Kinship Verification through Transfer Learning

Siyu Xia[*]
CSE, SUNY at Buffalo, USA
and Southeast University, China
xsy@seu.edu.cn

Ming Shao[*]
CSE
SUNY at Buffalo, USA
mingshao@buffalo.edu

Yun Fu
CSE
SUNY at Buffalo, USA
yunfu@buffalo.edu

Abstract

Because of the inevitable impact factors such as pose, expression, lighting and aging on faces, identity verification through faces is still an unsolved problem. Research on biometrics raises an even challenging problem—is it possible to determine the kinship merely based on face images? A critical observation that faces of parents captured while they were young are more alike their children's compared with images captured when they are old has been revealed by genetics studies. This enlightens us the following research. First, a new kinship database named UB KinFace composed of child, young parent and old parent face images is collected from Internet. Second, an extended transfer subspace learning method is proposed aiming at mitigating the enormous divergence of distributions between children and old parents. The key idea is to utilize an intermediate distribution close to both the source and target distributions to bridge them and reduce the divergence. Naturally the young parent set is suitable for this task. Through this learning process, the large gap between distributions can be significantly reduced and kinship verification problem becomes more discriminative. Experimental results show that our hypothesis on the role of young parents is valid and transfer learning is effective to enhance the verification accuracy.

1 Introduction

Face recognition [Zhao *et al.*, 2003; Jain and Li, 2005; Stone *et al.*, 2010], as an essential problem in pattern recognition and social media computing, attracts many researchers for decades. Differed from other biometric problems, e.g., fingerprint or iris based recognition, face recognition inherently relies on the un-controlled environment and inevitably suffers from degrading factors such as illumination, expression, pose and age variations. Successful approaches proposed during past decades can be found in [Turk and Pentland, 1991; Belhumeur *et al.*, 2002; Bartlett *et al.*, 2002] and meanwhile some popular databases, e.g., Yale B, CMU PIE and FERET [Jain and Li, 2005], have been set up for researchers to test or refine existing methods. Though including many unsure factors mentioned above, these databases are constructed under controlled environments with no exception.

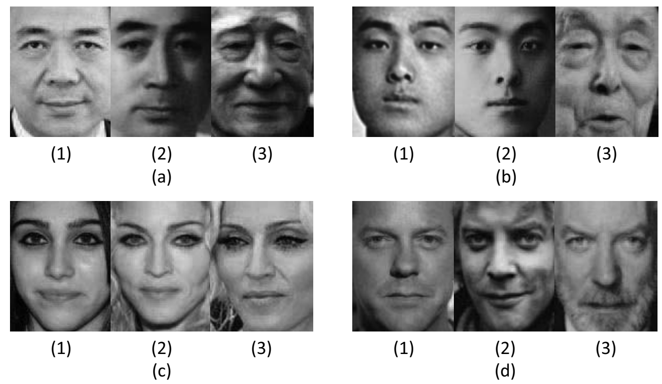

Figure 1: Four groups of typical kinship images (from (a) to (d)). Images (from (1) to (3)) represent the faces of children, their parents captured while they were young, and their parents captured while they are old, respectively.

A more unconstrained database called "Labeled Faces in the Wild" [Huang *et al.*, 2007] was set up recently, aiming at providing a data set applied to the real application. By collecting substantial typical face images from the real world, this database manages to include natural variations rather than artificial controls or designs. Challenging as it is, [Kumar *et al.*, 2009] proposed to utilize a group of local features to build two classifiers, namely, "attribute" and "simile" by which the recognition rate of this database is significantly improved. Besides, contextual information [Berg *et al.*, 2004; Gallagher and Chen, 2009] was considered as well to facilitate face recognition in the real world such as text around images, relationship of a group of people and even more, kinship [Fang *et al.*, 2010]. Figure 1 shows some typical examples of kinship images[1]. Frequently, people with kin appear in the same image, e.g., family album, and if used properly it is able to assist identification for image database management, e.g.,

[*]indicates equal contributions. This work was performed at State University of New York (SUNY) at Buffalo, USA.

