Kinship Verification from Facial Images Under Uncontrolled Conditions *

Xiuzhuang Zhou[§,‡], Junlin Hu[*], Jiwen Lu[†], Yuanyuan Shang[§,‡], Yong Guan[§,‡]
[§]College of Information Engineering, Capital Normal University, Beijing, China
[‡]Beijing Engineering Research Center of High Reliable Embedded System, Capital Normal University, China
[*]College of Information Science and Technology, Beijing Normal University, Beijing, China
[†]School of Electrical and Electronics Engineering, Nanyang Technological University, Singapore
E-mail: zxz@xeehoo.com; hujunlin@msn.com; lujiwen@pmail.ntu.edu.sg; syy@bao.ac.cn;
guanyong@mail.cnu.edu.cn

ABSTRACT

In this paper, we present an automatic kinship verification system based on facial image analysis under uncontrolled conditions. While a large number of studies on human face analysis have been performed in the literature, there are a few attempts on automatic face analysis for kinship verification, possibly due to lacking of such publicly available databases and great challenges of this problem. To this end, we collect a kinship face database by searching 400+ pairs of public figures and celebrities from the internet, and automatically detect them with the Viola-Jones face detector. Then, we propose a new spatial pyramid learning-based (SPLE) feature descriptor for face representation and apply support vector machine (SVM) for kinship verification. The proposed system has the following three characteristics: 1) no manual human annotation of face landmarks is required and the kinship information is automatically obtained from the original pair of images; 2) both local appearance information and global spatial information have been effectively utilized in the proposed SPLE feature descriptor, and better performance can be obtained than state-of-the-art feature descriptors in our application; 3) the performance of our proposed system is comparable to that of human observers.

Categories and Subject Descriptors

I.4.9 [**Computing Methodologies**]: Image Processing and Computer Vision–Applications

General Terms

Algorithm, Performance, Experimentation.

Keywords

Face Analysis, Kinship Verification, Image Descriptors.

*Area chair: Mor Naaman.

Figure 1: Samples of our kinship database. From top to bottom are the Father-Son (FS), Father-Daughter (FD), Mother-Son (MS) and Mother-Daughter (MD) kinship relations, respectively.

1. INTRODUCTION

Facial image analysis have been widely investigated in the computer vision and multimedia computing community, and a large number of such methods have been proposed for various practical applications, such as face recognition [12], facial expression recognition [6], gender classification [9], ethnic classification [8] and human age estimation [7]. While great successes have been made in these areas, there are a few attempts on automatic face analysis for kinship verification, possibly due to lacking of such publicly available databases and great challenges of this problem.

Recently, Fang *et al.* [5] attacked the challenge of kinship verification by using facial feature extraction and selection methods and their empirical results has verified this possibility. While encouraging results were obtained, there are still two shortcomings of their method:

- Their method was evaluated on a comparatively small dataset (150 pairs), such a small dataset may be insufficient to demonstrate the effectiveness of face image analysis-based kinship verification.
- Facial images in [5] were collected under a controlled environment such that only upright frontal images with normal illumination and natural facial expression were accepted for verification. In many real world applications, there may be some variations on pose, expression, occlusion and illumination of the acquired face images, and it is of great significant to investigate kinship verification under uncontrolled conditions.

Different from previous work in [5], the main purpose of this research is to drive the kinship verification from facial images research more towards real scenarios. Instead of being limited to well-cropped faces with costly manual annotations, we focus on fully-automatic kinship verification from face images under uncontrolled conditions that imposes no restrictions in terms of pose, lighting, background, expression and ethnicity on the face images (Figure 1 shows several example images), such that a fully automatic and real-time system which takes general face images as inputs can be achieved. Moreover, we further propose a new spatial pyramid learning-based (SPLE) feature descriptor for face representation. Since SPLE can utilize both local appearance and global spatial face information for feature representation, better performance can be obtained than state-of-the-art feature descriptors in our system. Experimental results are presented to demonstrate the efficacy of our proposed approach.

2. PROPOSED APPROACH

To automatically classify kinship relation from facial images, we are facing two questions. First, what characteristics are crucial for kinship verification? Second, how to make decisions based on these characteristics? To answer these questions, we will discuss several feature representation methods and present a new one in Section 2.1 together the SVM classifier in Section 2.2.

