2018 13th IEEE International Conference on Automatic Face & Gesture Recognition

Kinship Classification through Latent Adaptive Subspace

Yue Wu[*], Zhengming Ding[*], Hongfu Liu, Joseph Robinson, Yun Fu
Department of Electrical & Computer Engineering, Northeastern University
College of Computer & Information Science, Northeastern University
Boston, MA, USA
{*yuewu,allanding,yunfu*}@*ece.neu.edu*, {*liu.hongf, robinson.jo*}@*husky.neu.edu*

***Abstract*—We tackle the challenging kinship classification problem. Different from kinship verification, which tells two persons have certain kinship relation or not, kinship classification aims to identify the family that a person belongs to. Beyond age and appearance gap across parents and children, the difficulties of kinship classification lie in that any data of the children to be classified are unavailable in advance to help training. To handle this challenge, an auxiliary database with complete parents and children modalities is employed to uncover the parent-children latent knowledge. Specifically, we propose a Latent Adaptive Subspace learning (LAS) to uncover the shared knowledge between two modalities so that the unseen test children are implicitly modeled as latent factors for kinship classification. Moreover, person-wise and family-wise constraints are designed to enhance the individual similarity and couple the parents and children within families for discriminative features. Comprehensive experiments on two large kinship datasets show that the proposed algorithm can effectively inherit knowledge from different databases and modalities and achieve the state-of-the-art performance.**

***Keywords*-kinship classification; latent subspace learning; person-wise and family-wise constraints;**

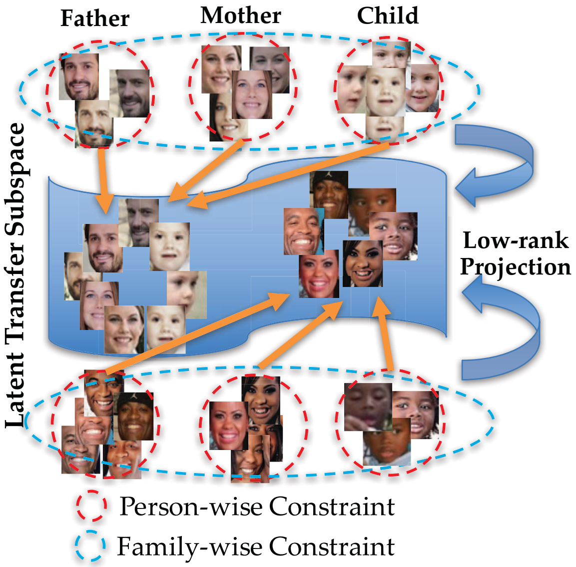


Figure 1: Illustration of the proposed Latent Adaptive Subspace framework for kinship classification. Cross knowledge from the auxiliary database is learned via transfer learning. The person/family-wise constraint help reduce the intra-person and intra-family variations.

I. INTRODUCTION

Mining relationships between people are a hot topic that has caught rising attention in the literatures. Kinship is one of the most obvious among the possible relationships. Parents pass down genes to their offspring. Inheritance and resemblance take place. Facial features are doubtlessly crucial clues of a human heritage, *e.g.*, gender, age, race, *etc*. Automatic kinship recognition has been applied in many emerging applications, such as automatic photo browsing and organizing, missing children searching and image understanding *etc*, which can roughly be divided into two categories, kinship verification and classification. Different from kinship verification [1], [2], [3], [4], which tells two persons have certain kinship relation or not, kinship classification aims to identify the family that a person belongs to [1], [5], [4], [6], [7]. Compared with face identification or verification [8], [9], which aims at recognizing the same person, kinship classification is more challenging. (1) Age and appearance variations resulting from different ages and various identities [10], [11] pose intrinsic challenges for kinship recognition; (2) Kinship classification contains several different relationships within families; (3) There exist limited or even no training samples for the test subjects [12], which makes it hard to directly use current face identification or verification technologies [13], [14], [15].

Traditional kinship classification algorithms [5], [16] build models to capture the parent-children relationship with complete parents and children data. These models were applied to the test data that are not involved in training. However, in the real-world applications that aim to find missing children, the parents data are available, which could be utilized to provide more information for the model. To address challenges of ad-hoc, we employ an auxiliary database with

---

[*] indicates equal contribution.

978-1-5386-2335-0/18/$31.00 ©2018 IEEE
DOI 10.1109/FG.2018.00030

both parents and their children for kinship classification. Generally speaking, we treat face images of parents and children as two separate modalities. By this means, we could tackle this challenge by transferring knowledge from one modality to the other. To this end, missing modality of the test subjects or children can be implicitly recovered by the knowledge between parents and children transferred from the auxiliary database.

