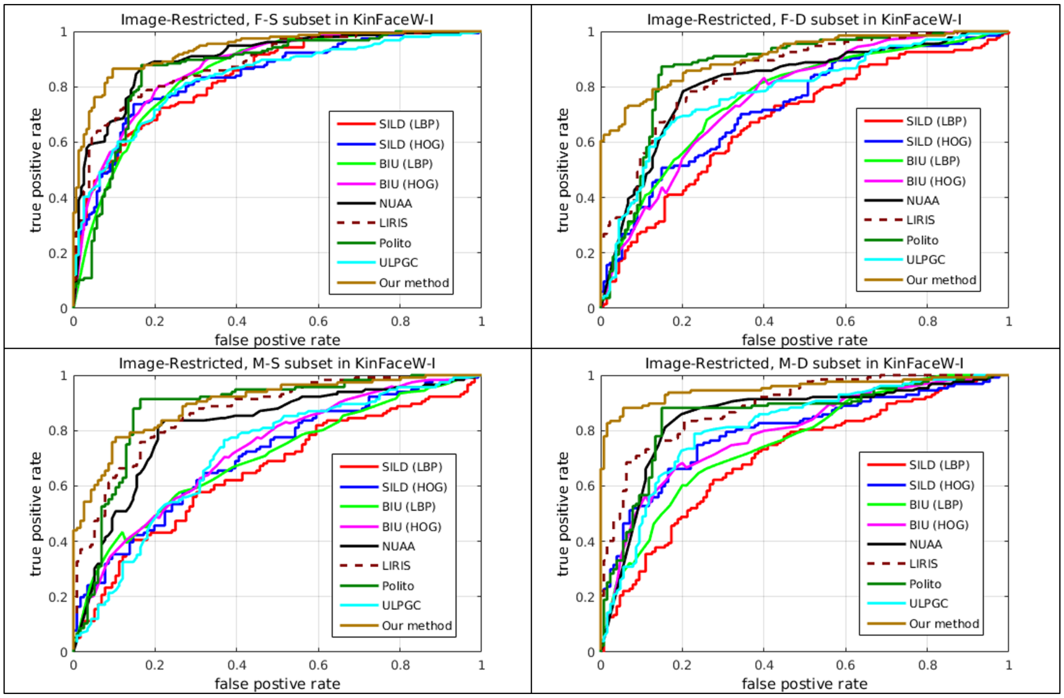

Figure 4. The ROC curve for the proposed MAFK method and other learning methods on the KinFaceW-I data set.

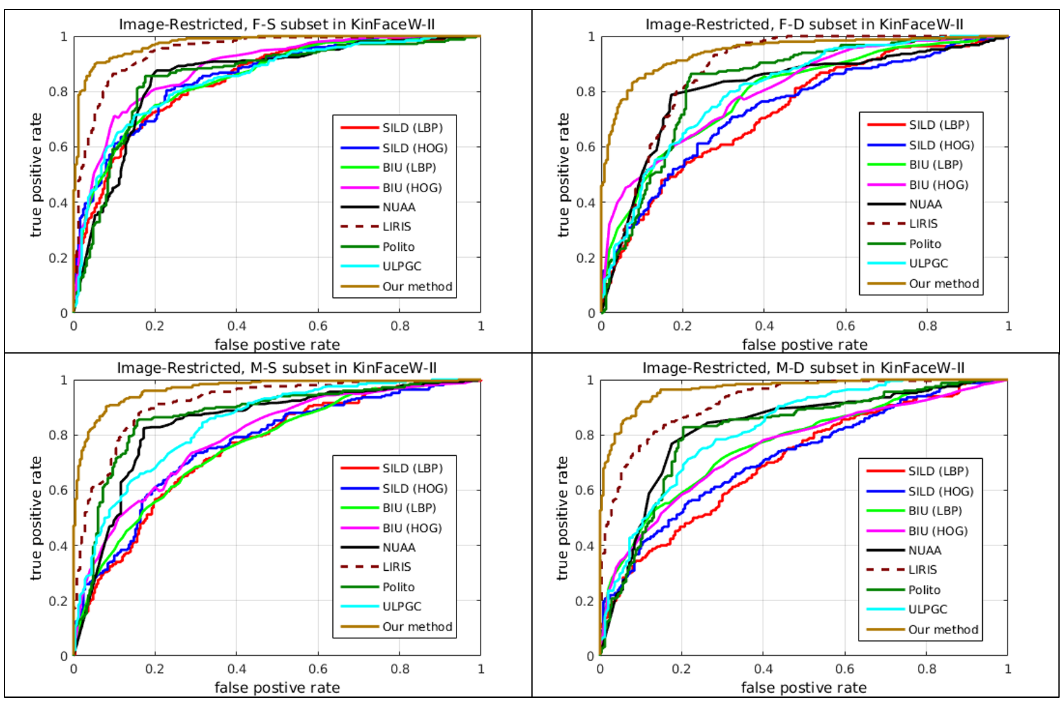

Figure 5. The ROC curve for the proposed MAFK method and other learning methods on the KinFaceW-II data set.