2.1 Representation

Raw pixel is the most straightforward representation for face images. However, it may be not a good choice for face analysis task because it usually suffers from the illumination and expression variations. To overcome these effects, many local descriptors have been proposed in the literature. For example, Ahonen *et al.* [1] proposed a local binary pattern (LBP) method to describe the micro-structure of the face. Since LBP is encoded by a handcrafted design, many LBP variants [1, 11] have been proposed to improve the performance of the original LBP method. However, these methods still apply the handcrafted strategy, and it is very difficult to obtain an optimal encoding method manually. To address this problem, Cao *et al.* [3] recently proposed a learning-based method by using a unsupervised learning approach to encode the local micro-structures of the face into a set of discrete codes. While better performance can be obtained, this feature descriptor still suffer from one shortcoming: the spatial information were totally discarded in the LE method because a conventional clustering technique, e.g. k-means, was used to obtain the histogram and the spatial information cannot be effectively reflected in such feature histograms.

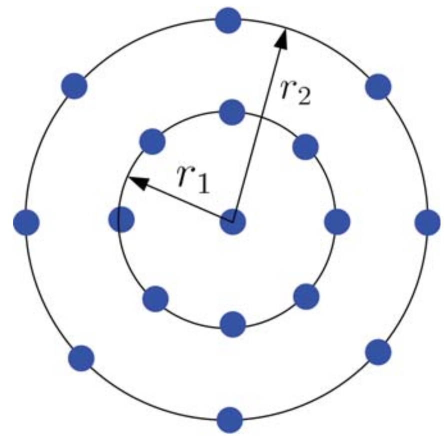

Figure 2: The sampling method used in our SPLE method.

To make better use of the spatial information, we propose a spatial pyramid learning-based (SPLE) feature descriptor for face representation, works as below

Step 1: For each pixel, we sample its neighboring pixels in a ring-based pattern to form a low-level feature vector. Similar to LE, we sample $r * 8$ pixels at even intervals on the ring of radius r. In our experiments, we empirically set $r_0 = 0$, $r_1 = 1$ and $r_2 = 2$, such that 25 (1+8+16) pixels are sampled to construct a feature vector for each pixel, as shown in Figure 2.

Step 2: Normalize the sampled feature vector into unit length such that it can be more robust to illumination variant.

Step 3: Randomly select $2N_1$ pairs of face images to construct a training set, in which N_1 pairs are with kinship relations and the other N_1 pairs are not with kinship relations, respectively.

Step 4: Perform K-means on the the training set to quantize all feature vectors into M discrete types, and make the simplifying assumption that only features of the same type can be matched to one another. Now, for each face image with size of $p \times q$, it will be encoded as a $1 \times M$ histogram-type feature vector.

Step 5: Construct a sequence of grids at resolution $0, \cdots, L$, such that the grid at level l has 2^l cells along each dimension. We extract the LE feature in each cell and concatenate them into a long feature vector. In our experiments, M and L are empirically set to be 200 and 2, respectively, such that the final feature vector is 4200-dimensional (200 + 200*4+200*16).

2.2 Classifier

After obtaining feature representation of each face image, different classifiers can be applied. Since our kinship verification is a binary classification problem and support vector machine (SVM) has demonstrated excellent performance for such tasks, we here apply SVM for classification. Different from the previous SVM-based face verification approaches which use the difference of any two positive or negative face samples as features, we use an normalized absolute histogram difference (NAHD) of a pair of samples as features for SVM learning.

Let $F_1 = [F_{11}, F_{12}, \cdots, F_{1D}]$ and $F_2 = [F_{21}, F_{22}, \cdots, F_{2D}]$

be two SPLE feature representations of two face images I_1 and I_2, we define NAHD as follows:

$$V(I_1, I_2) = \frac{|F_1 - F_2|}{\sum_{t=1}^{D} |F_{1t} - F_{2t}|} \quad (1)$$

We can observe that the elements in $V(I_1, I_2)$ are non-negative and hence the histogram intersection kernel can be used for calculate the similarity of two feature vectors V and U, described below as:

$$sim(V, U) = \sum_{t=1}^{D} \min(V_t, U_t) \quad (2)$$

Having obtained the kernel matrix, we can apply the conventional SVM for kinship verification. In our experiments, the RBF kernel was used for similarity measure of each pair of samples.

3. EXPERIMENTAL RESULTS

3.1 Dataset

Currently, there is no publicly available face image database with kinship relations. In this study, we collected 1000 images from the internet through an online search for images of public figures or celebrities and their parents or children. We pose no restrictions in terms of pose, lighting, background, expression, age, ethnicity and partial occlusion on the images used for training and testing. We adopt the Viola-Jones face detector to detect the face region in each image, and correctly detect 800+ face images, each is size of 64×64. We selected 100 pairs of face images for each of the four kinship relations (Father-Son (FS), Father-Daughter (FD), Mother-Son (MS) and Mother-Daughter (MD)) to construct the dataset, such that 800 images are totally used.

For each of the four subset, we construct 100 pairs of positive (true) samples and 100 pairs of negative (false) samples. The positive samples are the true pairs and the false samples are each parent with a randomly selected child from the children images who is not his/her true child. Figure 3 shows some positive and negative samples we used in our experiments.