Targeting at the kinship classification, we propose a Latent Adaptive Subspace (LAS) framework by modeling the test subjects or children as the missing modality during training (Figure 1). By joining latent factor recovery and the hierarchical family structure preserving, we build a latent transfer model with low-rank constraint to adapt knowledge from auxiliary complete database to target database. Consequently, LAS builds a shared feature subspace for which knowledge can successfully be transferred between different families, and then further adapted between parents and children. The person-wise and family-wise constraints (Figure 1) within LAS are designed to enhance the individual similarity and couple the parents and children within families for discriminative features. In summary, our major contributions are highlighted as:

- The complete parent-children family knowledge from external database is adapted to the target families through low-rank transfer subspace model. Furthermore, unseen test children are implicitly recovered as latent factors during model training.
- Features shared between parents and children are learned through the person-wise and family-wise constraints, so that such hierarchical family knowledge makes the latent adaptive subspace more robust in recognizing missing children.
- We evaluate our model on two large kinship datasets and achieve state-of-the-art performance.

II. THE PROPOSED APPROACH

In this section, we first introduce the preliminaries, then show the latent adaptive subspace learning with person/family-wise constraint and corresponding solutions.

A. Preliminaries

As stated in the *Kinship Classification* problem, we may not have the test children data available during training. Hence, we turn to the help of an auxiliary database X_s with complete parent-children pair, denoted as $\{X_{s,p}, X_{s,c}\}$. We further denote our target database $X_t = \{X_{t,p}, X_{t,c}\}$, where $X_{t,c}$ is blind in the model learning. Specifically, $X_{s,p} \in \mathbb{R}^{d\times n_{p_s}}, X_{s,c} \in \mathbb{R}^{d\times n_{c_s}}, X_{t,p} \in \mathbb{R}^{d\times n_{p_t}}, X_{t,c} \in \mathbb{R}^{d\times n_{c_t}}$, where d is the original feature dimension and $n_{p/c_{s/t}}$ are the sample size of the parents/children in auxiliary/target database, respectively. An illustration of these notations can be found in Table I.

Table I: Notations

Variable	Domain	Description
d	$\mathbb{R}$	Dimension of original features
n_s	$\mathbb{R}$	Number of samples in auxiliary database
n_t	$\mathbb{R}$	Number of samples in target database
p	$\mathbb{R}$	Dimension of features in subspace
X_s	$\mathbb{R}^{d\times n_s}$	Auxiliary data matrix
X_t	$\mathbb{R}^{d\times n_t}$	Target data matrix
Y_s	$\mathbb{R}^{p\times n_s}$	Low-dimensional auxiliary data matrix
P	$\mathbb{R}^{d\times p}$	Subspace projection
Z	$\mathbb{R}^{n_t\times n_s}$	Reconstruction coefficents
L	$\mathbb{R}^{p\times p}$	Latent factor

Note: $n_t = n_{p_t} + n_{c_t}, n_s = n_{p_s} + n_{c_s}$.

Traditional kinship classification techniques manage to build a model from the complete auxiliary database X_s to capture the parent-children relationship. Inspired by recent missing modality transfer learning [17], we also attempt to treat the unseen test data $X_{t,c}$ as a latent factor. In this way, complete X_s would assist the knowledge discovery of X_t when $X_{t,c}$ is unavailable during training. So far, the kinship classification turns to solve the problem of how to adapt knowledge from complete X_s to X_t by uncovering the latent information of $X_{t,c}$. Thus, we propose a latent adaptive subspace for effective knowledge transfer to tackle with kinship classification challenge.

B. Latent Adaptive Subspace Learning

Targeting at the kinship classification, we model an adaptive subspace learning by jointing latent factor recovery and hierarchical family structure preserving. Specifically, we treat the test children as the missing modality and build a latent transfer model to adapt knowledge from other complete database. To capture more effective information, we explore a family-person hierarchical structure to guide the subspace projection learning. In this way, we could uncover more knowledge from external database and family-wise structure to boost the kinship classification.

We first assume the missing modality $X_{t,c}$ is available to derive its formulation under the latent adaptive subspace learning framework. In the beginning, we pre-learn a low-dimensional feature from all available data $X_{s,p}, X_{s,c}$ by supervised subspace learning methods [18]. This helps to initialize the projection matrix and achieve stable solutions. Specifically, we obtain low-dimensional features $Y_{s,p} \in \mathbb{R}^{p\times n_{p_s}}, Y_{\mathrm{s,c}} \in \mathbb{R}^{p\times n_{c_s}}$. Then, we could exploit low-rank transfer learning for both auxiliary and target databases by seeking a domain-invariant subspace projection P.

