

Appearance-based Face Recognition using Aggregated 2D Gabor Features

King Hong Cheung¹, Jane You^{1,*} Qin Li¹ and Prabir Bhattacharya²

¹Department of Computing
The Hong Kong Polytechnic University, Kowloon, Hong Kong

²Concordia Institute for Information Systems Engineering,
Concordia University, Montreal, Quebec, H3G 1T7, Canada

Abstract. Current holistic appearance based face recognition methods require a high dimensional feature space to attain fruitful performance. In this paper, we have proposed a relatively low feature dimensional, template-matching scheme to cope with the transformed appearance-based face recognition problem. We use aggregated Gabor filter responses to represent face images. We investigated the effect of “duplicate” images (images from different sessions) and the effect of facial expressions. Our results indicate that the proposed method is more robust in recognizing “duplicate” images with variations in facial expression than the Principal Component Analysis method.

1 Introduction

Biometric recognition, or simply biometrics, is concerned with the automatic recognition of individuals based on their biological/physiological or behavioral characteristics. Although we do not expect that a single biometric to satisfy all identification requirements, the use of such unique, reliable and stable personal features has attracted considerable interest in the development of biometrics-based identification systems for civilian, military, and forensic applications that has attained certain level of maturity [8]. Among the many body characteristics that have been used, face is one of the most commonly used characteristics and has been studied across a number of research fields, e.g. computer vision, pattern recognition [6],[9],[17],[18]. Face recognition is a non-intrusive method that captures still images and/or video sequences - from controlled, static environment to uncontrolled, cluttered environment; we can perform recognition using 2D images and/or 3D models with approaches that are holistic/global, or feature-based, or structural, or hybrid [19],[8]. Holistic/global approach generally refers to those methods that take the whole face region as the input to a face recognition system [19]. A well-known holistic/global approach to face recognition is applying the principal component analysis (PCA) on preprocessed, or even transformed, face training images. A subset of principal components forms the global representation of face for recognition with a suitable (dis)similarity measure.[15],[11],[5] Another common holistic/global approach is the introduction of artificial neural networks as the classifier for face recognition [6].

In the feature-based structural approach, we first need to locate the feature points and the regions of interest, such as eyes, noses and mouth; then one extracts the features with the use of various operators or transforms based on their geometric or appearance characteristics [19]. In [16], this approach is used to locate a set of fiducial points as nodes of the elastic bunch graph. A small set of Gabor jets (i.e., filter responses) from an individual is stored at the relevant nodes of the elastic bunch graph for face recognition. Hybrid approach takes the global (whole face) and local features consideration that is arguably to be potentially the best approach. This approach is also used in [6] to allocate each local feature region a support vector machine (SVM) detector to decompose the face into a set of facial components.

There are three key, interdependent subtasks need to be addressed in automatic face recognition: (1) detection and rough normalization of faces, (2) feature extraction and accurate normalization, and (3) identification and/or verification. That is to say, given a still image or video sequence, the system can first localize face(s) within the image or video sequence; features are extracted from potential face(s) for identification/verification [19],[8].

In many commercial and civilian systems, since the environment is static and under control, full elasticity of the automatic face recognition system may not be required, i.e. with cooperative subjects, proper 2D frontal image can be obtained. In such a case, the automatic face recognition problem can be simplified to the classical pattern recognition/image retrieval problem, which deals mainly with feature extraction and identification/verification. Texture analysis is one of the main streams in image processing/analysis and retrieval, especially for monochromatic images. To perform texture analysis, filtering is one of the popular techniques. 2D Gabor filters proposed by Daugman [3] are one of the (multi-channel) filtering techniques adopted for texture analysis because of spatial-frequency localization, tunable in orientations, radial frequency bandwidths and center frequencies [2]. Characteristic textures contained in an image can be used to distinguish it. Therefore, we employ texture information for the retrieval of images containing identical and/or similar attributes.

2D Gabor filters (also known as 2D Gabor wavelets) are the key element in both holistic/global approach and feature-based/structural approach to face recognition [16],[17]. They were developed originally to model the receptive field profiles of mammalian cortical simple cells. Because of biological relevance and computational properties, they are used for image analysis, especially texture analysis [3],[4],[2],[9].

