Tri-Subjects Kinship Verification: Understanding the Core of A Family

Xiaoqian Qin[1,2], Xiaoyang Tan[1] and Songcan Chen[1]
[1]Department of Computer Science and Technology
Nanjing University of Aeronautics and Astronautics
#29 Yudao Street, Nanjing 210016, P.R.China. Email: x.tan@nuaa.edu.cn
[2]Huaiyin Normal University, #111 Changjiang Road, Huaian 223300, P.R.China

Abstract

Recent research has demonstrated that computer vision algorithms have understood individual face image fairly well. However, one major challenge in computer vision is to go beyond that and to investigate the bi- or tri- relationship among multiple visual entities, answering such questions as whether a child in a photo belongs to given parents. Indeed parents-child relationship plays a core role in a family and understanding such kind of kin relationship would have fundamental impact on the behavior of an artificial intelligent agent working in a human being world. In this work, we tackle the problem of tri-subjects kinship verification by effectively exploiting the dependence structure between child and parents in a few aspects: similarity measure, feature selection and classifier design. State-of-the-art results are reported with the proposed method on our newly released kinship database characterized by over 1,000 parents-child groups.

1 Introduction

Kinship verification from facial images is an emerging problem in computer vision. Applications of kin relationships include increasing face recognition rates [1] [2], social media analysis [3], and so on.

Besides its wide applications, kinship learning is also motivated by the long-term goal of computer vision to go beyond the understanding of a single visual entity and to that of the bi- or tri- relationship among multiple visual entities, answering such questions as whether a child in a photo belongs to given parents. Actually, recent research has demonstrated that computer vision algorithms have understood individual face image fairly well - the best result on the challenging LFW verification database has reached an accuracy as high as 99.15% [4]. However, extending those techniques to handle the complex relationship among multiple entities is not trivial. One major reason is due to the fact the appearance gap encountered for example in a kinship problem is much larger than that in a conventional face recognition setting. In this case, one may be given two face images with different sex and different ages.

In this sense, kinship learning is a step towards such a trend to capture mutual information among different visual entities, particularly multiple face images. Most of current research on this [5][6][7][8][9][10], however, focus on the kinship involving only two subjects such as father-son or mother-daughter, while in practice, kin relationship involving more subjects are desirable. Motivated by this, Fang et. al. [11] collected the Family101 kinship dataset for more general family membership classification beyond father and son.

Inspired by Fang et al's work [11], in this paper we focus on tri-subjects kinship verification(i.e., parents-son or parents-daughter). This is at a lower level to the whole-family relationship, as parents-child is the core and basic unit formed in a family.

We propose a new discriminative bilinear classifier to model the similarity between the parents and the child, with the dependence between them captured by a covariance-like matrix learnt from the data. The key idea of our method is based on the observation that compared to the case with only one image from one of the parents, images from both parents could provide richer information about the kinship relation regards to a child, since there exists genetic overlapping between both parents and the child. Furthermore, we propose a vote-based feature selection method, which jointly selects the most discriminative features for the parents-child pair, while taking local spatial information into account. Finally, we release a new tri-subjects kinship database, characterized by over 1,000 parents-child groups. State-of-the-art results are achieved using our method. Interestingly, our experimental results also show that the accuracy of bi-kinship verification could be benefit from reformulating it as a specific case of tri-kinship problem.

2 Related Work

Several attempts have been made to develop machine learning approaches to kinship verification.

Various feature descriptors are extracted for determining kinship similarity, including the skin and local/holistic texture [5], DAISY [12], GGOP [3], SSR [13], attributes [7], gated autoencoders[14] and dense stereo matching[2]. [9] propose to describe facial appearance over smile expression.

LU et al. [8] introduce NRML by learning a distance metric under which the samples with a kinship relation are pulled close and those without a kinship relation are pushed away.

While most the above works focus on the bi-subjects kinship verification, Fang et al. [11] extend this to more general family membership classification, i.e., given a query face image, asking which family it belongs to. This work is closely related to ours but we focus on the verification of more basic unit that forms a family, that is, the parents-child relationship. The methods developed here can be potentially extended to handle the entire family by treating a family tree as an ensemble of tri-relationships.

