Gabor-Based Gradient Orientation Pyramid for Kinship Verification Under Uncontrolled Environments

Xiuzhuang Zhou
Capital Normal University
Beijing, China
zxz@xeehoo.com

Jiwen Lu
ADSC
Singapore, Singapore
jiwen.lu@adsc.com.sg

Junlin Hu
School of EEE, NTU
Singapore, Singapore
hujunlin@msn.com

Yuanyuan Shang
Capital Normal University
Beijing, China
syy@bao.ac.cn

ABSTRACT

This paper presents a Gabor-based Gradient Orientation Pyramid (GGOP) feature representation method for kinship verification from facial images. First, we perform Gabor wavelet on each face image to obtain a set of Gabor magnitude (GM) feature images from different scales and orientations. Then, we extract the Gradient Orientation Pyramid (GOP) feature of each GM feature image and perform multiple feature fusion for kinship verification. When combined with the discriminative support vector machine (SVM) classifier, GGOP demonstrates the best performance in our experiments, in comparison with several state-of-the-art face feature descriptors. Experimental results are presented to show the efficacy of our proposed approach. Moreover, the performance of our proposed method is also 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, Experimentations

Keywords

Face Analysis, Kinship Verification, Image Descriptors.

1. INTRODUCTION

This paper considers the problem of kinship verification from facial images, which aims to determine whether there is a kinship relation between a pair of facial images or not. This research topic has several potential practical applications such as family album organization, missing children searching, image annotation, and social media analysis. It has become an important research topic in multimedia analysis and computer vision in recent years. However, very limited studies has been systematically conducted along this direction, possibly due to lacking of such publicly available kinship databases and great challenges of this problem.

To our best knowledge, Fang *et al.* [5] proposed the first computational method to tackle the challenge of kinship verification from facial images. Their method first localized sev-

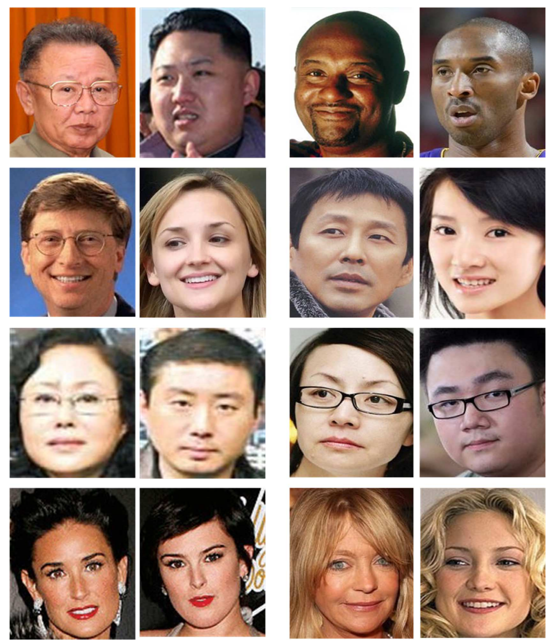

Figure 1: Image samples of our kinship database. From top to bottom are facial images with the Father-Son (FS), Father-Daughter (FD), Mother-Son (MS) and Mother-Daughter (MD) kinship relations, respectively.

eral facial components and extracted some features such as gray intensity value, histogram of gradient, skin color, and facial structure information to describe facial images. Then, the k-nearest-neighbor (KNN) with the Euclidean metric was applied to kinship verification. More recently, Xia *et al.* [10] proposed a transfer subspace learning method for kinship verification. Their basic idea is to adopt an intermediate young parent facial image set to reduce the divergence between the young children and old parent facial images based on the assumption that the children and young parents possess more facial resemblance in facial appearances. While some promising results have been obtained, facial images in their studies were collected under a constrained 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 pose, expression, occlusion, and illumination variations of the acquired face images (Figure 1 shows several example images), and it is of great significant to investigate kinship verification under uncontrolled environments.

In this paper, we propose a new Gabor-based Gradient Orientation Pyramid (GGOP) feature representation method to address this problem. First, we perform Gabor wavelet on each face image to obtain a set of Gabor magnitude (GM) feature images from different scales and orientations. Then, we extract the gradient orientation pyramid (GOP) feature of each GM feature image and perform multiple feature fusion for kinship verification. When combined with the discriminative support vector machine (SVM) classifier, GGOP demonstrates the best performance in our experiments, in comparison with several state-of-the-art face feature representation methods. Experimental results are presented to show the efficacy of our proposed method. Moreover, the performance of our proposed approach is also comparable to that of human observers.

