KINSHIP VERIFICATION BASED ON STATUS-AWARE PROJECTION LEARNING

Haijun Liu, Jian Cheng, Feng Wang

University of Electronic Science and Technology of China
School of Electronic Engineering
haijunliu518128@gmail.com

ABSTRACT

Kinship verification for parent-child is considered to be an asymmetric metric process, in which parents and children are associated with different status where the parents are priorly known to be significantly older than the children. To address the asymmetric metric learning, a status-aware projection learning (SaPL) method is proposed for facial image-based kinship verification, especially for the parent-child kinship. SaPL learns two status-specific projections to capture the significant appearance commonality between parents and children, respectively. Each status-specific projection consists of two components: a common component shared by the two status projections and a status-specific component. SaPL generally outperforms the one Mahalanobis distance metric. Extensive experimental results and comparisons with state-of-the-art approaches demonstrate the effectiveness of the proposed SaPL for kinship verification.

Index Terms— kinship verification, status-aware projection, asymmetric metric learning

1. INTRODUCTION

Kin relationship has been well investigated in psychology community over the past decades [1, 2]. More recently, kinship verification using facial images has attracted substantial attention in biometrics society, mainly motivated by the observation that children generally resemble their parents more than other persons in term of facial appearance, due to the genetic overlapping. Simultaneously, evidence from psychology has demonstrated that facial appearance is an important and reliable cue to measure the genetic similarity between children and their parents [3, 4]. Moreover, kinship verification has many real-world applications, such as missing children search [5] and social media analysis [6–8].

There have been several early attempts on kinship verification using facial images over the past a few years. Among them most of the metric learning methods are focused on learning a Mahalanobis distance metric to measure the squared distance between a pair of samples x_i and y_i,

$$d_M(x_i, y_i) = \sqrt{(x_i - y_i)^T M (x_i - y_i)}, \tag{1}$$

where $x_i, y_i \in R^d$ are a pair of facial images in the original feature space, and $M \succeq 0$ is a positive semidefinite (PSD) matrix. A Mahalanobis distance is in essence equivalent to the Euclidean distance after some linear projection of the data samples. When M is factorized into $M = A^T A$, ($A \in R^{r \times d}$, $r \leq d$), Eqn. (1) can be rewritten as,

$$\begin{aligned} d_M(x_i, y_i) &= \sqrt{(x_i - y_i)^T A^T A (x_i - y_i)} \\ &= \sqrt{(a_i - b_i)^T (a_i - b_i)}, \end{aligned} \tag{2}$$

where $a_i = Ax_i$, $b_i = Ay_i$ are in the learned feature space.

Several representative metric learning methods have been proposed to measure the kin similarity of facial images. Lu et al. [6] proposed a neighborhood repulsed metric learning (NRML) method for kinship verification. NRML learns a distance metric under which the intraclass samples (with a kinship relation) are pulled as close as possible and interclass samples lying in a neighborhood are repulsed and pushed away as far as possible. To boost the verification performance, Yan et al. [7] proposed a discriminative multi-metric learning method to exploit complementary information from multi-view features. Zhou et al. [8] proposed a new ensemble similarity learning (ESL) method for kinship verification. ESL, whose similarity function parameterized by a diagonal matrix, generates an ensemble of similarity models with the aim of achieving strong generalization ability.

The aforementioned methods, along with the Mahalanobis distance metric, are all symmetric metric, since $d_M(x_i, y_i) = d_M(y_i, x_i)$. In kinship verification, those symmetric metric methods treat the parents images and children images equally. Namely, they learn a Mahalanobis distance metric between image pair, or learn one projection matrix to map the image pair into a common feature space (the learned feature space). However, the kinship verification for parent-child is considered to be an asymmetric metric process, in which parents and children are associated with different status where the parents are priorly known to be significantly

This work was supported by the National Natural Science Foundation of China (61671125, 61201271, 61301269), and the State Key Laboratory of Synthetical Automation for Process Industries (NO. PAL-N201401).