3 Tri-Subjects Kinship Verification

Assume that we are given a set of N training samples $\{(x_i^f, x_i^m, x_i^c, y_i)\}_{i=1}^N$, where $x_i^f, x_i^m, x_i^c \in R^d$ respectively denotes the i-th sample of a father, a mother and a child, d is the dimension of the feature representation of a sample, and $y_i \in \{+1, -1\}$ indicates whether this child is produced by the corresponding two parents. Our goal is to learn a function $f : (x^f, x^m, x^c) \rightarrow \{+1, -1\}$ to verify whether such a kinship is validated for three previously never seen images (x^f, x^m, x^c).

3.1 Discriminative bilinear method for tri-subjects verification

The verification function $f(x^f, x^m, x^c)$ is modeled as a three-layered network-like architecture. Particularly, middle-level feature descriptors from each image serve as the input layer of the network. The second layer is composed of two groups of nodes roughly corresponding to the "covariance" between x_i^f and x_j^c and that of x_i^m and x_j^c, respectively. Finally, these covariance factors are linearly combined in the third layer to make the prediction after a sigmoid transformation.

Note that the nodes of the second layer encode the similarity between parents and a child. For this, we explore two ways. The first, we decompose the triples of (x^f, x^m, x^c) into two pairs (x^f, x^c) and (x^m, x^c), and hence the pairwise similarity between them is,

$$\begin{aligned} s^f(x^f, x^c) &= (x^f)^T W_f x^c \\ s^m(x^m, x^c) &= (x^m)^T W_m x^c \end{aligned} \tag{1}$$

where the transformation matrix W_f, W_m essentially encode the "covariance" relationship between a parent and a child. Since both W_f, W_m are $d \times d$ matrix, we call this Symmetric Bilinear Model (SBM) inspired by the definition of mahalanobis distance.

Alternatively, one can treat the parents and the child as samples from two domains. Let us denote the parents domain as $\mathcal{P}$, with data points $(x_1^f, x_1^m), (x_2^f, x_2^m), ..., (x_N^f, x_N^m)$, and the child domain as $\mathcal{C}$, with data points $x_1^c, x_2^c, ..., x_N^c$. With these notations, one can model the similarity between a child x^c and his/her parents $x^p = (x^f, x^m)$ as,

$$s^p(x^p, x^c) = (x^p)^T W_p x^c \tag{2}$$

where W_p is a $2d \times d$ matrix. This model is called Asymmetric Bilinear Model (ABM) in what follows.

To this end, our tri-subjects kinship verifier can be modeled as follows. Particularly, for the SBM similarity measure, we have,

$$f(x^f, x^m, x^c) = sign(\beta_1 s^f(x^f, x^c) + \beta_2 s^m(x^m, x^c) + b) \tag{3}$$

where the combination coefficients β_1 and β_2 are two scalars and b is the similarity threshold term. All these parameters are learnt using a linear SVM.

For the ABM similarity measure, our verifier is defined as follows,

$$f(x^f, x^m, x^c) = \sigma(s^p(x^p, x^c) + b) \tag{4}$$

where σ is the sigmoid function.

The parameters $\{W_p, b\}$ are learnt using the following regularized logistic regression objective,

$$\min_{W_p, b} \sum_{i=1}^{N} log(1 + exp(-y_i((x_i^p)^T W_p x_i^c) + b)) + \lambda \|W_p\|_* \tag{5}$$

where b is the threshold, and $\|W_p\|_*$ is the trace norm. With appropriate parameter λ, the trace norm shall force a solution with many singular value of W_p being exactly zero. This allows a more compact representation of the data, thus being useful especially when the original feature space is high-dimensional. See [15] for details on an efficient implementation of this.

3.2 Vote-based feature selection

Particularly, our algorithm has two steps. In the first step, we partition an image into overlapping patches, extract a middle-level feature descriptor from each patch, and utilize $a \in R^{2d}$ to represent the parent-child concatenated feature vector, then obtain a weight vector u using a sparse l_1 regularized logistic regression objective [16]. We use u^p and u^c to denote the first half and the second half of u respectively.

Then, we obtain the patches votes for parent through $v_k^p = \sum_{j \in k} u_j^p$, where u_j^p denotes the j-th element of vector u^p corresponding to patch k. After this, we select the first K patches with high v_k^p value for feature representation. Similarly, we obtain the first K patches with high v_k^c value for child feature. Note that, we use the union of selected patches from parent and child to represent each face.

3.3 Variants of our proposed method

The variants of our method include with/without feature selection and working at the block level i.e., selecting the most discriminative patches first, then learning the "covariance" relationship and making verification predictions based on each selected patches, and finally aggregating these meta-decisions through linear SVM for the final verification judgement.