Specifically, we assume auxiliary parent/children data could be reconstructed by target parent/children data, which are formulated as follows:

$$\min_{Z_p} \operatorname{rank}(Z_p), \quad \text{s.t.} \quad Y_{s,p} = P^{\mathrm{T}} X_{t,p} Z_p, \tag{1}$$

$$\min_{Z_c} \operatorname{rank}(Z_c), \quad \text{s.t.} \quad Y_{s,c} = P^{\mathrm{T}} X_{t,c} Z_c, \tag{2}$$

where $\mathrm{rank}(\cdot)$ is the rank of a matrix. $Z_p \in \mathbb{R}^{n_{t_p} \times n_{s_p}}$ and $Z_c \in \mathbb{R}^{n_{c_t} \times n_{c_s}}$, which are two low-rank coefficients matrix. Thus, Eq (1) and Eq (2) can be rewritten as

$$\min_{Z_s} \mathrm{rank}(Z_s), \quad \text{s.t.} \quad Y_s = P^{\mathrm{T}} X_t Z_s, \tag{3}$$

where $Y_s = [Y_{s,p} \quad Y_{s,c}]$, $X_t = [X_{t,p} \quad X_{t,c}]$, and

$$Z_s = \left[\begin{array}{cc} Z_p & 0 \\ 0 & Z_c \end{array} \right].$$

Obviously, $\mathrm{rank}(Z_s) = \mathrm{rank}(Z_p) + \mathrm{rank}(Z_c)$. Generally, nuclear norm is exploited to address the rank minimization NP-hard problem [17], [19] as:

$$\min_{Z_s} \|Z_s\|_*, \quad \text{s.t.} \quad Y_s = P^{\mathrm{T}} X_t Z_s, \tag{4}$$

where $\|\cdot\|_*$ represents the nuclear norm. In this case, Y_s can be spanned by $P^{\mathrm{T}} X_t$. However, when the $X_{t,c}$ is missing in the training stage, we could follow [17] to deduct our latent adaptive subspace to use auxiliary database to recover the missing $X_{t,c}$. Thus, the latent factor is able to be recovered by optimizing the following formulation:

$$\begin{array}{c} \min\limits_{Z,L,P} \|Z\|_* + \|L\|_*, \\ \text{s.t.} \quad Y_s = P^{\mathrm{T}} X_t Z + L Y_s, \quad P^{\mathrm{T}} P = I_p, \end{array} \tag{5}$$

where Z is a low-rank coefficients matrix, L is a latent matrix, and P is the learned subspace projection. The orthogonal constraint $P^{\mathrm{T}} P = I_p$ is imposed to avoid some trivial solutions.

C. Learning Discriminative Projection Through Person/Family-wise Constraint

In all training data, we have M identities. Each identity contains several samples. Specifically, we have n_m samples for the m-th person, $1 \le m \le M$. Therefore, it is essential to utilize the person-wise knowledge to make data points of the same person could still be lying closely in the latent space. Specifically, we define the person-wise constraint by enforcing within-person samples to be close to its mean as follows:

$$\Omega_m(P) = \frac{1}{n_m} \sum_{i=1}^{n_m} \left\| P^{\mathrm{T}} x_i - \mu_m \right\|_2^2, \tag{6}$$

where μ_m is the mean of the m-the person and can be calculated by $\frac{1}{n_m} \sum_{i=1}^{n_m} P^{\mathrm{T}} x_i$. Therefore, we have the person-wise constraint for all identities as follows:

$$\begin{array}{rcl} \Omega(P) = & \sum\limits_{m=1}^{M} \Omega_m(P) \\ = & \sum\limits_{m=1}^{M} \frac{1}{n_m} \sum\limits_{i=1}^{n_m} \left\| P^{\mathrm{T}} x_i - \mu_m \right\|_2^2 \\ = & \sum\limits_{m=1}^{M} \frac{1}{n_m} \sum\limits_{i=1}^{n_m} \left\| P^{\mathrm{T}} x_i - \frac{1}{n_m} \sum\limits_{i=1}^{n_m} P^{\mathrm{T}} x_i \right\|_2^2 \\ = & \sum\limits_{m=1}^{M} \mathrm{tr}(P^{\mathrm{T}} W_m P), \end{array} \tag{7}$$

where W_m can be written as following:

$$W_m = \frac{1}{n_m} \sum_{i=1}^{n_m} (x_i - \bar{x})(x_i - \bar{x})^{\mathrm{T}}, \tag{8}$$

where $\bar{x} = \frac{1}{n_m} \sum_{i=1}^{n_m} x_i$.

However, Eq. (7) only considers the discriminative information within each individual person, while ignoring the family group information.

Actually, the new presentations for each family should also be very similar, and thus, we could preserve the discriminative information with each family. For the n-th family, $1 \le n \le N$,we have K_n family members. The k-th member has the center u_k, $1 \le k \le K_n$. Specifically, we attempt to constrain the person-wise mean to be close to its family-wise mean. In this way, we could enforce family members to be close with each other in the latent space. We have the following expression for the family-n as:

$$\Psi_n(P) = \frac{1}{K_n} \sum_{k=1}^{K_n} \|\mu_k - c_n\|_2^2, \tag{9}$$

where c_n is the center of the family and can be calculated by $\frac{1}{K_n} \sum_{k=1}^{K_n} \mu_k$.