The holistic/global approach was adopted in Liu et al. [10] where an independent Gabor features (IGFs) method for face recognition was developed by performing PCA and then Independent Component Analysis (ICA) on downsampled Gabor filter responses. In the testing of face databases, they used either manually detected face images or no detection required and they concentrated on the feature extraction phase and identification/verification phase. In the feature extraction phase, they first applied Gabor filters (5 scales and 8 orientations) on the testing images in the frequency domain and then the filter responses are down-sampled by a factor of 64. The dimensionality is further reduced by performing PCA on down-sampled filter responses. IGFs are finally deduced by means of ICA. In the identification/verification phase, the MAP Bayes' rule is applied using the Probabilistic Reasoning Model (PRM) to perform classification.

Wu et al. [17] used the feature-based/structural approach to extract Gabor features in regions of interest such that computation complexity of Gabor features is reduced.

They detected the face first and then locate the feature points of interest within the face. Afterwards, they applied Gabor filters around the eight feature points of interest to generate filter responses as features for corresponding feature points of interest. Features points of interest in the query image are matched against the analogous regions in images registered in the database. The similarity measure between the query image and registered images is defined as the average of, regular correlation of all features of feature points of interest.

The feature dimension in [10] and [17] are both comparatively high. We, therefore, suggested a holistic template-matching approach of relatively low feature dimension that exploits aggregated Gabor filter responses and distance measures to deal with the transformed appearance-based face recognition problem.

The organization of the paper is as follows. Section 2 will briefly review the theory of 2D Gabor filters. Section 3 will outline the rationale of our suggested approach. Section 4 will discuss the results of our experiments. Section 5 will give our conclusion.

2 Background

2D Gabor filters proposed by Daugman [3],[4] extended from the original version (generally referred to as the 1D version) proposed by D. Gabor in communication engineering, has shown its capability in modeling the receptive-field profiles of simple cells in mammalian visual cortex. Simple cells in mammalian visual cortex are organized roughly in the polar Orientation-Frequency plane to form a receptive-field such that with their specific 2D locations in visual space, preferred orientations and spatial frequencies, localized 2D spectral information is captured [3]. A set of 2D Gabor filters, which are of various spatial dimensions and, spatial frequency and orientation bandwidths to imitate those empirical characteristics of receptive-field, can be used to extract different kinds of information from images.

2D Gabor filters have been shown to be effective for texture analysis on monochromatic images. They optimally achieve joint resolution in space and spatial frequency; their orientations, radial frequency bandwidths and center frequencies are all tunable. Thereby, it is adopted to perform texture analysis. [2],[4],[10]

2D Gabor filters, in general, are of the following functional form:

$$h(x, y) = g(x', y') \exp[2\pi i(Ux + Vy)] \quad (1)$$

where $(x', y') = (x \cos \phi + y \sin \phi, -x \sin \phi + y \cos \phi)$ and

$$g(x, y) = \left(\frac{1}{2\pi\lambda\sigma^2} \right) \exp \left[-\frac{(x/\lambda)^2 + y^2}{2\sigma^2} \right] \quad (2)$$

with $\sigma = \sigma_x = \sigma_y$ is the standard deviation of the Gaussian envelope.

The corresponding Fourier transform is

$$H(u, v) = \exp\{-2\pi^2\sigma^2 [(u' - U)^2\lambda^2 + (v' - V)^2]\} \quad (3)$$

where $(u', v') = (u \cos \phi + v \sin \phi, -u \sin \phi + v \cos \phi)$ and

$$(U', V') = (U \cos \phi + V \sin \phi, -U \sin \phi + V \cos \phi)$$

$$k(x, y) = k_c(x, y) + ik_s(x, y) \quad (4)$$

where

$$k_c(x, y) = \text{Re}\{k(x, y)\} = h_c(x, y) * t(x, y), \quad (5)$$

$$k_s(x, y) = \text{Im}\{k(x, y)\} = h_s(x, y) * t(x, y) \quad (6)$$

The amplitude and phase (envelopes) of $k(x, y)$ are, respectively,

$$m(x, y) = [k_c^2(x, y) + k_s^2(x, y)]^{1/2} \quad (7)$$

$$\psi(x, y) = \tan^{-1}[k_s(x, y)/k_c(x, y)] \quad (8)$$

The expanded form of the family of Gabor filters used in our case study is

$$h(x, y, \omega, \theta) = \frac{1}{2\pi\sigma^2} \exp\left(\frac{-\omega}{8\kappa^2}(4x^2 + y^2)\right) \left[\exp(i\omega x') - \exp\left(\frac{\kappa^2}{2}\right) \right] \quad (9)$$

for $\kappa = \sqrt{2 \ln 2} \left(\frac{2^s + 1}{2^s - 1} \right)$ and $\omega = \frac{\kappa}{\sigma}$

where $i = \sqrt{-1}$, ω is the sinusoidal wave frequency in spatial domain, θ is the orientation of the filter and σ is the standard deviation of the Gaussian envelope [9].