4 The TSKinFace Database and Evaluation Protocol

We have constructed a new kinship database named TSKinFace (Tri-Subjects Kinship Face Database) containing 1015 families, with 513 Father-Mother-Son families and 502 Father-Mother-Daughter families, including 2589 individuals. All images in the database are harvested from the internet based on knowledge of public figures family and photo sharing social network. During images collecting, we impose no restriction in terms of pose, lighting, expression, background, race etc. And the database includes around 66% Asians and 34% non-Asians to guarantee a diverse race distribution.This database will be made publicly available online[1].

The key difference between our database and other existing kinship database lies in our inclusion of family. Note that the Family101 database also has family structure, but does not focus on the core parents-child kin relationship. Another difference is that while

---

[1] Available at:
http://parnec.nuaa.edu.cn/xtan/data/TSKinFace.html

the number of images per family in Family101 is much larger than ours, the number of families contained in our database is about five times more than that of the Family101 database. Hence the two databases are complimentary to each other. Fig. 1 shows some image groups of parents-child pair from our TSKinFace database.

Figure 1. Some family image groups of our TSKinFace database, where each group consists of a family triple of a father, a mother and a child. The first row shows three Father-Mother-Daughter (FM-D) relation families, respectively and the second row are three Father-Mother-Son (FM-S) relation families, accordingly.

All images are cropped and scaled to a size of 64×64 grayscale pixels. In our experiments, each face image was divided into 7×7 overlapping patches and the size of each patch is 16×16. For each patch, we extracted a 128-dimensional SIFT feature which is adopted as our default feature descriptor for all experiments.

We design a verification protocol following [6] and [8]: the database is equally divided into five folds, which facilitates five-fold cross validation experiments. For face images in each fold, we consider all groups of face images with kinship relation as positive samples, while the negative samples are a random combination with a child image and two parents images subjected to the constraint that the child was not produced by them. In our experiments, each couple and child images appeared only once in the negative samples.

5 Experiments

5.1 Tri-subjects kinship verification

To tune the parameters of our method, we applied cross validation strategy to seek the optimal parameter. As a result, λ (E.q. 5), λ for block level and trade-off parameter for feature selection is set to be 5.0 , 0.1 and 0.08 respectively.

We compare our method to the following baselines:

Our first baseline is SVM. For this, we concatenate the feature vectors of the group and tune the parameters to reach the best possible performance.

Our second baselines include two classical metric learning algorithms i.e., ITML [17] and LMNN [18] and existing state-of-the-art bi-subjects kinship verification model (NRML [8] and gated autoencoder [14]) as the base models to score the similarity between a child and his/her parents separately, and then train a linear SVM over these to make the final prediction (c.f., E.q. 4).

Third, we follow [11] to treat images belonging to the same family as a group and to build a linear SVM-based kinship verifier by constructing a reconstruction errors-based representation for each face using sparse group lasso [11].

Table 1 summarizes our experimental results. One can see from the table that our proposed "SBM-block w FS" performs best. Particularly, when adopting the state of the art bi-subjects kinship verification for tri-subjects verification, the best performer is the Gated autoencoder [14].

Our method also significantly works better than the sparse group lasso based method proposed by [11] - one possible explanation for this is that for a core family group involved only three subjects, the assumption made in [11] that an image of a child could be best reconstructed by the images from his/her own parents may be too strong.

Thirdly, we see that simply adopting state of the art metric learning methods or learning a binary classifier through concatenating the tri-subjects feature vector for tri-subjects kinship verification is not enough as well, partly due to the fact that they fail to model the dependence structure among the three visual entities.

Table 1. Correct verification rates(%) for different methods on the TSKinFace database.