Therefore, we have the family-wise constraint for all families as follows:

$$\begin{array}{l} \Phi(P) = \sum\limits_{n=1}^{N} \Psi_n(P) \\ \quad = \sum\limits_{n=1}^{N} \frac{1}{K_n} \sum\limits_{k=1}^{K_n} \|\mu_k - c_n\|_2^2 \\ \quad = \sum\limits_{n=1}^{N} \frac{1}{K_n} \sum\limits_{k=1}^{K_n} \left\| \frac{1}{n_k} \sum\limits_{i=1}^{n_k} P^{\mathrm{T}} x_i - \frac{1}{K_n} \sum\limits_{k=1}^{K_n} \frac{1}{n_k} \sum\limits_{i=1}^{n_k} P^{\mathrm{T}} x_i \right\|_2^2 \\ \quad = \sum\limits_{n=1}^{N} \mathrm{tr}(P^{\mathrm{T}} F_n P). \end{array} \tag{10}$$

F_n can be expressed as follows:

$$F_n = \frac{1}{K_n} \sum_{k=1}^{K_n} (x_{ik} - \hat{x}_{ik})(x_{ik} - \hat{x}_{ik})^{\mathrm{T}}, \tag{11}$$

where $x_{ik} = \frac{1}{n_k} \sum_{i=1}^{n_k} x_i$ and $\hat{x}_{ik} = \frac{1}{K_n} \sum_{k=1}^{K_n} \frac{1}{n_k} \sum_{i=1}^{n_k} x_i$.

To this end, we have our final objective function by seeking an effective projection through latent factor recovery and discriminative information preserving as follows:

$$\begin{array}{l} \min\limits_{Z,L,P} \|Z\|_* + \|L\|_* + \beta \Omega(P) + \gamma \Psi(P), \\ \text{s.t.} \quad P^{\mathrm{T}} P = I_p, \quad Y_s = P^{\mathrm{T}} X_t Z + L Y_s. \end{array} \tag{12}$$

D. Solving the Optimization Problem

Problem (12) can be solved by some well-known algorithms, e.g., Augmented Lagrange Methods (ALM). However, the ALM has some complex matrix operations, *e.g.*inverse and multiplications, when the relax variables introduce the quadratic term. These operations are time-consuming. Hence the first order Taylor expansion like approximation is used to replace the quadratic term which causes the problem. This leads to a simpler solution. For clearance, the augmented Lagrangian function of problem (12) can be written as:

$$\begin{aligned}\mathcal{J} = \;& \|Z\|_* + \|L\|_* + \beta \mathrm{tr}(P^{\mathrm{T}} W P) + \gamma \mathrm{tr}(P^{\mathrm{T}} F P) \\ & + \langle Q, Y_s - P^{\mathrm{T}} X_t Z - L Y_s \rangle \\ & + \tfrac{\mu}{2}(\|Y_s - P^{\mathrm{T}} X_t Z - L Y_s\|_{\mathrm{F}}^2),\end{aligned} \tag{13}$$

where $\mu > 0$ is a penalty parameter, Q is the Lagrange multiplier and $\langle , \rangle$ is the inner product between matrixes. $W = \sum_m^M W_m$ and $F = \sum_n^N F_n$. The last two terms are merged into quadratic terms and the formulation becomes:

$$\begin{aligned}\mathcal{J} = \;& \|Z\|_* + \|L\|_* + \mathrm{tr}\big(P^{\mathrm{T}}(\beta W + \gamma F)P\big) \\ & + h(Z, L, P, Q, \mu) - \frac{1}{\mu}\|Q\|_{\mathrm{F}}^2,\end{aligned} \tag{14}$$

where $h(Z, L, P, Q, \mu) = \dfrac{\mu}{2}(\|Y_s - P^{\mathrm{T}} X_t Z - L Y_s + Q/\mu\|_{\mathrm{F}}^2)$. We follow the traditional ALM to not only solve the new formulation is over Z, L and P jointly, but also solve each of them by fixing the rest. Thus, one subproblem is solved at one time. The term h is approximated using the first order expansion of one variable, assuming other variables are constant. With iteration $t+1$ ($t \geq 0$), we have:

Update Z:

$$\begin{aligned}Z^{(t+1)} &= \arg\min_{Z} \|Z\|_* + h(Z, L^{(t)}, P^{(t)}, Q^{(t)}, \mu) \\ &= \arg\min_{Z} \frac{1}{\eta_z \mu}\|Z\|_* + \frac{1}{2}\|Z - Z^{(t)} + \nabla_Z h\|_{\mathrm{F}}^2,\end{aligned} \tag{15}$$

where $\nabla_Z h$ is the gradient of $h(\cdot)$ with respect to Z, which can be written as $\nabla_Z h(Z^{(t)}, L^{(t)}, P^{(t)}, Q^{(t)}, \mu) = X_t^{\mathrm{T}} P^{(t)}(Y_s - {P^{(t)}}^{\mathrm{T}} X_t Z^{(t)} - L^{(t)} Y_s + Q^{(t)}/\mu)$, $\eta_z = \|{P^{(t)}}^{\mathrm{T}} X_t\|_2^2$. Problem (15) is solvable using the singular value thresholding (SVT) operator [20].