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

building family trees. Meanwhile, as an important problem, kinship verification has applications in image retrieval and annotation, lost children search, etc. However, little research has been systematically conducted along this direction.

Recently, several important discoveries [Dal Martello and Maloney, 2006; Alvergne *et al.*, 2007; DeBruine *et al.*, 2009] encourage people to continue exploring in this domain. For example, facial resemblance among kinship members varies when children are aging. Analogously, a critical observation is that faces of parents captured while they were young are more alike their children's compared with images captured when they are old. This, promising in terms of statistics though, enlightens us the following research. First, a kinship database named UB KinFace consisting of child, young parent and old parent face images is collected from Internet search, most of which are with uncontrollable variations. Utilizing this database, the hypothesis in this paper based on genetic statistics is experimentally proved. Second, to eliminate the enormous divergence of distributions between children and old parents, we develop an extended transfer subspace learning approach. The key idea is to take advantage of an intermediate distribution close to both source and target distributions and naturally the young parent set is fairly suitable. Large gap between distributions can be significantly reduced through this learning process and child-old parent verification problem becomes more tractable.

The rest of the paper is organized as follows. Section 2 discusses related work. A new database and hypothesis are described in section 3. We then present the transfer subspace learning method for kinship verification. Section 4 shows experimental results. Conclusions are drawn in section 5.

2 Related Work

Differed from research based on the designed data set, real world's data generally consist of more than one impact factor that is combined in a natural way compared with other well-controlled experimental data sets. [Kumar *et al.*, 2009] proposed two methods, i.e., "attribute" and "simile", to significantly enhance the face recognition rates. The common part of their approaches, also adopted by many other papers, is to use local and structure information. As to the local information, not only color, gray value and texture but also shape, expression and action information that can be interpreted by human are extracted. Adaboost takes charge in the low-level feature selection and SVM determines if the feature exists. A group of results of such determination form a vector which can be further utilized by two methods to proceed with the final classification.

[Fang *et al.*, 2010] was the first attempt to tackle the challenge of kinship verification in the real world. To utilize local information of faces, they first localized key parts of faces by which facial features such as skin color, gray value, histogram of gradient as well as facial structure information are extracted. Then K-Nearest-Neighbor (KNN) with Euclidean metric is adopted to classify faces. Though simple the strategy is, it somehow works well for this problem. Different from the verification problem of the same person, kinship verification might not expect each feature pair is exactly the same due to genetic variations. Besides, similar parts or features on faces between kin are mostly fixed on eyes, nose, mouth, etc., according to genetic studies. Therefore, it is reasonable to concentrate only on features inherited from parents rather than the entire appearance of face. Though with these heuristic information, the kinship verification is still challenging due to tremendous differences between the query and gallery images. The most significant degrading factor, in terms of face recognition, is *aging*. Parent images used in kinship verification often contain elders and queries are usually young males or females. Texture distributions of these faces are quite different due to the aging effect, let alone the structure variations on faces from different identities. Meanwhile, in the real world, other factors are also out of control, as described in [Huang *et al.*, 2007]. These facts together lead to a complex problem in biometrics.

On the other hand, recently a more significant problem draws considerable attentions that common assumption of training and testing data from the same feature space and distribution is not always fully satisfied. This is natural for any new problems in classification. Manual labeling work is time-consuming and people hope to reuse the knowledge that has already been thoroughly explored. In such a case, knowledge transfer or transfer learning is desirable. In [Pan and Yang, 2010], several practical instances have been introduced to illustrate transfer learning role, e.g. Web-document classification, WiFi data remodeling and sentiment classification based on customers' reviews. The common thing, as described above, is how to reuse the knowledge learned before from other data or features. Specifically, transfer learning as we mentioned here can be further categorized into two classes, inductive transfer learning [Dai *et al.*, 2007a; 2007b; Quiñonero-Candela *et al.*, 2009] and transductive transfer learning [Joachims, 1999; Arnold *et al.*, 2008]. For the former one, the domains in which two sets of data embedded are the same or not, but learning targets are differed. Meanwhile, the latter one can tolerate different distributions between data sets and learning targets are quite identical. In this paper, as to our kinship verification problem, three distributions exist, namely, children, young parents and old parents. Differed from [Fang *et al.*, 2010], we introduce young parent data set as an intermediate distribution to facilitate transfer learning and clearly our approach falls into the transductive transfer learning category.