Method	FM-S	FM-D	avg.
Concatenated+SVM	53.5±0.2381	53.2±0.2037	53.4
Sparse Group Lasso[11]	71.6±0.9644	69.8±0.3485	70.7
NRML [8]	77.0±0.5831	71.4±0.5933	74.2
Gated autoencoder [14]	81.9 ±0.4433	79.6±0.3685	80.8
ITML [17]	76.6±0.3753	71.4±0.4087	74.0
LMNN [18]	75.4±0.7293	70.3±0.7372	72.9
ABM w/o FS (*proposed*)	78.5±0.3411	73.2±0.3888	75.9
ABM w FS (*proposed*)	78.6±0.3114	76.9±0.2927	77.8
ABM-block w FS (*proposed*)	83.4±0.2508	81.9±0.3025	82.7
SBM w/o FS (*proposed*)	82.4±0.3568	78.2±0.4105	80.3
SBM w FS (*proposed*)	82.8±0.2608	79.5±0.2550	81.2
SBM-block w FS (*proposed*)	**85.2±0.3031**	**83.5±0.2985**	**84.4**

5.2 Effectiveness of the vote-based feature selection

We compare our feature selection scheme with lasso and group lasso (GL). Table 2 gives the results. The mean performance of "GL" is better than L1, revealing that learning patches works better. Furthermore, the proposed feature selection method can reach a better performance than "GL", on average improving the performance by about 2.3% and 0.4%, respectively on both tasks.

Table 2. Correct verification rates(%) for different feature selection methods on the TSKinFace database (where "FS" denotes our vote based feature selection method while "L1" denotes lasso and "GL" denotes group lasso)

Method	FM-S	FM-D	avg.
ABM w L1	75.9±0.5100	74.8±0.5716	75.4
ABM w GL	76.0±0.4300	75.0±0.5523	75.5
ABM w FS (*proposed*)	78.6±0.3114	76.9±0.2927	77.8
SBM w L1	78.3±0.4980	79.2 ± 0.5629	78.8
SBM w GL	80.6±0.4972	80.9±0.5356	80.8
SBM w FS (*proposed*)	82.8±0.2608	79.5±0.2550	81.2

5.3 Enhancing bi-subjects kinship verification


(a)
(b)

Figure 2. It is difficult to determine the MD relation in (a), if the observations are limited to the pairs only. Intuitively, the ambiguity may decreases when the face of father is considered simultaneously as shown in (b), treating (a) as tri-subjects kinship verification and verifying the FM-D relation. Note that, for (b), since once a FM-D relationship is established, a MD and a FD relationship *must* be established as well.

Intuitively, having more information about one's parents is potentially useful to improve the performance of bi-subjects kinship verification, as shown in Fig. 2. Another series of experiments are conducted to verify this hypothesis. Table 3 compares the results of these two approaches for bi-subjects kinship

Table 3. Correct rates (%) of different methods for bi-subjects kinship verification with triple inputs(column 2 and 5) and pair inputs (column 3, 4, 6 and 7. "*" denotes that the result (P-values) of $t-$test for the performance comparison between pair inputs and triple inputs verification is less than 0.05).

Method	FM-S	FS	MS	FM-D	FD	MD
Sparse Group Lasso[11]	71.6±0.9644	69.1±0.6093	68.7±1.2204	69.8±0.3485	66.8±0.4627(*)	67.9±0.5977
NRML [8]	77.0±0.5831	74.8±0.7279(*)	72.2±0.3360(*)	71.4±0.5933	70.0±0.6716(*)	71.3±0.5853
Gated autoencoder [14]	81.9±0.4433	79.9±0.6790(*)	78.5±0.5963(*)	79.6±0.3686	74.2±0.3170(*)	76.3±0.2296(*)
ITML [17]	76.6±0.3753	75.6±0.3866(*)	72.1±0.3330(*)	71.4±0.4087	70.5±0.4000(*)	70.7±0.4435(*)
LMNN [18]	75.4±0.7293	72.7±0.7305	71.5±0.7455(*)	70.3±0.7372	69.8±0.7243(*)	70.1±0.3846
ABM-block w FS (*proposed*)	83.4±0.2508	83.0±0.5558	82.8±0.5037	81.9±0.3025	80.5±0.4301	81.1±0.4003
SBM-block w FS (*proposed*)	85.2±0.3031	83.0±0.5558(*)	82.8±0.5037(*)	83.5±0.2985	80.5±0.4301(*)	81.1±0.4003(*)

verification. One can see that treating this as a tri-subjects kinship verification problem hence potentially exploiting more information about one's parents is indeed beneficial. t-test analysis shows that this improvement is statistically significant. It is well known that a problem like MS or FD verification is quite difficult due to the different genders of two subjects to be verified. Our method essentially provides a new solution to this, and we consider it as one of the major motivations to study the tri-subjects kinship verification problem.