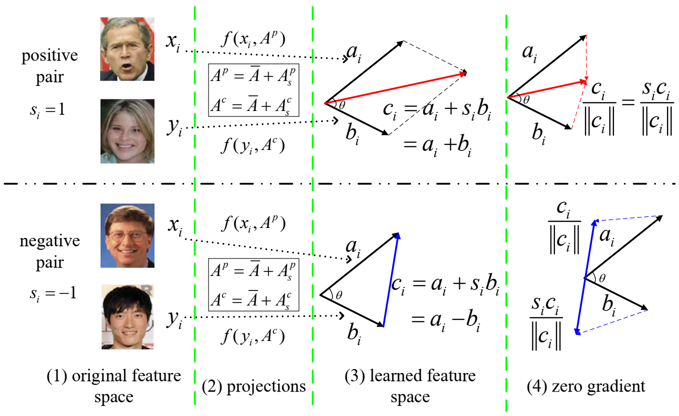


Fig. 1. A geometrical interpretation of our proposed method SaPL. (1) The example facial image pairs (triplet (x_i, y_i, s_i)) in the original feature space. (2) The two status-specific projections (A^p and A^c), each of which consists of two components: a common component ($\bar{A}$) shared by the two status projections and a status-specific component (A_s^p for A^p and A_s^c for A^c). (3) In the learned feature space, the objective is to make positive (kin, $s_i = 1$) pair vectors parallel and make negative (non-kin, $s_i = -1$) pair vectors opposite. (4) At zero gradient point, positive pair a_i and b_i are set as the same unit diagonal vector $\frac{c_i}{\|c_i\|}$, while negative pair a_i and b_i are set as the opposite unit diagonal vectors $\frac{c_i}{\|c_i\|}$ and $\frac{s_i c_i}{\|c_i\|}$, respectively.

older than the children. So we argue that the parents and children facial images should be separately processed to learn the significant appearance commonality.

To address the aforementioned problem, the effect of "age" in the facial image-based kinship verification, we propose a status-aware projection learning (SaPL) method for facial image-based kinship verification, especially for the parent-child kinship. SaPL learns two status-specific projections to capture the significant appearance commonality between parents and children, respectively. Each status-specific projection consists of two components: a common component and a status-specific component. The two status-specific projections are connected by the common component and optimized simultaneously. While the status-specific component allows our method to be more sensitive to the status of facial images. SaPL adopts the triangular similarity [9] as objective function to measure the similarity of the pair of facial images. The L-BFGS [10] optimization algorithm is adopted to compute the optimal status-specific projection matrices. Extensive experimental results and comparisons with state-of-the-art methods demonstrate the effectiveness of the proposed status-aware projection learning method for kinship verification.

2. PROPOSED METHOD

2.1. Status-aware projection

Traditional metric learning methods, such as side-information based linear discriminant analysis (SILD) [11], neighborhood repulsed metric learning (NRML) [6]and triangular similarity metric learning (TSML) [9], all only learn one linear transformation A to project image pairs (both parents and children images in kinship verification) into the learned feature space. It's not reasonable since parent-child kinship verification is considered to be an asymmetrical metric process. In a positive kinship image pair, the parent (father or mother) and child (son or daughter) are associated with different status where parent is priorly known to be significantly older than the child, resulting in significant appearance differences.

To address the asymmetrical metric problem (also to mitigate the effect of "age"), according to different status in kinship we propose to learn two status-specific projections A^p and A^c for parent and child, respectively. The triplet (x_i, y_i, s_i) is adopted to represent the kin (or non-kin) image pair, where x_i is the parent vector, y_i is the child vector in original feature space, and $s_i = 1$ (resp. $s_i = -1$) means that the two vectors are kinship (resp. non-kinship) (Fig.1 (1)). Using a linear transformation function $f(x, A)$, we project the triplet (x_i, y_i, s_i) in original feature space into another triplet (a_i, b_i, s_i) in the learned feature space through the two following status-specific projections,

$$a_i = f(x_i, A^p) = A^p x_i, \tag{3}$$
$$b_i = f(y_i, A^c) = A^c y_i, \tag{4}$$

where A^p is the linear projection for parents and A^c is the linear projection for children (Fig.1 (2)).