Update L:

$$\begin{aligned}L^{(t+1)} &= \arg\min_{L} \|L\|_* + h(Z^{(t+1)}, L, P^{(t)}, Q^{(t)}, \mu) \\ &= \arg\min_{L} \frac{1}{\eta_l \mu}\|L\|_* + \frac{1}{2}\|L - L^{(t)} + \nabla_L h\|_{\mathrm{F}}^2,\end{aligned} \tag{16}$$

where similarly, the gradient $\nabla_L h = \nabla_L h(Z^{(t+1)}, L^{(t)}, P^{(t)}, Q^{(t)}, \mu)$ is $(Y_s - {P^{(t)}}^{\mathrm{T}} X_t^{(t+1)} - L^{(t)} Y_s + Q^{(t)}/\mu) Y_s^{\mathrm{T}}$, $\eta_l = \|Y_s\|_2^2$.

Update P:

$$\begin{aligned}P^{(t+1)} = \;& \arg\min_{P^{\mathrm{T}}P = I_p} \alpha\|P\|_{2,1} + \beta \mathrm{tr}(P^{\mathrm{T}} \mathcal{M} P) + \gamma \mathrm{tr}(P^{\mathrm{T}} \Psi P) \\ & + \frac{\mu}{2}(\|\bar{P}^{(t)} - P^{\mathrm{T}} X_t Z^{(t+1)}\|_{\mathrm{F}}^2,\end{aligned}$$

where $\bar{P}^{(t)} = Y_s - L^{(t+1)} Y_s + Q^{(t)}/\mu$, which can be rewritten into the equivalent problem [21]:

$$\begin{aligned}P^{(t+1)} = \;& \arg\min_{P^{\mathrm{T}}P = I_p} \mathrm{tr}\big(P^{\mathrm{T}}(\beta W + \gamma F)P\big) \\ & + \frac{\mu}{2}(\|\bar{P}^{(t)} - P^{\mathrm{T}} X_t Z^{(t+1)}\|_{\mathrm{F}}^2,\end{aligned} \tag{17}$$

```
Algorithm 1 Latent Adaptive Subspace Learning
Input: $X_s, X_t, \gamma, \beta, Y_s, W_m, F_n$
Initialize: $L^{(0)} = 0, Z^{(0)} = 0, Q^{(0)} = 0,$
            $\mu = 10^{-6}, \epsilon = 10^{-6}, \rho = 1.2,$
            $\mathrm{maxIter} = 30, \max_\mu = 10^6, t = 0.$
while not converged or $t \leq \mathrm{maxIter}$ do
  1. Update $Z^{(t+1)}$ by fixing others according to (15);
  2. Update $L^{(t+1)}$ by fixing others according to (16);
  3. Update $P^{(t+1)}$ by fixing others according to (17);
  4. Update the multipliers $Q^{(t+1)}$
       $Q^{(t+1)} = Q^{(t)} + \mu(Y_s - {P^{(t+1)}}^{\mathrm{T}} X_t Z^{(t+1)}$
                $- L^{(t+1)} Y_s);$
  5. Update $\mu$ using $\mu = \min(\rho\mu, \max_\mu);$
  6. Check converge using conditions
       $\|Y_s - {P^{(t+1)}}^{\mathrm{T}} X_t Z^{(t+1)} - L^{(t+1)} Y_s\|_\infty < \epsilon.$
  7. $t = t + 1.$
end while
output: $P, Z, L$
```

Due to theorthogonal constraints, we address the difficult non-convex problem (17) using a gradient descent optimization procedure [22]. The gradient of $\mathcal{J}$ w.r.t P is first calculated as $\frac{\partial \mathcal{J}}{\partial \mathrm{P}} = (2\beta W + 2\gamma F + \mu X_t Z^{(t+1)}(X_t Z^{(t+1)})^{\mathrm{T}})P - \mu X_t Z^{(t+1)}(Y_s - L^{(t+1)} Y_s + Q^{(t)}/\mu)^{\mathrm{T}}$. Then the skew-symmetric matrix is calculated and P is optimized until Armijo-Wolfe conditions meet.

The procedure of the solutions is shown in **Algorithm 1**. The parameters ϵ, $\max_\mu$, ρ, μ and maxIter are set empirically. The parameters γ, β are tuned using grid search. For the initialization of Y_s, we aim to preserve family-wise and person-wise knowledge so that we adopt the following expression:

$$Y_s = \arg\min_{Y_s} \mathrm{tr}\big(Y_s^{\mathrm{T}}(\beta W + \gamma F)Y_s\big), \quad \text{s.t.} \quad Y_s^{\mathrm{T}} Y_s = I_p.$$

E. Complexity Analysis

Assume X_s and X_t are $d \times n$ matrixes, and P is a $d \times p$ matrix ($p \ll d$), then time-consuming components of **Algorithm 1** have: 1) Subspace optimization in Line 3 and 2) Trace norm computation in Line 1.

