First published in:

2010 International Conference on Pattern Recognition

# Multi-resolution Local Appearance-based Face Verification

Hua Gao, Hazım Kemal Ekenel, Mika Fischer, Rainer Stiefelhagen
*Institute for Anthropomatics*
*Karlsruhe Institute of Technology*
*Karlsruhe, Germany*
*Email: {hua.gao, ekenel, mika.fischer, rainer.stiefelhagen}@kit.edu*

***Abstract*—Facial analysis based on local regions / blocks usually outperforms holistic approaches because it is less sensitive to local deformations and occlusions. Moreover, modeling local features enables us to avoid the problem of high dimensionality of feature space. In this paper, we model the local face blocks with Gabor features and project them into a discriminant identity space. The similarity score of a face pair is determined by fusion of the local classifiers. To acquire complementary information in different scales of face images, we integrate the local decisions from various image resolutions. The proposed multi-resolution block based face verification system is evaluated on the experiment 4 of Face Recognition Grand Challenge (FRGC) version 2.0. We obtained 92.5% verification rate @ 0.1% FAR, which is the highest performance reported on this experiment so far in the literature.**

***Keywords*-local appearance representation; face verification; multi-resolution; gabor filter;**

## I. INTRODUCTION

Face recognition / verification has been an attractive research domain for decades, and it becomes a more challenging problem due to uncontrolled condition in modern security applications. The local appearance-based face representation is one of the most promising approaches that can successfully deal with appearance variations caused by illumination or expression changes [1].

An initial study of the local representation for face recognition can be found in [2], where the local facial regions such as eyes, nose, and mouth are modeled with separate linear subspaces. However, this local component based approach requires a precise localization of salient facial features which is not an easy task. A more generic local appearance based approach was proposed later in [3], [4]. This approach divides an input image into non-overlapping local blocks and performs subspace or frequency analysis separately on the local blocks. Experiments have proved that this approach is superior to the holistic approaches as well as the local component based approaches.

However, the non-overlapping block partition does not include the spatial correlation information between the neighboring blocks. There are some studies which model local appearance with overlapping blocks [5], [6]. In this paper, we integrate part of the neighboring frequency information by utilizing the nature of the Gabor wavelet filters, in which the local filter response is smoothed with neighboring pixels. On the other hand, the Gabor wavelet transformation usually results in very high dimensional feature vectors, which may result in singular within-class scatter matrices if we want to perform dimensionality reduction with linear discriminant analysis (LDA) [7]. Some solutions such as the $nullspace$ methods have been proposed to potentially solve this dimensionality problem [8]. Dividing the Gabor responses into local blocks has also been proven to be an effective solution [7]. It implicitly incorporates the spatial correlation information from the neighboring blocks and solves the dimensionality problem decently.

As a major contribution, we combine the local Gabor feature modeling with a multi-resolution face representation, which enable us to acquire complementary information from image spatial space as well as scale space. The proposed face modeling is evaluated on the FRGC experiment 2.0.4, which is considered to be the most challenging experiment among the FRGC experiments. With the multi-resolution local appearance-based face representation, we achieved $92.5\%$ verification rate (VR) @ $0.1\%$ false acceptance rate (FAR), which is the best result on this experiment reported in the literature.

## II. METHODOLOGY

Our face image processing pipeline is illustrated in Fig. 1. A given face image is first normalized with the available eye locations in a specific resolution defined by the inter-ocular distance. The aligned face image is then preprocessed to remove the effects of illumination variations, especially local shadowing and highlights. We follow the sequence of steps in [9] with Gamma correction, Difference of Gaussian (DoG) filtering, and highlights suppression. Then we transform the preprocessed face image with a 2D Gabor wavelet filter bank and obtain a set of filter responses with different scale and orientation parameters. We divide the 2D Gabor responses into local non-overlapping blocks and build local experts with the extracted local Gabor features. By merging the local experts in different alignment resolutions we obtain the final classification score for face verification.

### A. Local Gabor Feature Representation

2D Gabor wavelets are considered to be one of the most successful local descriptors for face representation due to

1051-4651/10 $26.00 © 2010 IEEE
DOI 10.1109/ICPR.2010.371

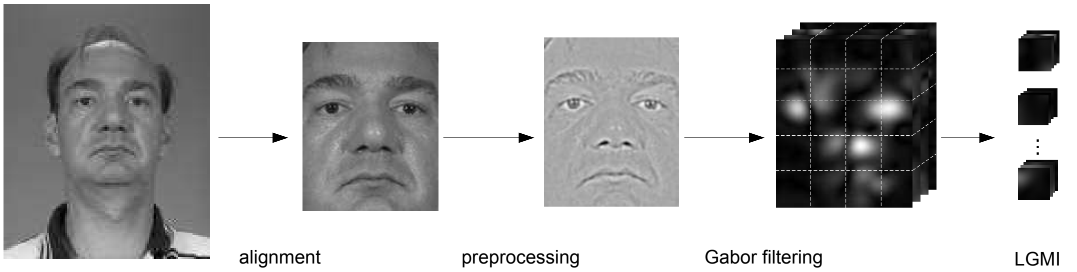


Figure 1. The procedure of LGMI extraction

their biological relevance. The Gabor wavelets are the mathematical model of visual cortical cells of mammalian brains, which decompose an input image into multiple scales and multiple orientations. The extracted feature representation has been widely used in the computer vision domain due to its optimal localization properties in both spatial and frequency domain.