Assume each status-specific projection could be factorized into two components: a common component shared by the two status-specific projections and a status-specific component.

$$A^p = \bar{A} + A_s^p, \tag{5}$$
$$A^c = \bar{A} + A_s^c, \tag{6}$$

where $\bar{A}$ is the common component which connects the two status-specific projections, A_s^p and A_s^c are the status-specific component respectively for parents and children, which make our method be sensitive to the status of facial images.

2.2. The objective function

With the above defined status-aware projection, we model the status-aware projection learning (SaPL) method as following. Similar to TSML [9], in the learned feature space we can formulate the minimizing objective function through the triangular similarity as $J_i = \|a_i\| + \|b_i\| - \|c_i\|$, where $c_i = a_i + s_i b_i$ is a diagonal of the parallelogram formed by a_i and b_i. When

Table 1. The mean verification accuracy (%) of different methods with LBP feature on KinfaceW-I.

Method	F-S	F-D	M-S	M-D	Mean
SILD [11]	78.2	69.4	66.8	70.1	71.1
NRML [6]	81.4	69.8	67.2	72.9	72.8
ESL [8]	81.7	71.1	69.6	74.3	74.1
TSML [9]	80.5	71.6	72.8	74.1	74.8
SaPL	80.8	73.5	73.6	74.5	**75.6**

Table 2. The mean verification accuracy (%) of different methods with HOG feature on KinfaceW-I.

Method	F-S	F-D	M-S	M-D	Mean
SILD [11]	80.5	72.4	69.8	77.1	74.9
NRML [6]	83.9	74.6	71.6	80.0	77.5
ESL [8]	83.9	76.0	73.5	81.5	**78.6**
TSML [9]	83.7	74.6	71.1	80.7	77.5
SaPL	84.3	75.1	72.4	81.5	**78.3**

Table 3. The mean verification accuracy (%) of different methods with LBP feature on KinfaceW-II.

Method	F-S	F-D	M-S	M-D	Mean
SILD [11]	78.2	70.0	71.2	67.8	71.8
NRML [6]	79.2	71.6	72.2	68.4	72.9
ESL [8]	80.5	72.2	72.8	71.6	74.3
TSML [9]	80.2	71.6	74.0	73.8	74.9
SaPL	82.4	78.0	78.0	77.6	**79.0**

Table 4. The mean verification accuracy (%) of different methods with HOG feature on KinfaceW-II.

Method	F-S	F-D	M-S	M-D	Mean
SILD [11]	79.6	71.6	73.2	69.6	73.5
NRML [6]	80.8	72.8	74.8	70.4	74.7
ESL [8]	81.2	73.0	75.6	73.0	75.7
TSML [9]	81.0	71.0	75.0	73.2	75.1
SaPL	82.4	72.4	75.8	74.2	**76.2**

we constrain the norms $\|a_i\|$ and $\|b_i\|$ unchanged (e.g. $\|a_i\|$ and $\|b_i\|$ are approaching to 1) during optimization, minimizing J_i is equivalent to maximizing $\|c_i\|$. Moreover, it is also equivalent to minimizing the angle θ inside a positive pair ($s_i = 1$) or maximizing the angle θ inside a negative pair ($s_i = -1$), as shown in Fig.1 (3). This is also same to the core thought of cosine metric between a_i and b_i.