3 Proposed Method

We use aggregated 2D Gabor features, as a relatively low dimensional (space) representation of face because current appearance based face recognition techniques require a high dimensional feature space. Liu et al. [10] reported that over 180 features were needed for the recognition using the FERET data set [14] containing 600 images (of 200 subjects) in order to attain similar performance of using 85 features with the ORL dataset containing 400 images (of 40 subjects). In Wu et al. [17], there are 96 dimensions in total; under each dimension, however, it is a filter response of size 20×20 . Therefore, the actual dimension used to calculate similarity is much higher. We therefore proposed a holistic template-matching approach with lower feature dimension to deal with the transformed appearance-based face recognition problem; our proposed approach uses the mean and standard deviation of Gabor filters responses, which is the result of the convolution of Gabor filters with grayscale face images, as features. Since Gabor filters of three scales and six orientations are adopted for feature extraction, for the mean and the standard deviation of filter responses, each is of dimension thirty-six (eighteen each for real and imaginary part of filters' responses); that is there are totally seventy-two dimensions/features. Feature vector of a lower dimensional space can reduce the computation complexity, i.e. increasing speed, as well as the storage required for face recognition process. Moreover, our proposed method is a template-matching approach that does not require any training/learning beforehand.

Assuming the image is of size $Y \times X$ and based on Eqs. (5), (6) and (9), the mean (M) of the real part and the imaginary part of filter responses are

$$M_c(\omega, \theta) = \frac{\sum_x \sum_y k_c(x, y, \omega, \theta)}{X \cdot Y} \quad (10)$$

$$M_s(\omega, \theta) = \frac{\sum_x \sum_y k_s(x, y, \omega, \theta)}{X \cdot Y} \quad (11)$$

Correspondingly, the standard deviation (SD) of the real part and the imaginary part of filter responses are as follows.

$$SD_c(\omega, \theta) = \sqrt{\frac{\sum_x \sum_y [k_c(x, y, \omega, \theta) - M_c(\omega, \theta)]^2}{(X \cdot Y) - 1}} \quad (12)$$

$$SD_s(\omega, \theta) = \sqrt{\frac{\sum_x \sum_y [k_s(x, y, \omega, \theta) - M_s(\omega, \theta)]^2}{(X \cdot Y) - 1}} \quad (13)$$

$$FV = [M_c(\omega_1, \theta_1), \dots, M_c(\omega_3, \theta_6), \\ M_s(\omega_1, \theta_1), \dots, M_s(\omega_3, \theta_6), \\ SD_c(\omega_1, \theta_1), \dots, SD_c(\omega_3, \theta_6), \\ SD_s(\omega_1, \theta_1), \dots, SD_s(\omega_3, \theta_6)] \quad (14)$$

L_1 - and L_2 -norm are adopted as the (dis)similarity measures to define how similar is one image, in our testing databases, when it is compared to another. Based on the query face image, the nearest neighbour rule with majority voting scheme helps to determine the best matched individual within the database.

4 Experimental Results

We used the AR face database [12] from Purdue University in our experiment. After careful examination of the database, we found that there are images from 135 subjects (76 men and 59 women) instead of the stated 126, nonetheless. We also found that only 120 subjects (65 men and 55 women) out of the 135 subjects have been taken images in the two sessions. Therefore, we used only those images from the 120 subjects who have been taken photographs in two sessions (with 14 days apart between the two sessions). As there are 13 images are taken in each session, the testing set we used contains 3,120 images (1,560 in each session); each of the 13 images is taken under various conditions: different facial expressions, illumination settings and occlusions (sunglasses and scarf). There is no limitation on the participant's clothes, make-up, ornaments, hairstyles, etc. [12],[10].