2. PROPOSED APPROACH

2.1 Proposed GGOP

Raw intensity pixel is the most straightforward feature representation of facial images. However, it is not a good descriptor for facial images collected under uncontrolled environments because it suffers from some variations such as varying pose, illuminations, expressions and misalignments. To address this, a number of local feature descriptors have been proposed in recent years, such as local binary pattern (LBP) [1], learning-based (LE) [3] and spatial pyramid LE (SPLE) [11]. Since face images collected are under uncontrolled environments in our scenario, these feature representation methods are not robust enough to these variations. Motivated by recent advances in face analysis that gradient-based feature representation methods are more robust to practical variations [6], we propose a new GGOP feature descriptor for robust face representation, works as follows:

Step 1: For each face image, we perform Gabor wavelet with different scales and directions. For each Gabor filtered component, we compute the Gabor magnitude (GM) and normalize it into a gray-scale feature image. Specifically, the GM feature image is defined as follows

$$m_{\mu,\nu}(x,y) = \sqrt{Im^2(O_{\mu,\nu}(x,y)) + Re^2(O_{\mu,\nu}(x,y))} \quad (1)$$

where $Re(O_{\mu,\nu}(x,y))$ and $Im(O_{\mu,\nu}(x,y))$ denote the real and imaginary parts of $O_{\mu,\nu}(x,y)$, where $O_{\mu,\nu}(x,y)$ is the Gabor filtered component from the μth orientation and the νth scale, $m_{\mu,\nu}(x,y)$ and $p_{\mu,\nu}(x,y)$ are the GM and GP features of the pixel located at (x,y) of orientation μ and scale ν, respectively. Similar to that of [7], we also choose Gabor functions with eight orientations $\mu = \{0, 1, \cdots, 7\}$ and five scales $\nu = \{0, 1, \cdots, 4\}$, and obtain a total of 40 Gabor kernel functions. The value of each parameter follows the setting in [7]: $\sigma = 2\pi$, $k_{max} = \pi/2$, $f = \sqrt{2}$. Hence, we extract 40 GM features from each face image.

Step 2: For each GM feature image $I(\mathrm{p})$, where $\mathrm{p} = (x,y)$ indicates pixel locations, we extract the gradient orientation pyramid (GOP) feature for representation. We define the pyramid of I as $\mathcal{P}(I) = \{I(\mathrm{p};\sigma)\}_{\sigma=0}^{s}$ with

$$\begin{aligned} &I(\mathrm{p};0) = I(\mathrm{p}) \\ &I(\mathrm{p};\sigma) = [I(\mathrm{p};\sigma-1) * \Phi(\mathrm{p})] \downarrow_2, \ \sigma = 1, \ldots, s \end{aligned} \quad (2)$$

where $*$ denotes the convolution operator, $\Phi(\mathrm{p})$ is the Gaussian kernel, which is set 0.5 as the standard deviation in our experiments, $\downarrow_2$ denotes half size downsampling, and s is the number of pyramid layers. Note that in Eq. (2) the notation I is used for all the GM feature images at different scales for convenience. The reason why GP is not utilized is GP is not stable, especially when face images collected under uncontrolled conditions.

Step 3: The gradient orientation at each scale σ is defined by its normalized gradient vectors at each pixel.

$$g(I(\mathrm{p};\sigma)) = \begin{cases} \frac{\nabla(I(\mathrm{p};\sigma))}{\|\nabla(I(\mathrm{p};\sigma))\|}, & \text{if } \|\nabla(I(\mathrm{p};\sigma))\| > \tau \\ (0,0)^\top, & \text{otherwise} \end{cases} \quad (3)$$

where τ is a threshold for dealing with flat pixels. The GOP of I is naturally defined as $\mathcal{G}(I) = \mathrm{stack}(\{g(I(\mathrm{p};\sigma))\}_{\sigma=0}^{s}) \in \mathbb{R}^{d\times 2}$ that maps I to a $d \times 2$ representation, where stack($\bullet$) is used for stacking GOs of all pixels across all scales and d is the total number of pixels.