To prevent $\|a_i\|$ and $\|b_i\|$ from degenerating to 0, we constrains the norms to be 1. Adding the constraint $\frac{1}{2}(\|a_i\|-1)^2$ and $\frac{1}{2}(\|b_i\|-1)^2$ to the objective function,

$$\begin{aligned} J_i &= \frac{1}{2}(\|a_i\|-1)^2 + \frac{1}{2}(\|b_i\|-1)^2 + \|a_i\| + \|b_i\| - \|c_i\| \\ &= \frac{1}{2}\|a_i\|^2 + \frac{1}{2}\|b_i\|^2 - \|c_i\| + 1. \end{aligned} \tag{7}$$

For all possible positive and negative pairs, the overall objective function is,

$$J = \frac{1}{n}\sum_{i=1}^{n} J_i = \frac{1}{n}\sum_{i=1}^{n}(\frac{1}{2}\|a_i\|^2 + \frac{1}{2}\|b_i\|^2 - \|c_i\| + 1), \tag{8}$$

where n is the total number of image pairs.

2.3. Solve the problem

To simultaneously learn the common component $\bar{A}$ and status-specific components A_s^p and A_s^c, we reformulate Eqns.(3) and (4),

$$a_i = A^p x_i = A\tilde{x}_i, \tag{9}$$

$$b_i = A^c y_i = A\tilde{y}_i, \tag{10}$$

where $A = [\bar{A}, A_s^p, A_s^c]$, $\tilde{x}_i = [x_i; x_i; 0]$ and $\tilde{y}_i = [y_i; 0; y_i]$.

To prevent the above objective function (Eqn.(8)) from over-fitting, a regularization term is added,

$$J = \frac{1}{n}\sum_{i=1}^{n}(\frac{1}{2}\|a_i\|^2 + \frac{1}{2}\|b_i\|^2 - \|c_i\| + 1) + \frac{\lambda}{2}\|A - A_0\|^2, \tag{11}$$

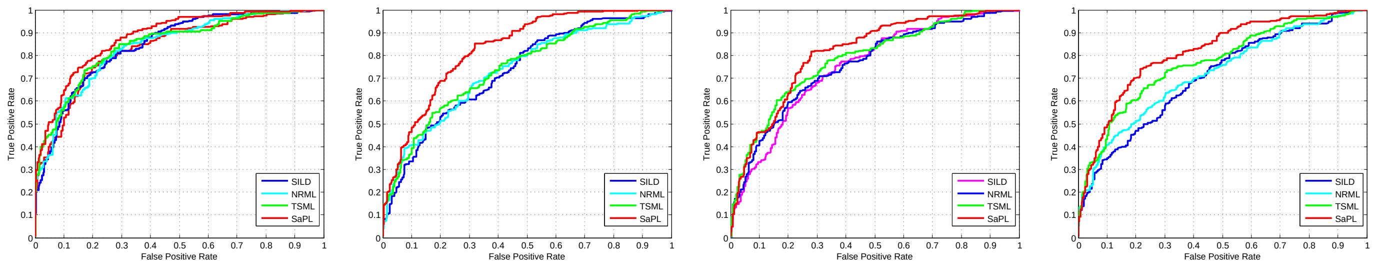
where $A_0 = [\bar{A}_0, A_{s0}^p, A_{s0}^c]$ is the predefined constraint matrix, and the initialization for $\bar{A}$, A_s^p and A_s^c, respectively. $\lambda > 0$ is the tradeoff parameter.