A 2D Gabor wavelet can be considered as an excellent bandpass filter which consists of a planar sinusoid multiplied by a two dimensional Gaussian. Such a complex filter can be defined as follows:

$$\Psi_{u,v} = \frac{\|k_{u,v}\|^2}{\sigma^2} e^{(-\|k_{u,v}\|^2\|z\|^2/2\sigma^2)} \left[ e^{i\vec{k}_{u,v}z} - e^{-\sigma^2/2} \right] \tag{1}$$

where $k_{u,v} = k_v e^{i\Phi_u}$, $k_v = \frac{k_{max}}{f^v}$ is the parameter for the frequency, $\Phi_u = \frac{u\pi}{8}$, $\Phi_u \in [0, \pi)$ is the parameter for the orientation. The sinusoid wave defined in Equation 1 is activated by frequency information in the image. The Gaussian envelope ensures that the convolution is dominated by the image region that is close to the wavelet center. This means, when a signal is convolved with the Gabor wavelet, the frequency information near the frequency of the sinusoid wave is captured and the other frequency information will be neglected.

When varying the parameter $u$ and $v$, for example, in the range $[0, 7]$ and $[0, 4]$ respectively, we get $40$ Gabor wavelets in different forms with $8$ orientations and $5$ scales. A filter with a certain orientation and scale parameter captures corresponding edge information to a specific degree of detail. Instead of convolving the filter kernels in image space, we filter the input image in Fourier space by transforming the image and convolution kernels with the FFT, and transforming the filtered Fourier feature back to into image space with the inverse FFT. After filtering we obtain 40 Gabor magnitude images (GMI) in different scales and orientations. In order to reserve location information, the Gabor features are spatially partitioned into $N$ local blocks, each of which corresponds to a local patch of the face image. Fig. 1 illustrates the feature partition on the GMIs.

We perform separate PCLDA (principal component and linear discriminant analysis) on each of the extracted local GMIs (LGMI) to reduce dimensionality and transform it to a discriminative feature space. The local expert $C_i$ for a certain local block is then based on the nearest neighbor classification in the trained PCLDA space. The final decision is obtained by fusing the $N$ local experts $C_i$ in a simple weighted form:

$$C = \sum_{i=1}^{N} w_i \cdot C_i \tag{2}$$

where $w_i$ is the weight of $C_i$.

### B. Multi-resolution Face Models

Face registration is an important step for accurate face recognition. In addition to the precision of the registration, the proportion of the face region is also a crucial parameter which can be adjusted to achieve better performance. As we use the location of the eye centers to register faces, the aligned image (with fixed size) represents different details of the face depending on the inter-ocular distance parameter used for alignment. As shown in Fig. 2, if we align a face image into a low resolution (with low inter-ocular distance) more information is included in the image such as face contour, hair, etc. Aligning the face image into a high resolution results in a more close view of the face and only the inner part of the face is present in the aligned image. Although the face alignment with a smaller inter-ocular distance provides more discriminant information about the face (e.g face contour), it also includes some background clutter which adds noise to the face representation. The medium resolution in the middle of Fig. 2 is an empirical trade-off between information content and noise. Since different resolutions focus on its own view of analysis, combining the local experts on different resolution may potentially increase the robustness for classification. This idea has been successfully exploited in [10], [11].

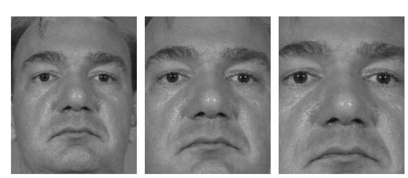

Figure 2. Face alignment with different interocular distances resulting in different resolutions : low, medium, and high

## III. Experiments

We evaluated our method on the experiment 4 in the Face Recognition Grand Challenge (FRGC) v2.0 [12]. The training set for this experiment consists of $12,776$ images from 222 individuals. The gallery and probe set only consist of a single still image per subject. This experiment contains $8,014$ uncontrolled query images and $16,028$ controlled target images from 466 subjects. It is the most challenging experiment due to uncontrolled conditions including large illumination changes, out of focus, and partial occlusions. The performance is reported as verification rate (VR) at $0.1\%$ FAR. There are three Receiving Operator Characteristic (ROC) curves can be generated, which correspond to three different time gaps. ROC I corresponds to the images collected within a semester, ROC II within a year, and ROC III between semesters.