2.2 Classifier

Given an image pair (I_1, I_2) and the corresponding GOPs $(G_1 = \mathcal{G}(I_1),\ G_2 = \mathcal{G}(I_2))$, the feature vector $x = \mathcal{F}(I_1, I_2)$ is computed as the cosines of the difference between GOs at all pixels

$$x = \mathcal{F}(I_1, I_2) = (G_1 \odot G_2) \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad (4)$$

where $\odot$ is the element-wise product. Next, we apply the Gaussian kernel to the extracted feature x to be used with the SVM framework. Specifically, our kernel is defined as

$$K(\mathrm{x}_1, \mathrm{x}_2) = \exp(-\gamma \|x_1 - x_2\|^2) \quad (5)$$

where γ is a parameter determining the size of RBF kernels.

Since there are 40 GM features in our GGOP feature representation method, we need to combine these multiple image descriptors effectively in the SVM classifier. We introduce a new kernel learning approach in the following [9].

Given L sub-kernel $K_1, K_2, \ldots, K_L$ computed on different image features, the kernel K can be formulated as:

$$K(x_i, x_j) = \sum_{\ell=1}^{L} w_\ell K_\ell(x_i, x_j) \quad (6)$$

where L is the number of Gabor images (in our work, $L = 40$), $w_\ell \geq 0$ and $\sum_{\ell=1}^{L} w_\ell = 1$. The weight w_ℓ measures the importance for discrimination. Our task is to learn the weight w_ℓ before training SVM. If we use all the sub-kernel K_ℓ ($1 \leq \ell \leq L$) directly, we may encounter the scale and transformation problems. In order to deal with these problems, we normalize each sub-kernel K_ℓ to constrain the sub-kernel K_ℓ margin as 1:

$$\begin{aligned} \overline{K_\ell} &= K_\ell/(a_\ell - b_\ell) \\ \overline{a_\ell} &= a_\ell/(a_\ell - b_\ell) \\ \overline{b_\ell} &= b_\ell/(a_\ell - b_\ell) \end{aligned} \tag{7}$$

where a_ℓ is the largest value of the sub-kernel K_ℓ, b_ℓ is the minimum value of K_ℓ,respectively. And $\overline{K_\ell}$ is the normalized variant of K_ℓ.

Kernel value can be treated as measuring the similarity between two features. In the ideal case, all the images from the same categories should be similar, and those from different categories are dissimilar.

Our aim is to learn the weights of normalized sub-kernels

$$K_{opt}(x_i, x_j) = \sum_{\ell=1}^{L} w_\ell \overline{K_\ell}(x_i, x_j) \tag{8}$$

The alignment loss of $\overline{K_\ell}(x_i, x_j)$ is defined as:

$$\xi_\ell(x_i, x_j) = \begin{cases} a_\ell - \overline{K_\ell}(x_i, x_j), & \text{if } y_i = y_j \\ \overline{K_\ell}(x_i, x_j) - b_\ell, & \text{otherwise} \end{cases} \tag{9}$$

The alignment loss defines how well the kernel is closed to its ideal form. Therefore, The total alignment loss of image x_i and x_j is defined as:

$$\xi(x_i, x_j) = \sum_{\ell=1}^{L} w_\ell \ \xi_\ell(x_i, x_j) \tag{10}$$

If the alignment loss is small, it implicates that the kernel K is close to its ideal form. In order to make kernel K closer to its ideal form, we minimize the square of the alignment loss:

$$\sum_{i,j=1}^{N} t_{ij} \ \xi^2(x_i, x_j) + \lambda \sum_{\ell=1}^{L} w_\ell \tag{11}$$

where t_{ij} is the weight of each image pairs, and λ is a tradeoff parameter to prevent overfitting problem. t_{ij} is designed to tackle the imbalanced dataset problem. In this paper, $t_{ij} = 1/(n_i \times n_j)$, where n_i is the number of images whose labels are to y_i.

Denote $w = [w_1, w_2, \ldots, w_L]^\top$ as the weights vector, and $\xi_{ij} = [\xi_1(x_i, x_j), \xi_2(x_i, x_j), \cdots, \xi_L(x_i, x_j)]^\top$ as the column vector of alignment loss of all sub-kernels, the optimization problem can be rewritten as:

$$\begin{aligned} \min \quad & w^\top (\sum_{i,j=1}^{N} t_{ij} \xi_{ij} \xi_{ij}^\top + \lambda I) w \\ s.t. \quad & w_\ell \geq 0 \ (\ell = 1, 2, \ldots, L) \\ & \sum_{\ell=1}^{L} w_\ell = 1 \end{aligned} \tag{12}$$

It is a quadratic programming problem with L unknowns, which can be solved efficiently. Here, I is identical matrix, and the parameter λ is set to $0.1 \sum_{i,j=1}^{N} t_{ij} \xi_{ij} \xi_{ij}^\top$ for all the experiments.