We have converted the color images in to gray scale images of size 192×144 as images of our testing database. Since there is only one face per image and the face is roughly located at the center of the image, no face detection is performed.

In our experiment, the images of 50 randomly chosen subjects were used consisting of 25 males and 25 females. For the image set used, the facial features [11], such as the center of eyes, in each of the image are first localized. Then each face is aligned upright based on the center of eyes and the region containing the face is cropped to be of size 55×71; this is analogous to the planar transform suggested in [1]. They are finally warped to a "standard" face [1]; the warping procedure, which is designed to normalize the face and to align the facial features to approximately the same pixels, has been shown to improve recognition results [11],[13]. Afterwards, an oval mask is applied to remove the highly probable background area to extract only responses (for aggregated 2D Gabor Features) or pixels (for PCA) that are within the mask for feature extraction [13].

We experimented with our method under four cases: Cases 1-4: i th session match against j th session (s_i vs. s_j) for $i=1, 2$ and $j=1, 2$. For Cases 1 (S1 vs. S1) and 4 (S2 vs.

S2), images of either 1st session or 2nd session are used as registered/training and query images, i.e. no unseen image is expected. For Cases 2 (S1 vs. S2) and 3 (S2 vs. S1), however, either 1st or 2nd session is the registered set while the other one is the query set, i.e. matching “duplicates” [13]. Another candidate under testing is the first 72 principle components, which is associated with the largest variances, trained with all images obtained in one session, i.e. either S1 or S2. Since we use PCA as a reference for performance comparison, evaluation cases described above are tested such that images in the training sets are from the same session. Moreover, we choose, for PCA, the same feature dimension as that of aggregated 2D Gabor features because we can determine the effectiveness of our proposed method directly from the system performance; we only use one feature dimensionality for experiment as it already captures over 99.75% of the total variability for either session of the two testing sets.

To study the effect of facial expressions, we conducted the experiment in two stages: First, only the four images of the neutral expression (under various illumination conditions) from both sessions, image sets S1n and S2n, are used in stage one; Second, the seven images without occlusion, image sets S1 and S2 (i.e. S1n/S2n plus the three with various facial expressions other than neutral, S1f/S2f), are used in stage two.

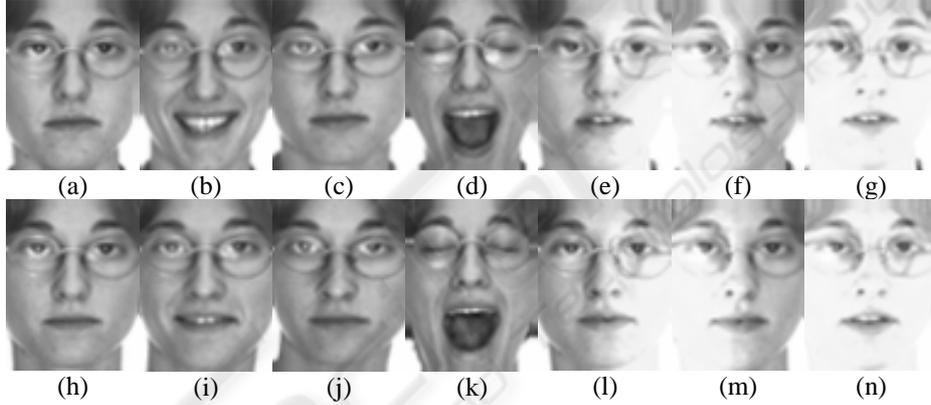


Fig. 1. A set of warped images of one subject in our testing set; (a)-(g) are from 1st session while (h)-(n) are from 2nd session.

A sample set of the fourteen preprocessed (warped) images used is shown in Figure 1. In Figure 1, the first row, i.e. (a)-(g), are images obtained in the first session (S1) and the second row, i.e. (h)-(n), are those obtained in the second session (S2). In stage one of experiment, image sets S1n and S2n, i.e. Figure 1 (a), (e)-(g) and (h), (l)-(n) respectively (400 images in total) are used to test against our proposed method and PCA similarly. In stage two of experiment, image sets S1 and S2, i.e. Figure 1 (a)-(g) and (h)-(n) respectively (700 images in total) are used to test against the two methods.

In stage one of experiment, one test, Test I, is conducted using image sets S1n and S2n to evaluate the performance of the two methods under the four evaluation cases; it is mainly to determine the performance of the two methods on variations in illumination.