[Si *et al.*, 2010] adopted the Bregman divergence to minimize the distribution distance in the subspace. [Su *et al.*, 2010] extended this framework to the cross data set age estimation and successfully improved the accuracy by knowledge transferred from other data set. Apparently, their methods share the same learning target but different distributions and the goal of their problem is finding a common subspace. Sometimes, however, transfer learning may fail due to large discrepancy between source and target data sets [Pan and Yang, 2010]. We therefore consider introducing intermediate data as a link to abridge the divergence between source and target domains. These intermediate samples, namely face images of young parents, are close to both children and old parents and appropriately designed to fit in our formulation. Details of the database can be found in the following sections.

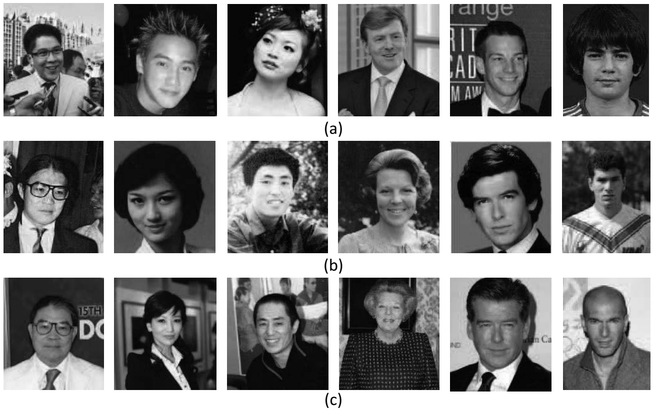

Figure 2: Example images from UB KinFace database Ver1.0. Three rows (from (a) to (c)) represent the images of children, young parents and old parents, respectively.

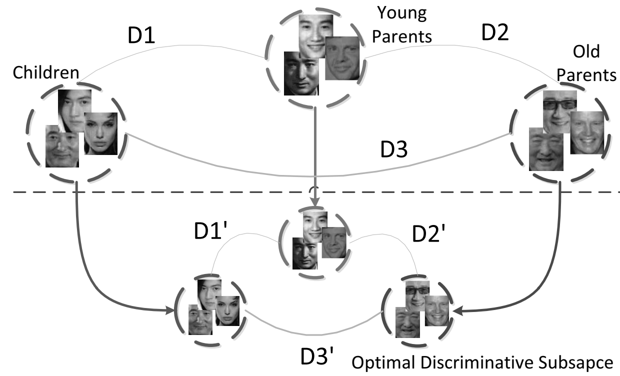

Figure 3: Red, green and blue circles represent source, intermediate and target samples. These samples are projected into an optimal discriminative subspace (under the black dot line). The distance between the distributions of the source and the target samples in the optimal discriminative subspace (marked as D3') is shorter than that in the original space (marked as D3) due to the role of the intermediate samples.

3 Our Approach

3.1 Database and Hypothesis

To the best of our knowledge, it is the first time when a database of children, young parents and old parents is collected for the purpose of kinship verification. UB KinFace[2] database Ver1.0 comprises 270 images of 180 people which can be separated into 90 groups. Each group is composed of child, young parent and old parent images. All images in the database are real-world images of public figures (celebrities in entertainment, sports and politicians) downloaded from the Internet. We simply use people's names as the query for image search. Some sample images are shown in Figure 2. Our hypothesis is that faces of parents captured while they were young are more alike their children's compared with images captured when they are old. With this assumption and its proof (see section 4.1), usage of transfer learning in the next subsection can be intuitively proposed.

3.2 Transfer Subspace Learning

To fully use the young parent set, an extended transfer subspace learning method is proposed in this paper. The proposed method aims at finding a subspace where distribution similarity of three different data sets is maximized while the low-dimensional representation is still discriminative. Essentially, our approach is different from [Si *et al.*, 2010] in that:

1) Our method adopts intermediate set to prevents the failure of transfer; and

2) Differed from the traditional transfer learning problem, ours involves source, target and intermediate sets.