For the computation complexity, the SVD computation in Line 1 takes $O(n^3)$, and it can be improved to $O(rn^2)$ by accelerations of SVD, where r is the rank of the low-rank matrix Z. For Line 3, since we calculate the gradient descent of P first, it usually costs $O(d^2p)$. The skew-symmetric matrix calculation also takes $O(d^2p)$. Consider this step would cost t_1 iterations to converge, the complexity of Line 3 is $O(t_1 d^2 p)$. In total, the whole time complexity would be $O(t_2(t_1 d^2 p + rn^2))$ if there are t_2 iterations in Algorithm 1.

III. EXPERIMENTS

In this section, the datasets and experimental settings are introduced first. The influence of model parameters and convergence is discussed later, followed by results and analysis on two large kinship datasets.

A. Datasets and Experimental Settings

Experiments are conducted on two kinship datasets, Families In the Wild (FIW) and Family 101 (FM101).

Families In the Wild (FIW) [4] is the largest kinship recognition dataset up to date. 276 families that have more than six family members are utilized. These family members are split into two parts, parents and children. The children part consists of the last generation of each family while the parents part is the rest family members. All 276 families are divided into auxiliary and target database randomly and equally. Thus, we have 138 families for the auxiliary database and the rest 138 families for the target database. In auxiliary database, parents have 2,612 images and children have 1,730 images. In target database, parents have 2,480 images and children have 1,832 images.

Family 101 (FM101) [5] contains 101 different family trees, including 206 nuclear families, 607 individuals. Similar to FIW, we re-label all family members into two parts, parents and children. Then 50 families are randomly selected as the auxiliary database and the rest 51 families are used for the target database. In auxiliary database, parents have 2,308 images and children have 2,224 images. In target database, parents have 3,396 images and children have 5,413 images.

We do face detection and facial points detection using MTCNN [23] for all the images in these datasets. Then face alignment is employed before feeding these faces into deep Convolutional Neural Networks(CNNs) for feature extraction. Three state-of-the-art deep face models (Centerface [15], VGGFace [14], DSDA [24]) are utilized to extract deep face features. These features are used for comparisons and the inputs for subspace learning methods. For Centerface, we use the alignment method provided by authors. For VGGFace and DSDA, we tried different alignment methods and reported the best result. The nearest-neighbor classifier with Euclidean distance metric is exploited for classification to evaluate the ability of different face representations.

We adopt the standard Cumuative Match Characteristic (CMC) curve as the performance measurement for kinship classification. The recognition rate is defined as the proportion of queries correctly matched to a corresponding gallery entity and it is displayed as a function of the rank.

B. Parameter and Convergence Analysis

Parameters analysis is conducted on FIW dataset with deep face features from Centerface[15]. We observe a stable range for parameters β and γ in term of the recognition rate. The hyper parameter β dominates the intra-person variations and the other hyper parameter γ controls the intra-family variations. Both of them are essential to our model. We conduct two experiments to investigate the sensitiveness of two parameters.

In the experiment of parameter β, we fix γ to 0.02 and vary β from 0 to 0.02. The recognition rates of rank 10, 30 and 50 are shown in Figure 2a. We can see that the results of small β are very stable and show small improvements. But the recognition rate drops when it gets larger. This is brought from that the original face features have a very good representation ability of each person. The term Eq. (7) aims to reduce the intra-person variations. It has the same goal with the original face identification task [15], [24], [14]. But these deep features were learned from large scale face data and already have very small intra-person variations. Large β makes the model overfit to the training data and this leads to bad generalization ability. We fix β to 0.003 in all experiments in section III-C.

In the experiment of parameter γ, we fix β to 0.003 and vary γ from 0 to 4. The results are shown in Figure 2b with the $\log$ scale. It is very clear that simply using the person-wise constraint with little family information incorporated (in these cases, γ is set to small values) is not a good choice, leading to poor recognition performance. Properly choosing the value of can improve the recognition accuracy of the learned subspace projection P in Eq. (7). The recognition performance remains stable with γ at the near range of 0.02 and starts dropping when γ is larger than 0.04. But even the performance drops, the recognition rate is still better than these small γ. In section III-C, we fix γ to 0.02.

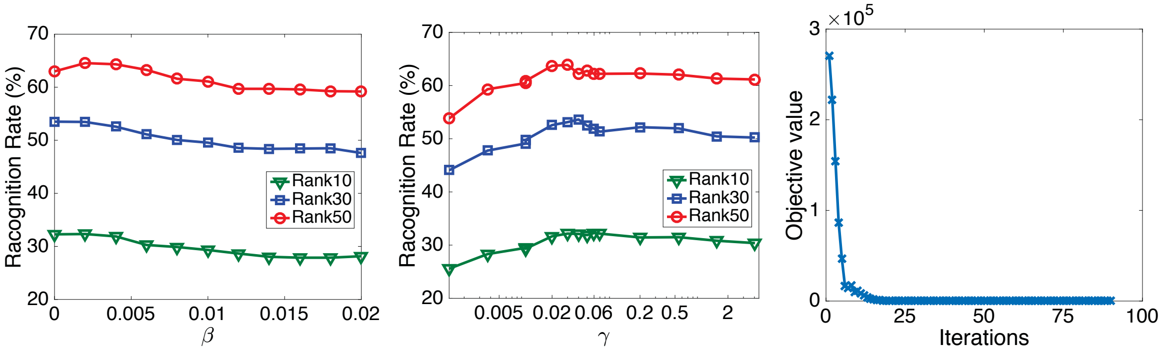
We also test the convergence with the objective value (Eq. (12)) over different iterations. From Figure 2c, we can observe our method converges after about 20 iterations. We set maximum iteration to 30 as the default setting.