The gradient function over matrix A is,

$$\frac{\partial J}{A} = \frac{1}{n}\sum_{i=1}^{n}\left((a_i - \frac{c_i}{\|c_i\|})\tilde{x}_i^T + (b_i - \frac{s_i c_i}{\|c_i\|})\tilde{y}_i^T\right) + \lambda(A - A_0). \tag{12}$$

Discarding the regularization term, the optimal cost can be obtained at the zero gradient, $a_i - \frac{c_i}{\|c_i\|} = 0$ and $b_i - \frac{s_i c_i}{\|c_i\|} = 0$. Just as what we constrain above that the norms of a_i and b_i should be approaching to 1. $\frac{c_i}{\|c_i\|}$ is the optimal solution for a_i in the learned feature space, while $\frac{s_i c_i}{\|c_i\|}$ is for b_i. As shown in Fig.1 (4), for a positive pair, a_i and b_i are projected to the same unit vector along the diagonal (the red solid line); for a negative pair, a_i and b_i are projected to the opposite unit vector along the other diagonal (the blue solid line).

The L-BFGS [10] optimization algorithm is adopted to compute the optimal projection matrix $A = [\bar{A}, A_s^p, A_s^c]$, which has no need to manually pick a learning rate and it is much faster compared to the standard gradient descent algorithm. Finally, the two status-specific projections can be obtained through $A^p = \bar{A} + A_s^p$ and $A^c = \bar{A} + A_s^c$ (Eqns.(5) and (6)), respectively.



(a) F-S (b) F-D (c) M-D (d) M-S

Fig. 2. ROC curves of different methods with the LBP feature on KinFaceW-II dataset.

3. EXPERIMENTS AND ANALYSIS

3.1. Settings

Datasets: In the experiments we utilized two kinship datasets, the KinFaceW-I and KinFaceW-II [6, 12] (available at: http://www.kinfacew.com). There are four kinship relations: father-son (F-S), father-daughter (F-D), mother-son (M-S) and mother-daughter (M-D), all of them are the parent-child kinship. In the two datasets, all the facial images have been aligned and cropped into 64×64 pixels to remove the photo background. As for the feature representation, we utilized the two feature descriptors, local binary patterns (LBP) [13] and histogram of gradient (HOG) [14], provided in the datasets for kinship verification.

Compared methods: Our SPML method was compared to the following representative kinship verification methods: 1) side-information based linear discriminant analysis (SILD) [11], 2) neighborhood repulsed metric learning (NRML) [6], 3) ensemble similarity learning (ESL) [8] and 4) triangular similarity metric learning (TSML) [9].

Following the settings in [6, 8, 12], we perform 5-fold cross validation based on the pre-specified training/testing split on the two kinship datasets. For training, PCA was applied to project each LBP and HOG feature into a low-dimensional feature space, and then those metric learning methods were employed to learn the projection matrices on the training datasets. Our SaPL only focuses on the positive (kin) pairs (in each triplet $s_i = 1$) in training setting. For testing, each test pair facial images were mapped into the learned feature space by the learned projection matrices, and then the cosine similarity of each test pair in the learned feature space was computed, based on which the mean accuracy and the receiver operating characteristic (ROC) curves were adopted for evaluation.

The results of SILD [11] and NRML [6] are from http://www.kinfacew.com/results.html, and the results of ESL [8] are from the original paper. We fine tuned the tradeoff parameters λ and initialized $A_0 = [I, 0, 0]$ in SaPL method to achieve the best performance.

3.2. Experimental results and analysis

Tables 1 and 2 respectively present the mean verification accuracy of different methods with LBP and HOG features on KinFaceW-I dataset, while Tables 3 and 4 respectively present those results on KinFaceW-II dataset. From those results, we can draw the following conclusions: (1) Our proposed SaPL method significantly outperforms all the compared methods with different features on the two datasets, except one case with HOG feature on KinFaceW-I obtaining a comparable results to ESL (78.3% vs 78.6% in Table 2). This demonstrates the effectiveness of our proposed SaPL method for kinship verification which mainly attribute to the introduction of status-aware projection. (2) Compared to TSML, our SaPL obtained better results in all the four cases. TSML equally treats both the parents and children facial images, which is a symmetrical model, while SaPL learns two status-specific projections for parents and children respectively, which is an asymmetrical model. It demonstrates the effectiveness of our two status-specific projections to capture more significant appearance commonality in parents and children facial image pairs, compared to only one projection.