Back to the kinship problem in this paper, three different data sets, namely child, young and old parent sets, are source, intermediate and target in terms of transfer learning. Intuitively, since the children and young parents possess more facial resemblance based on our assumption, transfer learning should work on children-young parents and young parents-old parents. Specifically, our objective is to seek a subspace projection W where the projected features of the intermediate and the other two domains share the same distribution. Consequently, in an implicit way, the source and the target domains are close to each other in this subspace. Figure 3 illustrates our transfer subspace learning approach. Based on these objectives, we formulate the problem as,

$$W = \arg\min_{W}\{F(W) + \lambda_{\mathbf{1}} D_W(P_L||P_U) + \lambda_2 D_W(P_L||P_V)\} \quad (1)$$

where $F(W)$ is a general subspace learning objective function, e.g., PCA [Turk and Pentland, 1991], LDA [Belhumeur *et al.*, 2002], DLA [Zhang *et al.*, 2008] and P_U, P_L and P_V represent the distributions of source samples, intermediate samples and target samples respectively. $D_W(P_L||P_U)$ and $D_W(P_L||P_V)$ are the Bregman divergence-based regularizations that measure the distance between two different distributions in the projected subspace. λ_1 and λ_2 are the weights for the regularizations. This optimization problem can be solved iteratively by the gradient descent method as follows,

$$W_{k+1} = W_k - \eta(k)(\frac{\partial F(W)}{\partial W} + \lambda_1 \sum_{i=1}^{l+u} \frac{\partial D_W(P_L||P_U)}{\partial \vec{y}_i}\frac{\partial \vec{y}_i}{\partial W} + \lambda_2 \sum_{i=1}^{l+v} \frac{\partial D_W(P_L||P_V)}{\partial \vec{y}_i})\frac{\partial \vec{y}_i}{\partial W} \quad (2)$$

where $\eta(k)$ is the learning rate, $y = W^T x$ is the low-dimensional representation of the sample and l, u, v are the numbers of samples contained in distributions of P_L, P_U, P_V. Particularly, the gradient of the first term $F(W)$ in Eq.(1) can be obtained based on the general subspace learning method.

[2] UB KinFace has been updated to Ver2.0 with more data collected when this paper was under review.

Take DLA [Zhang *et al.*, 2008] for instance, we have

$$\frac{\partial F(W)}{\partial W} = (XLX^T + (XLX^T)^T)W \qquad (3)$$

where L encapsulates the local geometry and the discriminative information and X is the matrix of input samples including all training images. This formulation essentially aims to preserve as much discriminative information as possible by assigning a margin degree to each sample. The margin degree quantifies the importance of a sample for discriminative subspace selection. Moreover, the distributions P_U, P_L and P_V in the projected subspace can be obtained by Kernel Density Estimation (KDE) technique. Then the Bregman divergence based on KDE (also see [Si *et al.*, 2010]) is given by,

$$\begin{aligned} D_W(P_L||P_U) = &\frac{1}{l^2}\sum_{s=1}^{l}\sum_{t=1}^{l} G_{\Sigma_{11}}(\vec{y}_t - \vec{y}_s) \\ &+\frac{1}{u^2}\sum_{s=l+1}^{l+u}\sum_{t=l+1}^{l+u} G_{\Sigma_{22}}(\vec{y}_t - \vec{y}_s) \\ &-\frac{2}{lu}\sum_{s=1}^{l}\sum_{t=l+1}^{l+u} G_{\Sigma_{12}}(\vec{y}_t - \vec{y}_s) \end{aligned} \qquad (4)$$

where $G_\Sigma(\vec{y})$ is the d-dimensional Gaussian kernel with the covariance matrix Σ. According to Eq.(4), the derivative of the second and third term in Eq.(1) with respect to W is