C. Kinship Classification

In this section, we compare with several state-of-the-art face recognition algorithms and subspace learning methods, including: Centerface [15] , VGGFace [14], DSDA [24], LAC [17] and SRRS [25]. We directly extract deep face features from Centerface, VGGFace and DSDA models on target data for testing. For subspace learning methods LAC and SRRS, deep face features from Centerface are taken as input and a subspace is learned from each method. The CMC curves are shown in Figure 3a and Figure 3b. Matching rate comparisons of FIW and FM101 are reported in Table II.

From the results, we can see that all methods have better performance in FM101 than FIW. This observation results from: 1) FM101 has less families than FIW. For test data, FM101 has 51 families while FIW has 136 families. 2) The average number of images for each family in FM101 is 132 which is larger than the average number 31 in FIW.

For FIW, compared with three kinds of deep face features, we could notice that Centerface, VGGFace and DSDA have similar performance on FIW. Thus we take deep face



(a) Parameter β (b) Parameter γ (c) Convergence

Figure 2: Parameters and convergence analysis. The recognition rate is calculated with rank 10, 30 and 50. Each parameter is evaluated by fixing others. Objective value decreases quickly and the model converges after about 20 iterations.

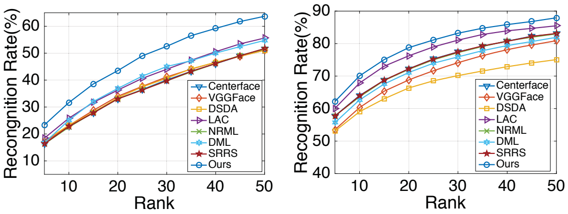
features from Centerface as the input for subspace learning for other methods. This is due to the Centerface features have less dimension (1,024 compared with 4,096 in VGGFace and DSDA). Compared with either face recognition algorithms (Centerface, VGGFace and DSDA) or subspace learning methods (LAC and SRRS), our approach achieves the best performance. LAC gained some improvement from Centerface features due to it also tried to recover the unseen children modality in target database which is similar to us. But it did not incorporate the identity information and the family constraint. Thus, we can see that an auxiliary database with complete parents and children data and the person-wise family-wise constraint do help on learning a better subspace to represent faces in one family.

Similar observations can be found on FM101 results. But on FM101, Centerface shows superiority compared with VGGFace and DSDA. We think this is due to the different generalization ability of these features on this dataset. Then we also take the deep face features as the input for subspace learning since it has better performance. Our approach also beat other competitors. This demonstrated that the learned subspace features are good at kinship classification again.

Some samples for kinship classification are shown in Figure 4. From the top 10 candidates of each query face image, we can observe that our model can handle the large appearance variations between parents and children. Especially, in the first, fourth and fifth rows, our model returns the query's different parents at the same time. We also see that there are some near-duplicated and wrong labeled faces in these kinship datasets. For examples, in the second row, the second, third and fifth results in the top 10 candidates are almost the same face. Similar findings are found in the fourth row. Two Tom Hanks faces are exactly the same image but exist as different images in the dataset. The last row is all correct due to the query image is wrong labeled as Cameron Douglas but actually is Michael Douglas that is exactly the same person in returned results.

Table II: Matching rate comparison on FIW and FM101.

	Rank	10	20	30	40	50
FIW	Centerface [15]	0.2285	0.3305	0.4008	0.4610	0.5162
	VGGFace [14]	0.2313	0.3400	0.4119	0.4660	0.5111
	DSDA [24]	0.2324	0.3344	0.4069	0.4660	0.5106
	LAC [17]	0.2586	0.3629	0.4404	0.5056	0.5563
	SRRS [25]	0.2285	0.3305	0.3974	0.4615	0.5173
	NRML [2]	0.2252	0.3300	0.3974	0.4615	0.5151
	DML [6]	0.2514	0.3690	0.4504	0.5006	0.5475
	Ours	**0.3166**	**0.4353**	**0.5256**	**0.5920**	**0.6371**
FM101	Centerface [15]	0.6401	0.7222	0.7746	0.8077	0.8301
	VGGFace [14]	0.6031	0.6881	0.7404	0.7815	0.8089
	DSDA [24]	0.5904	0.6635	0.7015	0.7289	0.7507
	LAC [17]	0.6787	0.7624	0.8106	0.8393	0.8548
	SRRS [25]	0.6393	0.7224	0.7741	0.8070	0.8313
	NRML [2]	0.6361	0.7206	0.7722	0.8080	0.8298
	DML [6]	0.6267	0.7111	0.7591	0.7939	0.8193
	Ours	**0.7006**	**0.7876**	**0.8319**	**0.8585**	**0.8794**



(a) CMC curves on FIW. (b) CMC curves on FM101.