To depict the detailed differences of verification results, Fig.2 plots the ROC curves of different methods with the LBP features on KinFaceW-II datasets. We can see that the ROC curves of our proposed SaPL are consistently higher than the other compared methods in most cases.

4. CONCLUSION

Status-aware projection learning (SaPL) has been proposed to address the asymmetrical problem caused by the different status associated with parents and children, where parents are priorly known to be significantly older than the children. SaPL learning two status-specific projections could capture more significant appearance commonality between parents and children, compared to one Mahalanobis distance metric. Each status-specific projection consists of a common component and a status-specific component. The experimental results have demonstrated the effectiveness of SaPL method for kinship verification using facial images.

5. REFERENCES

[1] A. Pashos and D. H. McBurney, "Kin relationships and the caregiving biases of grandparents, aunts, and uncles : A two-generational questionnaire study," *Human nature*, vol. 19, no. 3, pp. 311–30, 2008.

[2] K. Kim, S. H. Zarit, E. E. Femia, and J. Savla, "Kin relationship of caregivers and people with dementia: stress and response to intervention," *International journal of geriatric psychiatry*, vol. 27, no. 1, pp. 59–66, 2012.

[3] M. F. Dal Martello and L. T. Maloney, "Lateralization of kin recognition signals in the human face," *Journal of vision*, vol. 10, no. 8, pp. 9, 2010.

[4] G. Kaminski, F. Ravary, C. Graff, and E. Gentaz, "Firstborns' disadvantage in kinship detection.," *Psychological science*, vol. 21, no. 12, pp. 1746–50, 2010.

[5] J. Lampinen, J. D. Arnal, J. Adams, K. Courtney, and J. L. Hicks, "Forensic age progression and the search for missing children," *Psychology, Crime & Law*, vol. 18, no. 4, pp. 405–415, 2012.

[6] J. Lu, X. Zhou, Y. Tan, Y. Shang, and J. Zhou, "Neighborhood repulsed metric learning for kinship verification," *IEEE Trans. pattern analysis and machine intelligence*, vol. 36, no. 2, pp. 331–345, 2014.

[7] H. Yan, J. Lu, W. Deng, and X. Zhou, "Discriminative multimetric learning for kinship verification," *IEEE Trans. Information forensics and security*, vol. 9, no. 7, pp. 1169–1178, 2014.

[8] X. Zhou, Y. Shang, H. Yan, and G. Guo, "Ensemble similarity learning for kinship verification from facial images in the wild," *Information Fusion*, vol. 32, pp. 40–48, 2016.

[9] L. Zheng, K. Idrissi, C. Garcia, S. Duffner, and A. Baskurt, "Triangular similarity metric learning for face verification," in *Proc. Int'l Conf. Automatic Face and Gesture Recognition*, 2015, vol. 1, pp. 1–7.

[10] D. C. Liu and J. Nocedal, "On the limited memory bfgs method for large scale optimization," *Mathematical Programming*, vol. 45, no. 1, pp. 503–528, 1989.

[11] M. Kan, S. Shan, D. Xu, and X. Chen, "Side-information based linear discriminant analysis for face recognition," in *British Machine Vision Conference*, 2011, pp. 125.1–125.0.

[12] J. Lu, J. Hu, V. E. Liong, X. Zhou, A. Bottino, I. U. Islam, T. F. Vieira, X. Qin, X. Tan, S. Chen, et al., "The fg 2015 kinship verification in the wild evaluation," in *Proc. Int'l Conf. Automatic Face and Gesture Recognition*, 2015, vol. 1, pp. 1–7.

[13] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. pattern analysis and machine intelligence*, vol. 28, no. 12, pp. 2037–2041, 2006.

[14] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. Int'l Conf. Computer Vision and Pattern Recognition*, 2005, vol. 1, pp. 886–893.