$$\begin{aligned} &\lambda_1 \sum_{i=1}^{l+u} \frac{\partial D_W(P_L||P_U)}{\partial \vec{y}_i}\frac{\partial \vec{y}_i}{\partial W} + \lambda_2 \sum_{i=1}^{l+v} \frac{\partial D_W(P_L||P_V)}{\partial \vec{y}_i}\frac{\partial \vec{y}_i}{\partial W} \\ =\ &\lambda_1 \sum_{i=1}^{l} \frac{\partial D_W(P_L||P_U)}{\partial \vec{y}_i}\vec{x}_i^T + \lambda_1 \sum_{i=l+1}^{l+u} \frac{\partial D_W(P_L||P_U)}{\partial \vec{y}_i}\vec{x}_i^T \\ &+\lambda_2 \sum_{i=1}^{l} \frac{\partial D_W(P_L||P_V)}{\partial \vec{y}_i}\vec{x}_i^T + \lambda_2 \sum_{i=l+1}^{l+v} \frac{\partial D_W(P_L||P_V)}{\partial \vec{y}_i}\vec{x}_i^T \\ =\ &\frac{2(\lambda_1+\lambda_2)}{l^2}\sum_{i=1}^{l}\sum_{t=1}^{l} G_{\Sigma_{11}}(\vec{y}_i - \vec{y}_t)(\Sigma_{11})^{-1}(\vec{y}_t - \vec{y}_i)\vec{x}_i^T \\ &-\frac{2\lambda_1}{lu}\sum_{i=1}^{l}\sum_{t=l+1}^{l+u} G_{\Sigma_{12}}(\vec{y}_t - \vec{y}_i)(\Sigma_{12})^{-1}(\vec{y}_t - \vec{y}_i)\vec{x}_i^T \\ &-\frac{2\lambda_1}{lu}\sum_{i=l+1}^{l+u}\sum_{t=1}^{l} G_{\Sigma_{12}}(\vec{y}_t - \vec{y}_i)(\Sigma_{12})^{-1}(\vec{y}_t - \vec{y}_i)\vec{x}_i^T \\ &+\frac{2\lambda_1}{u^2}\sum_{i=l+1}^{l+u}\sum_{t=l+1}^{l+u} G_{\Sigma_{22}}(\vec{y}_i - \vec{y}_t)(\Sigma_{22})^{-1}(\vec{y}_t - \vec{y}_i)\vec{x}_i^T \\ &-\frac{2\lambda_2}{lv}\sum_{i=1}^{l}\sum_{t=l+1}^{l+v} G_{\Sigma_{13}}(\vec{y}_t - \vec{y}_i)(\Sigma_{13})^{-1}(\vec{y}_t - \vec{y}_i)\vec{x}_i^T \\ &-\frac{2\lambda_2}{lv}\sum_{i=l+1}^{l+v}\sum_{t=1}^{l} G_{\Sigma_{13}}(\vec{y}_t - \vec{y}_i)(\Sigma_{13})^{-1}(\vec{y}_t - \vec{y}_i)\vec{x}_i^T \\ &+\frac{2\lambda_2}{v^2}\sum_{i=l+1}^{l+v}\sum_{t=l+1}^{l+v} G_{\Sigma_{33}}(\vec{y}_i - \vec{y}_t)(\Sigma_{33})^{-1}(\vec{y}_t - \vec{y}_i)\vec{x}_i^T \end{aligned} \qquad (5)$$

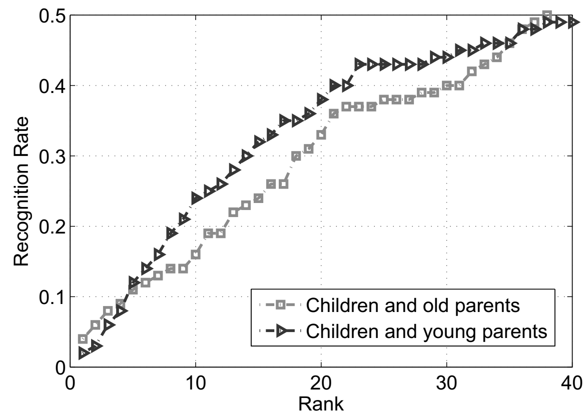

Figure 4: Cumulative Match Characteristic (CMC) for kinship classfication. Rank 1 happens when query is the nearest neighbor of the corresponding person in the gallery.