In stage two of experiment, three tests, Tests II to IV, are conducted. Test II introduces variations in facial expression in addition to those of Test I. In Test III,

image sets containing only the three with various facial expressions other than neutral, S1f and S2f, are used to query against S1 and S2. Test III is mainly to show that the high accuracy in recognition of PCA is due to the neutral expression images in query sets. Further, in Test IV, we would like to know whether the two methods can withstand facial variations without a priori knowledge by using S1f and S2f to query against S1n and S2n. The average recognition rates for all tests and cases of the two methods are plotted in Figure 2. The k nearest neighbours chosen are $k = \{5, 10, 20\}$ but only $k = 5$ is shown as it always performs the best.

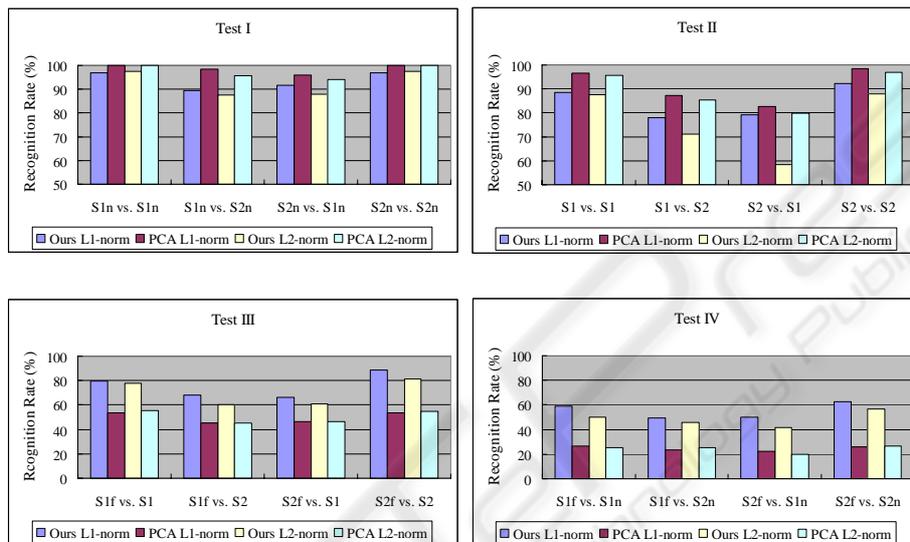


Fig. 2. Results (the average recognition rates) of Tests I to VI in (a)-(f) respectively.

It is expected that PCA will outperform our proposed method in Test I. Since PCA is sensitive to variations in pixels, the warping (that aligns the features), the oval mask (that removes the noisy background) and the testing images, S1n and S2n (only the neutral expression is used, i.e. no variation introduced by facial expression), are all in greater favor to PCA. Both the methods, nevertheless, are expected to attain high recognition rate, normally over 90 percent with a testing image set at this scale [11].

From Tests I and II, it is confirmed that our presumption about the performance of our proposed method and PCA in Test I is correct; PCA outperforms our proposed method in stage one of this experiment. Nonetheless, with the introduction of variations in facial expression, Test II, a larger drop in performance of matching “duplicates” using PCA is observed by comparing Test I with Test II while a relatively mild degradation in performance is recorded for our proposed method. The overall performance of PCA is better, which is mainly due to the composition of the query image set. In the query image set, there are more than half (four out of seven) of the images are of neutral facial expression. The high recognition rate of PCA for neutral faces (see Tests I and II) dominates the performance evaluation. This can be confirmed by Test III.

From Test III, It is seen that our proposed method is more robust than PCA against variations in facial expression. We can further show that our proposed method is more robust to variations in facial expression by Test IV using faces with facial expressions other than neutral as query images to match against a database that contains only faces with neutral facial expression. Since the database does not contain any faces with facial expression other than neutral, they are unseen to PCA, i.e. not trained. Test IV shows that PCA is ineffective to adapt variations in facial expression if no faces with varied facial expression are presented in training whereas our proposed method can still achieve a moderate accuracy, even matching the “duplicates”.