3.2 Experimental Settings

For each of the four kinship subsets, we perform 5-fold cross validation for the 100 pairs positive and 100 pairs negative samples. We compared the performance of SPLE with the following four face feature descriptors:

- **PCA** [10]: This method has been widely used for a large number of face analysis tasks, and it can also work for the scenario when the face images are collected under uncontrolled conditions. The feature dimension for PCA was empirically set to be 100 in our experiments.

- **LBP** [1]: We followed the parameter setting of LBP in [1] and used 59 bins to describe each face image.

- **HOG** [4]: HOG (histogram of gradient) was originally proposed for human detection and was recently applied to face recognition [2]. Here, we followed the parameter setting of HOG in [3] and used a b-bin HOG descriptor for face representation.

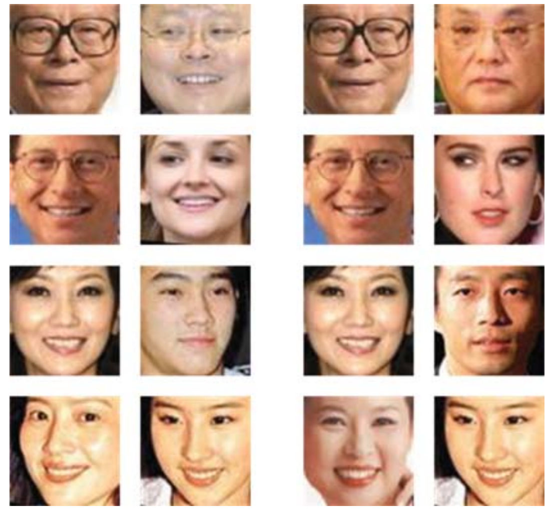

Figure 3: Positive (left) and negative (right) samples used in our experiments.

Table 1: The classification accuracy (%) of different feature descriptors on different subsets.

Method	F-S	F-D	M-S	M-D	Mean
PCA	59.00	51.50	61.50	61.50	58.38
LBP	62.50	59.75	63.25	60.75	61.56
HOG	57.50	50.50	59.50	58.00	56.38
LE	61.75	58.50	68.75	69.50	64.62
SPLE	63.50	61.50	72.50	73.50	67.75

- **LE** [3]: This method is a newly proposed face feature descriptor and shows better performance than other popular face feature descriptors such as LBP and HOG [3]. We used 200 bins for LE to encode a histogram feature for each image.

3.3 Results and Analysis

Table 1 shows the classification accuracy of different feature descriptors on different kinship subsets. As can be seen from this table, the proposed SPLE method outperforms PCA, LBP, HOG and LE with gains in accuracy of 4.50%, 1.00%, 2.00%, and 1.25% on the F-S subset, 10.00%, 1.75%, 11.00%, and 3.00% on the F-D subset, 11.00%, 9.25%, 13.00%, and 3.75% on the M-S subset, and 12.00%, 12.75%, 15.50%, and 4.00% on the M-D subset, respectively. Moreover, we can also observe from this table that all feature descriptors obtain worse performance on the Father-Daughter subset than those on the Mother-Son and Mother-Daughter subsets, which indicates that it is more challenging to recognize the Father-Daughter relation than others from facial images.

Now, we investigate the parameter sensitivity of the proposed SPLE feature descriptor. There are two important parameters: the discrete types M and the spatial pyramid L. Figures 4 & 5 show the classification accuracy of SPLE versus different number of discrete types and different spatial pyramid levels, respectively. We can easily see from the two figures that SPLE is stable to varying parameters. This shows the relative insensitivity of M and L and hence it is easy to select appropriate values of SPLE to obtain good performance.

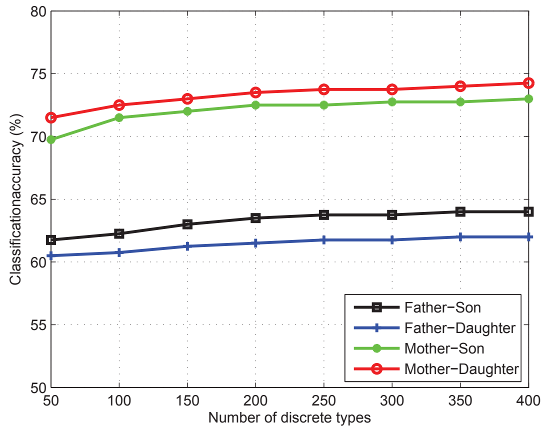

Figure 4: Classification accuracy (%) of SPLE versus different number of discrete types.