### A. Experimental Results

The FRGC data set provides labels of salient facial features such as eye centers and mouth corners. We use the provided eye labels to align the face images. After alignment, the size of the face image is $128 \times 160$ pixels, with the eye distance being 72 pixels, which corresponds to the medium resolution. After 2D Gabor filtering, the resulting GMIs are spatially partitioned into 20 non-overlapping patches of $32 \times 32$ pixel size. Since we use 40 Gabor wavelets for this experiment, the dimension of each LGMI is $32 \times 32 \times 40 = 40960$, which is very high compared to the number of subjects in the training set (222). So the LGMIs are uniformly down-sampled by averaging the magnitude values in an $8 \times 8$ grid before the subspace analysis. After down-sampling, the dimension of each LGMI is reduced to $640$ $(= 4 \times 4 \times 40)$. Each dimension of one LGMI is normalized with zero mean unit variance. Each individual normalized LGMI is then projected into a discriminant identity space with PCLDA. Normalized correlation is adopted as the distance metric to calculate the similarity of local blocks. Combining all the local experts, we get the final similarity score for one single resolution. The weights $w_i$ for each local classifier $C_i$ were assigned equally.

The ROC performance on experiment 4 for the medium resolution is plotted in Fig. 3. It can be noticed that the applied preprocessing contributes a performance improvement of about $5\%$ for all the three ROCs.

As ROC III evaluates the matching with large time gap (between semesters), we compared the ROC III performance in the later experiments to face the challenge. In Fig. 4, we compared the ROC performance between different alignment resolutions and their fusion. As expected, the VR of medium resolution ($RES\_M$) outperforms the low ($RES\_L$) and high ($RES\_H$) alignment resolution. As a compromise of information content and noise the $RES\_M$ achieved $88.1\%$ VR at FAR of $0.1\%$, while $RES\_L$ and $RES\_H$ achieved

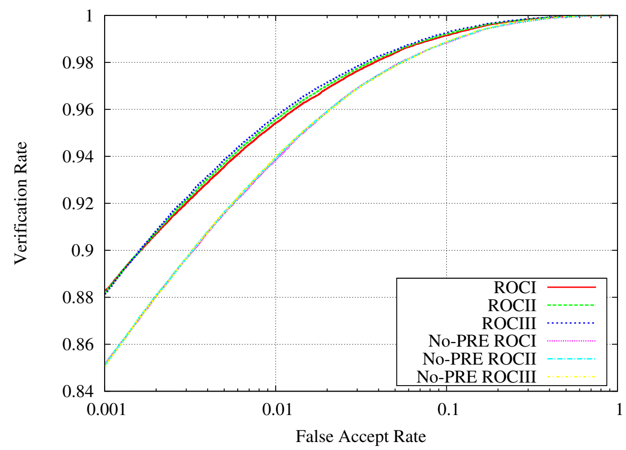


Figure 3. Effect of the preprocessing.

$86.8\%$ and $86.4\%$ respectively. However, as we believe that different resolution may contribute complementary views of the whole face appearance, the decision from each resolution is fused to a multi-resolution similarity score. As can be observed in Fig. 4, a noticable performance gain is achieved with this multi-resolution method (Multi-RES). Using sum-rule fusion, we obtained $92.0\%$ VR @ $0.1\%$ FAR in ROC III.

We carried out additional experiments to see whether an additional face representation contributes complementary information for verifying faces. We used another local face representation based on discrete cosine transformation (DCT) [3] as extra evidence. The multi-resolution study was also applied on this representation.We combined the similarity matrix obtained from this representation with the one with local gabor features on score-level, and the verification rate was improved further to $92.5\%$.

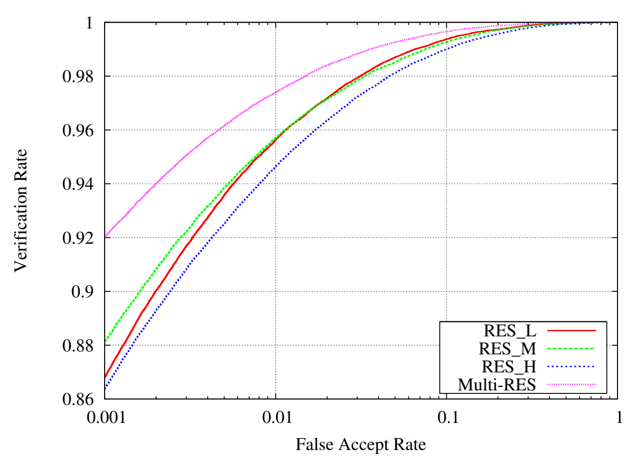


Figure 4. Multi-resolution performance in Exp. 4 of the FRGC data set (ROC III).