where $\Sigma_{11} = \Sigma_1 + \Sigma_1, \Sigma_{12} = \Sigma_1 + \Sigma_2, \Sigma_{13} = \Sigma_1 + \Sigma_3, \Sigma_{22} = \Sigma_2 + \Sigma_2, \Sigma_{33} = \Sigma_3 + \Sigma_3$. The iterative process of Eq.(2) can be obtained based on Eq.(3) and Eq.(5). Finally, optimized projection matrix W is achieved by this gradient descent based optimization. Through the induction of former formulas, one projection matrix is derived to connect three different distributions rather than two described in [Si *et al.*, 2010]. The experimental results in section 4 demonstrate that the verification accuracy with the intermediate domain is higher than that only with source and target domains.

4 Experiments

4.1 Pairwise Kinship Verification

Before following experiments, we conduct face and fiducial points detection on all images in the UB KinFace database to obtain cropped face images. Then we align faces according to corresponding points using an affine transform. For the whole face and its small regions in each image, we crop them to fixed sizes. A few image pre-processing techniques, e.g., histogram equalization, are implemented to mitigate irrelevant factors. A series of experiments are finely designed to verify the proposed hypothesis as well as the transfer subspace learning framework.

First, a kinship classification experiment is performed. Here, in terms of biometrics, classification means finding the proper identity for the query. Child images are used as query while young parent and old parent images are used as gallery images, respectively. Gabor [Adini *et al.*, 2002] features (5 scales and 8 directions) of the entire face are extracted after illumination normalization [Chen *et al.*, 2006]. Euclidean distance metric and Nearest-Neighbor are adopted for this task and Cumulative Match Characteristic (CMC) [Jain and Li, 2005] is shown in Figure 4. We can see that young parents possess more facial resemblance with their children than old parents and particularly it is very significant between rank 10 and 30. Clearly, this result experimentally suggests to accept our hypothesis.

Table 1: Verification results by five-fold cross validation. 1-5 means different folds. “C vs. Y” and “C vs. O” mean child-young parent and child-old parent verification respectively. “Raw” and “Struct” indicate the features used, which mean raw image and structure features respectively.

Verification 5 fold	C vs. Y (Raw)	C vs. O (Raw)	C vs. Y (Struct)	C vs. O (Struct)
1	47.22%	47.22%	50.00%	61.11%
2	58.88%	50.00%	58.33%	41.67%
3	55.56%	55.56%	69.44%	50.00%
4	55.56%	50.00%	52.78%	52.78%
5	52.78%	50.00%	52.78%	61.11%
Average	**53.89**%	50.56%	**56.67**%	53.33%

Second, we conduct kinship verification to further prove our assumption described in section 3: given two images of faces, determine if they are the true child-parent pair. In our experiments, we use the appearance as well as anthropometric models [Ramanathan and Chellappa, 2006] that are based on the measurements and proportions of the human faces for feature extraction. In order to classify the image pairs into true or false child-parent pairs, we use Euclidean distance and KNN classifier with five-fold cross validation, where 90 positive example pairs and 90 negative example pairs are used. The positive examples are true child-parent pairs and negative examples are children with randomly selected parents from non-corresponding parent images (who are not their true parents).

Experimental results of five-fold cross validation in Table 1 with raw image and structure features still supports our hypothesis. The kinship verification rate based on young parents is about three percents higher than that of based on old parents. Considering the poor quality of Internet-collected “wild” images containing young parents (It is impractical to acquire high quality images since the digital camera was not so popular back to that time), the improvements are significant enough to validate our assumption.