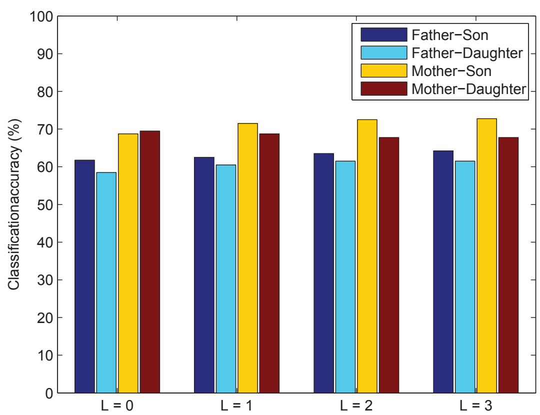

Figure 5: Classification accuracy (%) of SPLE versus different number of spatial pyramid levels.

As an important baseline, the human ability in kinship verification from facial images is also tested. From each of the above four subsets, we randomly selected 100 pairs of face samples, 50 are positive and the other 50 are negative, and presented them to 20 human observers (10 males and 10 females) with age of 20 to 30 years old. None of them received training on the task before the experiment. There are two stages in the experiment. The difference is that, in the first stage (HumanA), only the cropped face regions are shown, while, in the second stage (HumanB), the whole original color images are shown. HumanA intends to test kinship verification purely based on face, while HumanB intends to test kinship verification based on multiple cues including face, hair, skin color, and background. Note that the information provided in HumanA is the same as that provided to the algorithms. Table 2 shows the classification accuracy of human ability on kinship verification on different kinship subsets. We can observe from Tables 1 & 2 that our proposed automatic kinship verification approach can obtain better performance than HumanA, and performs slightly worse than HumanB, which further indicates that some other cues such as hair, skin color, and background also contribute to kinship verification.

Table 2: The classification accuracy (%) of human ability on kinship verification on different kinship subsets.

Method	F-S	F-D	M-S	M-D	Mean
HumanA	63.00	60.00	68.00	72.00	65.75
HumanB	68.00	66.00	76.00	78.00	72.00

4. CONCLUSION

To the best of our knowledge, this paper is the first attempt to investigate kinship verification from facial images under uncontrolled conditions. Our method does not require manual human annotation of face landmarks and the kinship information is automatically obtained from the original pair of images. Moreover, we further proposed a new spatial pyramid learning-based (SPLE) feature descriptor for face representation, and experimental results have shown that the performance of SPLE is not only significantly better than that of the state-of-the-art feature descriptors, but also comparable to that of the human observers.

5. ACKNOWLEDGE

This work was partially supported by the grant from the National Natural Science Foundation of China (No. 11178017) and Beijing Nova Program (No. 2008B57).

6. REFERENCES

[1] T. Ahonen, A. Hadid, et al. Face Description with Local Binary Patterns: Application to Face Recognition. *PAMI*, 28(12):2037–2041, 2006.

[2] A. Albiol, D. Monzo, A. Martin, J. Sastre, and A. Albiol. Face Recognition Using HOG-EBGM. *Pattern Recognition Letters*, 29(10):1537–1543, 2008.

[3] Z. Cao, Q. Yin, X. Tang, and J. Sun. Face Recognition with Learning-based Descriptor. In *CVPR*, pages 2707–2714, 2010.

[4] N. Dalal and B. Triggs. Histograms of Oriented Gradients for Human Detection. In *CVPR*, pages 886–893, 2005.

[5] R. Fang, K. Tang, N. Snavely, and T. Chen. Towards Computational Models of Kinship Verification. In *ICIP*, pages 1577–1580, 2010.

[6] B. Fasel and J. Luettin. Automatic Facial Expression Analysis: A Survey. *Pattern Recognition*, 36(1):259–275, 2003.

[7] G. Guo, Y. Fu, C. Dyer, and T. Huang. Image-Based Human Age Estimation by Manifold Learning and Locally Adjusted Robust Regression. *IEEE Trans. on Image Processing*, 17(7):1178–1188, 2008.

[8] S. Gutta, J. Huang, P. Jonathon, and H. Wechsler. Mixture of Experts for Classification of Gender, Ethnic Origin, and Pose of Human Faces. *IEEE Trans. on Neural Networks*, 11(4):948–960, 2002.

[9] B. Moghaddam and M. Yang. Learning Gender with Support Faces. *PAMI*, 24(5):707–711, 2002.

[10] M. Turk and A. Pentland. Eigenfaces for Recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[11] L. Wolf, T. Hassner, and Y. Taigman. Descriptor Based Methods in the Wild. In *ECCVW*, 2008.

[12] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld. Face Recognition: A Literature Survey. *ACM Computing Surveys*, 35(4):399–458, 2003.