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

[10] J. Goldberger, S. T. Roweis, G. E. Hinton, and R. Salakhutdinov. Neighbourhood components analysis. In *NIPS*, 2004.

[11] J. Hu, J. Lu, Y. P. Tan, J. Yuan, and J. Zhou. Local large-margin multi-metric learning for face and kinship verification. *IEEE Transactions on Circuits and Systems for Video Technology*, PP(99):1–1, 2017.

[12] 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.

[13] T. Jaakkola and D. Haussler. Exploiting generative models in discriminative classifiers. In *NIPS*, pages 487–493, 1998.

[14] H. Jegou, F. Perronnin, M. Douze, J. Sanchez, P. Perez, and C. Schmid. Aggregating local image descriptors into compact codes. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(9):1704–1716, Sept 2012.

[15] F. S. Khan, R. M. Anwer, J. van de Weijer, A. D. Bagdanov, A. M. Lopez, and M. Felsberg. Coloring action recognition in still images. *International journal of computer vision*, 105(3):205–221, 2013.

[16] C. Liu. Gabor-based kernel pca with fractional power polynomial models for face recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(5):572–581, May 2004.

[17] C. Liu. The bayes decision rule induced similarity measures. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 29(6):1086–1090, 2007.

[18] C. Liu and H. Wechsler. Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. In *IEEE Transactions on Image Processing*, pages 467–476, 2002.

[19] C. Liu, J. Yuen, and A. Torralba. SIFT flow: Dense correspondence across scenes and its applications. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(5):978–994, 2011.

[20] Q. Liu, A. Puthenputhussery, and C. Liu. Inheritable fisher vector feature for kinship verification. In *2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–6, Sept 2015.

[21] Q. Liu, A. Puthenputhussery, and C. Liu. A novel inheritable color space with application to kinship verification. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1–9, March 2016.

[22] J. Lu, J. Hu, V. E. Liong, X. Zhou, A. Bottino, I. U. Islam, T. F. Vieira, X. Qin, X. Tan, S. Chen, Y. Keller, S. Mahpod, L. Zheng, K. Idrissi, C. Garcia, S. Duffner, A. Baskurt, M. Castrillon-Santana, and J. Lorenzo-Navarro. The FG 2015 Kinship Verification in the Wild Evaluation. In *FG 2015*, pages 1–7, May 2015.

[23] J. Lu, J. Hu, and Y. P. Tan. Discriminative deep metric learning for face and kinship verification. *IEEE Transactions on Image Processing*, 26(9):4269–4282, Sept 2017.

[24] J. Lu, X. Zhou, Y.-P. Tan, Y. Shang, and J. Zhou. Neighborhood repulsed metric learning for kinship verification. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 36(2):331–345, 2014.

[25] F. B. Naini and J. P. Moss. Three-dimensional assessment of the relative contribution of genetics and environment to various facial parameters with the twin method. *American Journal of Orthodontics and Dentofacial Orthopedics*, 126(6):655 – 665, 2004.

[26] H. Nguyen and L. Bai. Cosine similarity metric learning for face verification. In *ACCV*, volume 6493, pages 709–720, 2011.

[27] A. Puthenputhussery, Q. Liu, and C. Liu. Sift flow based genetic fisher vector feature for kinship verification. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 2921–2925, Sept 2016.

[28] K. Simonyan, O. M. Parkhi, A. Vedaldi, and A. Zisserman. Fisher Vector Faces in the Wild. In *BMVC*, 2013.

[29] E. Tola, V. Lepetit, and P. Fua. Daisy: An efficient dense descriptor applied to wide-baseline stereo. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(5):815–830, May 2010.

[30] R. D. Wand M.P., Marron J.S. Transformations in density estimation. *Journal of the American Statistical Association*, 86(414):343–353, 1991.

[31] S. Wang, J. P. Robinson, and Y. Fu. Kinship verification on families in the wild with marginalized denoising metric learning. In *2017 12th IEEE International Conference on Automatic Face Gesture Recognition (FG 2017)*, pages 216–221, May 2017.

[32] K. Q. Weinberger and L. K. Saul. Distance metric learning for large margin nearest neighbor classification. *J. Mach. Learn. Res.*, 10:207–244, 2009.

[33] L. Wolf, T. Hassner, and Y. Taigman. Descriptor based methods in the wild. In *Workshop on Faces in'Real-Life'Images: Detection, Alignment, and Recognition*, 2008.

[34] S. Xia, M. Shao, J. Luo, and Y. Fu. Understanding kin relationships in a photo. *Multimedia, IEEE Transactions on*, 14(4):1046–1056, Aug 2012.

[35] E. P. Xing, M. I. Jordan, S. Russell, and A. Y. Ng. Distance metric learning with application to clustering with side-information. In *NIPS*, pages 505–512, 2002.

[36] H. Yan, J. Lu, W. Deng, and X. Zhou. Discriminative multimetric learning for kinship verification. *IEEE Transactions on Information Forensics and Security*, 9(7):1169–1178, 2014.

[37] H. Yan, J. Lu, and X. Zhou. Prototype-based discriminative feature learning for kinship verification. *IEEE Transactions on Cybernetics*, 2015.