4.2 Kinship Verification with Transfer Learning

In this section, we proceed with tests for the proposed transfer subspace learning method. In the training phase, the inputs for the first part of transfer subspace learning are images from the intermediate domain. For the second part, children, young parent and old parent image samples are used for source domain, intermediate domain and target domain. An optimal projective matrix W can be obtained after the training phase. Similarly, we also use five-fold cross validation as in section 4.1 as our test protocol. All the image samples are projected to the optimal subspace before KNN classification with Euclidean distance metric. Additionally, we compare the proposed method with Transfer Subspace Learning (TSL) [Si *et al.*, 2010] without the intermediate domain. Moreover, to illustrate that our method does not merely utilize extra label information, i.e. young parent images, a two-step TSL based method is implemented in the comparisons as well. In this method, we take advantage of young parent images, but put them into two different learning phases, one for child-young parent and another for young parent-old parent. Therefore, two subspaces can be learned instead of one in the proposed method and we use the data in these two subspaces for the verification. Figure 5 presents results of approaches mentioned above based on Gabor (region) features over different subspace dimensions and Table 2 is a comparison of TSL and proposed method over different features within the most discriminative subspace. Note we use DLA [Zhang *et al.*, 2008] as TSL’s general subspace learning algorithm.

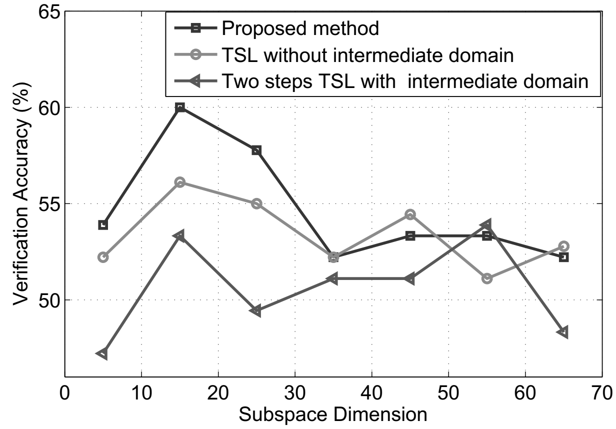

Figure 5: Verification accuracy vs. subspace dimensions.

Table 2: Verification result comparisons. “Face” means the raw image while “Region” means combination of raw image of different regions on face, i.e., eyes, nose, mouth. “Gabor” indicates the Gabor feature extracted from the corresponding raw image for verification.

Different features	Proposed method	TSL without intermediate domain
Face	55.00%	51.11%
Region	55.33%	52.78%
Face (Gabor)	56.11%	52.78%
Region (Gabor)	**60.00**%	**56.11**%

4.3 Discussion

From above experimental results, it can be concluded that the appearance variance is so significant that the child-parent verification is challenging. Nevertheless, our method outperforms TSL without young parent samples, which shows that the intermediate domain indeed functions as a bridge to reduce the difference between the source domain and the target domain. In addition, we find that the facial region information is more effective than the whole face and Gabor feature can improve the results as well. These phenomenons are consistent with the common sense that human tends to recognize the kinship more efficiently by facial region features, e.g., eyes, nose, etc.

Compared with the result of [Fang *et al.*, 2010], our experimental protocol is more challenging. First, the impacts of pose, lighting and aging of faces in UB KinFace database are more significant. Second, color information is not used in our experiment due to the poor quality of young parent images.

5 Conclusion and Future work

In this paper we exploit the kinship verification task through face images. We have collected the new UB KinFace database from Internet search. We also propose a transfer subspace learning method including the young parent set as the intermediate domain whose distribution is close to both source and target distributions. Experimental results support our hypothesis on the role of young parents and demonstrate transfer learning can further enhance the verification accuracy.

Future work includes extending the UB KinFace database and seeking more robust features to improve kinship verification. With more training or testing samples involved, chance factors will be eliminated and performance will be more robust. Local features not only capturing the kinship characteristics but also invariant to illumination, poses, etc., will be preferred in our future work. In addition, ethnicity classification should be considered in that facial resemblance of kinship varies over different races. Subjective test will also be conducted to compare the human performance and machine performance on kinship verification.

Acknowledgements

This work was partially supported by the SUNY Buffalo faculty startup funding and National Natural Science Foundation of China under Grant 90820009. We thank Si Si and Dacheng Tao for providing the source code and valuable discussions.

References

[Adini *et al.*, 2002] Y. Adini, Y. Moses, and S. Ullman. Face recognition: The problem of compensating for changes in illumination direction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):721–732, 2002.

[Alvergne *et al.*, 2007] 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.