Figure 3: CMC curves with different methods: a) results on FIW. b) results on FM101.

IV. CONCLUSION

In this paper, we proposed a Latent Adaptive Subspace framework by modeling the test children as the missing modality. Complete parent-children family knowledge from external databases was adapted to the target families through low-rank transfer subspace model. The core idea of LAS was to build a shared feature subspace such that knowledge can be successfully transferred between different families and between parents and children. A person-wise constraint

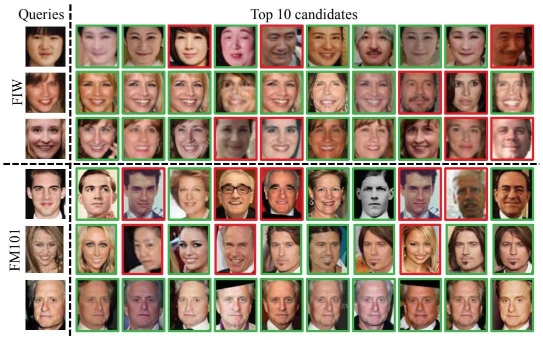


Figure 4: Samples for Kinship classification. The first left column is the query children. The rest columns are the top 10 results. Correct results are with green border. Error results are with red border. Top three rows are from FIW and bottom three rows are from FM101.

and a family-wise constraint were introduced to enhance the individual similarity and couple the parents and children within families for discriminative features.

ACKNOWLEDGMENT

This work is supported in part by the NSF IIS award 1651902, ONR Young Investigator Award N00014-14-1-0484, and U.S. Army Research Office Award W911NF-17-1-0367.

REFERENCES

[1] S. Xia, M. Shao, J. Luo, and Y. Fu, "Understanding kin relationships in a photo," *IEEE Transactions on Multimedia*, 2012.

[2] J. Lu, X. Zhou, Y.-P. Tan, Y. Shang, and J. Zhou, "Neighborhood repulsed metric learning for kinship verification," *IEEE transactions on pattern analysis and machine intelligence*, 2014.

[3] Z. Zhang, Y. Chen, and V. Saligrama, "Group membership prediction," in *ICCV*, 2015.

[4] J. P. Robinson, M. Shao, Y. Wu, and Y. Fu, "Families in the wild (fiw): Large-scale kinship image database and benchmarks," in *ACM MM*, 2016.

[5] R. Fang, A. C. Gallagher, T. Chen, and A. Loui, "Kinship classification by modeling facial feature heredity," in *ICIP*, 2013.

[6] S. Wang, J. P. Robinson, and Y. Fu, "Kinship verification on families in the wild with marginalized denoising metric learning," in *FG*, 2017.

[7] Y. Xu, X. Fang, J. Wu, X. Li, and D. Zhang, "Discriminative transfer subspace learning via low-rank and sparse representation," *IEEE Transactions on Image Processing*, 2016.

[8] 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, Tech. Rep., 2007.

[9] Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao, "Ms-celeb-1m: A dataset and benchmark for large-scale face recognition," in *ECCV*, 2016.

[10] Y. Fu, G. Guo, and T. S. Huang, "Age synthesis and estimation via faces: A survey," *IEEE transactions on pattern analysis and machine intelligence*, 2010.

[11] X. Geng, C. Yin, and Z.-H. Zhou, "Facial age estimation by learning from label distributions," *IEEE transactions on pattern analysis and machine intelligence*, 2013.

[12] A. J. Sedlack, D. Finkelhor, H. Hammer, and D. Schultz, "National estimates of missing children: An overview." *NISMART Series Bulletin*, 2002.

[13] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *CVPR*, 2015.

[14] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *BMVC*, 2015.

[15] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, "A discriminative feature learning approach for deep face recognition," in *ECCV*, 2016.

[16] J. Dong, X. Ao, S. Su, and S. Li, "Kinship classification based on discriminative facial patches," in *VCIP*, 2014.

[17] Z. Ding, M. Shao, and Y. Fu, "Missing modality transfer learning via latent low-rank constraint," *IEEE Transactions on Image Processing*, 2015.

[18] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on pattern analysis and machine intelligence*, 1997.

[19] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013.

[20] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, 2010.

[21] C. Hou, F. Nie, D. Yi, and Y. Wu, "Feature selection via joint embedding learning and sparse regression," in *IJCAI*, 2011.

[22] Z. Wen and W. Yin, "A feasible method for optimization with orthogonality constraints," *Mathematical Programming*, 2013.

[23] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Processing Letters*, 2016.

[24] I. Masi, A. T. Trn, T. Hassner, J. T. Leksut, and G. Medioni, "Do we really need to collect millions of faces for effective face recognition?" in *ECCV*, 2016.

[25] S. Li and Y. Fu, "Learning robust and discriminative subspace with low-rank constraints," *IEEE transactions on neural networks and learning systems*, 2016.