3. EXPERIMENTAL RESULTS

3.1 Dataset

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.

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 our GGOP with the following five face feature descriptors:

- **PCA** [8]: 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 P-CA 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 9-bin HOG descriptor for face representation.
- **LE** [3]: We follow the parameter settings in [3] and use 200 bins for LE to encode a histogram feature for each image.
- **SPLE** [11]: SPLE is the state-of-the-art face feature descriptor and shows better performance than other popular face feature descriptors [11]. We used 4200 bins for LE to encode a histogram feature for each image.

3.3 Results and Analysis

Comparisons with other feature descriptors: Table 1 shows the classification accuracy of different feature descriptors on different kinship subsets. As can be seen from this table, the proposed GGOP method outperforms the other compared five feature descriptors in all subsets of our experiments. 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 verify the Father-Daughter relation than others from facial images.

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
GGOP	65.50	65.50	73.50	74.50	69.75

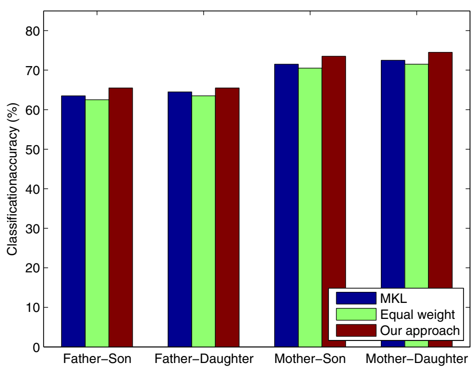

Figure 2: Classification accuracy (%) of different weighting methods.

Comparisons with different feature weighting methods: We also compare our feature weighting method with the widely used MKL method and the equal weighting of each GM feature. For the equal weighting method, each G-M feature is assigned the same weighting value: 1/40. Figure 2 shows the classification accuracy of different methods on different subsets. We can easily see from that our approach performs better than the other compared methods in all subsets of our kinship verification experiments.

Comparisons with humans: 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 Table 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
GGOP	65.50	65.50	73.50	74.50	69.75

4. CONCLUSION

In this paper, we proposed a new Gabor-based gradient orientation pyramid (GGOP) feature representation method for kinship verification from facial images under uncontrolled conditions. To make better of multiple feature information, we propose a new feature weighting method to exploit complementary information for verification. Experimental results have shown the efficacy of our proposed feature descriptor for kinship verification. Moreover, the performance of our proposed method is comparable to that of human observers. How to extract other contextual information from facial images to further improve the kinship verification performance appears an interesting future work.

Acknowledgement

This work is partially supported by the research grant from the National Natural Science Foundation of China (No. 11178017 and No. 60903141) and Beijing Nova Program (No. 2008B57), and the research grant of the Human Sixth Sense Program at the Advanced Digital Sciences Center from the Agency for Science, Technology and Research (A*STAR) of Singapore. Xiuzhuang Zhou and Yuanyuan Shang are also supported by funding of Beijing Engineering Research Center of High Reliable Embedded System Lab, Capital Normal University, China.

5. 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] H. Ling, S. Soatto, N. Ramanathan, and D. Jacobs. Face verification across age progression using discriminative methods. *IEEE TIFS*, 5(1):82–91, 2010.

[7] J. Lu and Y.-P. Tan. Gait-Based Human Age Estimation. *IEEE TIFS*, 5(4):961–970, 2010.

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

[9] S. Wang, S. Jiang, Q. Huang, and Q. Tian. Multiple kernel learning with high order kernels. In *ICPR*, pages 2138–2141, 2010.

[10] S. Xia, M. Shao, and Y. Fu. Kinship verification through transfer learning. In *IJCAI*, pages 2539–2544, 2011.

[11] X. Zhou, J. Hu, J. Lu, Y. Shang, and Y. Guan. Kinship verification from facial images under uncontrolled conditions. In *ACM MM*, pages 953–956, 2011.