[Arnold *et al.*, 2008] A. Arnold, R. Nallapati, and W.W. Cohen. A comparative study of methods for transductive transfer learning. In *IEEE International Conference on Data Mining Workshops*, pages 77–82, 2008.

[Bartlett *et al.*, 2002] M.S. Bartlett, J.R. Movellan, and T.J. Sejnowski. Face recognition by independent component analysis. *IEEE Transactions on Neural Networks*, 13(6):1450–1464, 2002.

[Belhumeur *et al.*, 2002] 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*, 19(7):711–720, 2002.

[Berg *et al.*, 2004] T. L. Berg, A. C. Berg, J. Edwards, M. Maire, R. White, Y.-W. Teh, E. Learned-Miller, and D. A. Forsyth. Names and faces in the news. In *IEEE conference on Computer vision and pattern recognition*, pages 848–854, 2004.

[Chen *et al.*, 2006] T. Chen, W. Yin, X.S. Zhou, D. Comaniciu, and T.S. Huang. Total variation models for variable lighting face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(9):1519–1524, 2006.

[Dai *et al.*, 2007a] W. Dai, G.-R. Xue, Q. Yang, and Y. Yu. Transferring naive bayes classifiers for text classification. In *AAAI*, pages 540–545, 2007.

[Dai *et al.*, 2007b] W. Dai, Q. Yang, G.R. Xue, and Y. Yu. Boosting for transfer learning. In *International conference on Machine learning*, pages 193–200. ACM, 2007.

[Dal Martello and Maloney, 2006] M.F. Dal Martello and L.T. Maloney. Where are kin recognition signals in the human face? *Journal of Vision*, 6(12):1356–1366, 2006.

[DeBruine *et al.*, 2009] L.M. DeBruine, F.G. Smith, B.C. Jones, S. Craig Roberts, M. Petrie, and T.D. Spector. Kin recognition signals in adult faces. *Vision Research*, 49(1):38–43, 2009.

[Fang *et al.*, 2010] R. Fang, K.D. Tang, N. Snavely, and T. Chen. Towards computational models of kinship verification. In *IEEE International Conference on Image Processing*, pages 1577–1580, 2010.

[Gallagher and Chen, 2009] A.C. Gallagher and T. Chen. Understanding images of groups of people. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 256–263, 2009.

[Huang *et al.*, 2007] 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.

[Jain and Li, 2005] A.K. Jain and S.Z. Li. *Handbook of face recognition*. Springer, 2005.

[Joachims, 1999] T. Joachims. Transductive inference for text classification using support vector machines. In *International Conference on Machine Learning*, pages 200–209, 1999.

[Kumar *et al.*, 2009] N. Kumar, A.C. Berg, P.N. Belhumeur, and S.K. Nayar. Attribute and simile classifiers for face verification. In *IEEE International Conference on Computer Vision*, pages 365–372, 2009.

[Pan and Yang, 2010] S. J. Pan and Q. Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, 2010.

[Quiñonero-Candela *et al.*, 2009] J. Quiñonero-Candela, M. Sugiyama, A. Schwaighofer, and N.D. Lawrence. *Dataset Shift in Machine Learning*. The MIT Press, 2009.

[Ramanathan and Chellappa, 2006] N. Ramanathan and R. Chellappa. Modeling age progression in young faces. pages 387–394, 2006.

[Si *et al.*, 2010] S. Si, D. Tao, and B. Geng. Bregman divergence-based regularization for transfer subspace learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(7):929–942, 2010.

[Stone *et al.*, 2010] Z. Stone, T. Zickler, and T.J. Darrell. Toward large-scale face recognition using social network context. *Proceedings of the IEEE*, 98(8):1408–1415, 2010.

[Su *et al.*, 2010] Y. Su, Y. Fu, Q. Tian, and X. Gao. Cross-database age estimation based on transfer learning. In *IEEE International Conference on Acoustics Speech and Signal Processing*, pages 1270–1273, 2010.

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

[Zhang *et al.*, 2008] T. Zhang, D. Tao, and J. Yang. Discriminative locality alignment. In *European Conference on Computer Vision*, pages 725–738. Springer, 2008.

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