Finally, we compare our results with the FRGC baseline and other best known results [7], [9], [11], [13] in Table I.

Hwang et al. modeled separate frequency bands in a holistic way. Different face models in different resolutions were merged later [11]. In [9], Local Binary Patterns (LBP) and global Gabor features were fused in feature level and kernel-based subspace analysis was applied to extract discriminant nonlinear features. Our method is partially inspired by the work in [7], however, they differ in several aspects: *a*) Instead of using equal kernel mask size for different scale parameters, we set the mask size according to the $\sigma$ of Gaussian envelope. This avoids the distortion of filter response if the kernel mask size is too small for the largest scale. *b*) We normalize image contrast with some preprocessing steps, which boost our performance to $88.1\%$ with single resolution. *c*) The multi-resolution decision making improved the performance further. The result in [13] is close to our best result. However, the identity information in the target set was utilized, which did not follow the protocol of the experiment 4. Overall, our proposed method improves the baseline by $80.5\%$ in VR @ $0.1\%$ FAR.

Table I
PERFORMANCE COMPARISON ON EXP. 4 OF THE FRGC DATA SET (ROC III).

| Method | VR @ $0.1\%$ FAR |
|---|---|
| FRGC Baseline | 12.0% |
| Method in [11] | 74.3% |
| Method in [9] | 83.6% |
| Method in [7] | 86.0% |
| Method in [13] | 91.3% |
| Our Method | **92.5%** |

## IV. CONCLUSIONS

This paper presents a multi-resolution local appearance-based face verification system. The local facial appearance is modeled with Gabor features and they are separately projected into a discriminant identity space. The similarity score of a face pair is determined by merging the local experts. To acquire complementary information in different scales of face images, we integrate the local decisions from various image resolutions. The proposed system was evaluated on experiment 4 of the FRGC data set. We achieved $92.5\%$ VR @ $0.1\%$ FAR, which is the best result reported in the literature.

## ACKNOWLEDGMENT

This work was realized as part of the Quaero Programme, funded by OSEO, French State agency for innovation.

## REFERENCES

[1] H. K. Ekenel, "A Robust Face Recognition Algorithm for Real-World Applications," PhD. Dissertation, University of Karlsruhe (TH), 2009.

[2] A. Pentland, B. Moghaddam, and T. Starner, "View-based and Modular Eigenspaces for Face Recognition," in *Proc. of CVPR*, 1994.

[3] H. K. Ekenel and R. Stiefelhagen, "Local Appearance-based Face Recognition using Discrete Cosine Transform," in *Proc. of EUSIPCO*, 2005.

[4] R. Gottumukkal and V. K. Asari, "An Improved Face Recognition Technique based on Modular PCA Approach," *Pattern Recognition Letters*, vol. 25, no. 4, 2004.

[5] S. Eickeler, S. Mueller, and G. Rigoll, "Recognition of JPEG Compressed Face Images based on Statistical Methods," *Image Vision Computing*, vol. 18, no. 4, pp. 279–287, 2000.

[6] C. Sanderson and K. K. Paliwal, "Fast Features for Face Authentication under Illumination Direction Changes," *Pattern Recognition Letters*, vol. 24, pp. 2409–2419, 2003.

[7] Y. Su, S. Shan, X. Chen, and W. Gao, "Hierarchical Ensemble of Global and Local Classifiers for Face Recognition," in *Proc. of ICCV*, 2007, pp. 1–8.

[8] L.-F. Chen, H.-Y. Liao, M.-T. Ko, J.-C. Lin, and G.-J. Yu, "A new LDA-based face recognition system which can solve the small sample size problem," *Pattern Recognition*, vol. 33, no. 10, pp. 1713–1726, 2000.

[9] X. Tan and B. Triggs, "Fusing Gabor and LBP Feature Sets for Kernel-Based Face Recognition," in *Analysis and Modelling of Faces and Gestures*, vol. 4778, 2007, pp. 235–249.

[10] http://www.pittpatt.com/face_recognition/.

[11] W. Hwang, G. Park, J. Lee, and S.-C. Kee, "Multiple Face Model of Hybrid Fourier Feature for Large Face Image Set," in *Proc. of CVPR*, 2006, pp. 1574–1581.

[12] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in *Proc. of CVPR*, 2005, pp. 947–954.

[13] M. Savvides, J. Heo, R. Abiantun, C. Xie, and B. V. Kumar, "Class Dependent Kernel Discrete Cosine Transform Features for Enhanced Holistic Face Recognition in FRGC-II," in *Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, 2006.