6 Conclusions

In this work, we made the first attempt to investigate the tri-subjects kinship verification problem. For this we proposed a novel symmetric bilinear model (SBM) and a vote-based feature selection method, both incorporate prior knowledge about the dependence structure between a child and his/her two parents. Various experiments on our newly collected database "T-SKinFace" characterized by over 1,000 groups of triples demonstrate the effectiveness of our proposed method. Our experimental results also reveals that the proposed method could be used to significantly boost the performance of bi-subjects kinship verification.

Acknowledgment

This work was supported by the National Science Foundation of China (61073112, 61373060), Jiangsu Science Foundation ($BK2012793$), Qing Lan Project, Research Fund for the Doctoral Program (RFDP)(20123218110033), the Natural Science Foundation of the Jiangsu Higher Education Institutions of China ($13KJD520002$).

References

[1] G. Wang, A. Gallagher, J. Luo, D. Forsyth, Seeing people in social context: Recognizing people and social relationships, in: Computer Vision–ECCV 2010, Springer, 2010, pp. 169–182.

[2] M. Shao, S. Xia, Y. Fu, Identity and kinship relations in group pictures, in: Human-Centered Social Media Analytics, Springer, 2014, pp. 175–190.

[3] X. Zhou, J. Lu, J. Hu, Y. Shang, Gabor-based gradient orientation pyramid for kinship verification under uncontrolled environments, in: Proceedings of the 20th ACM international conference on Multimedia, ACM, 2012, pp. 725–728.

[4] Y. Sun, X. Wang, X. Tang, Deep learning face representation by joint identification-verification, arXiv preprint arXiv:1406.4773.

[5] R. Fang, K. D. Tang, N. Snavely, T. Chen, Towards computational models of kinship verification, in: Image Processing (ICIP), 2010 17th IEEE International Conference on, IEEE, 2010, pp. 1577–1580.

[6] S. Xia, M. Shao, Y. Fu, Kinship verification through transfer learning, in: Proceedings of the Twenty-Second international joint conference on Artificial Intelligence-Volume Volume Three, AAAI Press, 2011, pp. 2539–2544.

[7] S. Xia, M. Shao, Y. Fu, Toward kinship verification using visual attributes, in: Pattern Recognition (ICPR), 2012 21st International Conference on, IEEE, 2012, pp. 549–552.

[8] J. Lu, X. Zhou, Y.-P. Tan, Y. Shang, J. Zhou, Neighborhood repulsed metric learning for kinship verification, IEEE Trans. Pattern Anal. Mach. Intell. 36 (2) (2014) 331–345.

[9] H. Dibeklioglu, A. A. Salah, T. Gevers, Like father, like son: Facial expression dynamics for kinship verification, in: IEEE International Conference on Computer Vision, 2013, pp. 1497–1504.

[10] H. Yan, J. Lu, W. Deng, X. Zhou, Discriminative multi-metric learning for kinship verification, Information forensics and security, IEEE Transactions on 9 (7) (2014) 1169–1178.

[11] R. Fang, A. C. Gallagher, T. Chen, A. Loui, Kinship classification by modeling facial feature heredity, in: ICIP'13, 2013.

[12] G. Guo, X. Wang, Kinship measurement on salient facial features, Instrumentation and Measurement, IEEE Transactions on 61 (8) (2012) 2322–2325.

[13] N. Kohli, R. Singh, M. Vatsa, Self-similarity representation of weber faces for kinship classification, in: Biometrics: Theory, Applications and Systems (BTAS), 2012 IEEE Fifth International Conference on, IEEE, 2012, pp. 245–250.

[14] A. Dehghan, E. G. Ortiz, R. Villegas, M. Shah, Who do i look like? determining parent-offspring resemblance via gated autoencoders, in: Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on, IEEE, 2014.

[15] S. Ji, J. Ye, An accelerated gradient method for trace norm minimization, in: Proceedings of the 26th Annual International Conference on Machine Learning, ACM, 2009, pp. 457–464.

[16] J. Liu, J. Chen, J. Ye, Large-scale sparse logistic regression, in: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2009, pp. 547–556.

[17] J. V. Davis, B. Kulis, P. Jain, S. Sra, I. S. Dhillon, Information-theoretic metric learning, in: Proceedings of the 24th international conference on Machine learning, ACM, 2007, pp. 209–216.

[18] K. Weinberger, J. Blitzer, L. Saul, Distance metric learning for large margin nearest neighbor classification, Advances in neural information processing systems 18 (2006) 1473.