5 Conclusion

We have proposed a relatively low feature dimension scheme for face recognition for frontal face images captured indoor. We derived an aggregated Gabor filter responses representation for face images that is of lower feature dimension. A comparative study on the effect of using different (dis)similarity measures, L_1 - and L_2 -norm, and the effect of matching against “duplicates” is reported for our proposed method; a similar study is done using PCA for performance comparison. A study of the influence of the facial expressions has shown that PCA is not as effective as our proposed method in recognizing face with varied facial expressions. The better performance of PCA is a result of the uneven composition of images (of various facial expressions) in testing set and its ability in matching neutral expression images. The performance for PCA is even worse that if none of the images with variations in facial expression is included during the training phase, PCA can hardly recognize a person correctly.

Acknowledgement

The authors would like to thank for the partial support of the research grants from Hong Kong Government (UGC) and The Hong Kong Polytechnic University.

References

1. Beymer, D.: Vectorizing Face Images by Interleaving Shape and Texture Computations. AI Memo No. 1537, AI Lab., Massachusetts Inst. of Technology (1995)
2. Bovik, A.C. Clark, M., Geisler, W.S.: Multi-channel texture analysis using localized spatial filters. IEEE Trans. Pattern Anal. Machine Intell., Vol. 12 No. 1 (1990) 55-73
3. Daugman, J.G.: Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. Journal of the Optical Society of America A., Vol. 2 No. 7 (1985) 1160-1169
4. Daugman, J.G.: Complete discrete 2-d Gabor transforms by neural networks for image analysis and compression IEEE Trans. Acoust., Speech, Signal Processing, Vol. 36 No. 7 (1988) 1169-1179
5. Draper, B.A., Baek, K., Bartlett, M.S., Beveridge, J.R.: Recognizing faces with PCA and ICA. Computer Vision and Image Understanding, Vol. 91 No.1-2 (2003) 115-137
6. Heisele, B., Ho, P., Wu, J., Poggio, T.: Face Recognition: component-based versus global approaches. Computer Vision and Image Understanding, Vol. 91 No. 1-2 (2003) 6-21

7. Jain, A.K., Duin, R.P.W., Mao, J.: Statistical Pattern Recognition: A Review. *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 22 No. 1 (2000) 4-37
8. Jain, A.K., Ross, A., Prabhakar, S.: An Introduction to Biometric Recognition. *IEEE Trans. Circuits Syst. Video Technol.*, Vol. 14 No. 1 (2004) 4-20
9. Lee, T.S.: Image representation using 2D Gabor wavelet. *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 18 No. 10 (1996) 959-971
10. Liu, C., Wechsler, H.: Independent Component Analysis of Gabor Features for Face Recognition. *IEEE Trans. Neural Networks*, Vol. 14 No. 4 (2003) 919-928
11. Martinez, A.M.: Recognizing Imprecisely Localized, Partially Occluded, and Expression Variant Faces from a Single Sample per Class. *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 24 No. 6 (2001) 748-763
12. Martinez, A.M., Benavente, R.: The AR face database. CVC Tech. Report #24, Purdue Univ., Dept. of Electrical Eng. (1998); please see also: http://rv11.ecn.purdue.edu/~aleix/aleix_face_DB.html
13. Martinez, A.M. Kak, A.C.: PCA versus LDA. *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 23 No. 2 (2001) 228-233
14. Phillips, P.J., Moon, H., Rauss, P.J., Rizvi, S.: The FERET Evaluation Methodology for Face Recognition Algorithms. *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 22 No. 10 (2000) 1090-1104
15. Turk, M., Pentland, A.: Eigenfaces for recognition. *J. Cognitive Neuroscience*, Vol. 3 No. 1 (1991) 71-86
16. Wiskott, L., Fellous, J.M., Kruger, N., von der Malsburg, C.: Face Recognition by Elastic Bunch Graph Matching. *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 19 No. 7 (1997) 775-779
17. Wu, H., Yoshida, Y., Shioyama, T.: Optimal Gabor Filters for High Speed Face Identification. *Proc. of 16th International Conference on Pattern Recognition, Quebec City*, Vol. 1 (2002) 107-110
18. Zhang, D., Peng, H., Zhou, J., Sankar, K.P.: A Novel Face Recognition System Using Hybrid Neural and Dual Eigenspaces Methods. *IEEE Trans. Syst. Man, Cybern. A*, Vol. 32 No. 6 (2002) 787-793
19. Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A.: Face Recognition: A Literature Survey. *ACM Computing Surveys*, Vol. 35 No. 4 (2003) 339-458

