Computationally Efficient Statistical Face Model in the Feature Space

Mohammad Haghighat
Department of ECE
College of Engineering
University of Miami
Coral Gables, FL, USA
Email: haghighat@umiami.edu

Mohamed Abdel-Mottaleb
Department of ECE
College of Engineering
University of Miami
Coral Gables, FL, USA
and Adjunct, Effat University
Email: mottaleb@miami.edu

Wadee Alhalabi
Department of CS
College of Engineering
Effat University
Jeddah, Saudi Arabia
Email: walhalabi@effatuniversity.edu.sa

***Abstract*—In this paper, we present a computationally efficient statistical face modeling approach. The efficiency of our proposed approach is the result of mathematical simplifications in the core formula of a previous face modeling method and the use of the singular value decomposition. In order to reduce the errors in our resulting models, we preprocess the facial images to normalize for pose and illumination and remove little occlusions. Then, the statistical face models for the enrolled subjects are obtained from the normalized face images. The effects of the variations in pose, facial expression, and illumination on the accuracy of the system are studied. Experimental results demonstrate the reduction in the computational complexity of the new approach and its efficacy in modeling the face images.**

Fig. 1. Changes in the face images of an individual.

I. Introduction

Biometric identification has been one of the challenging topics in computer vision for the past few decades. Changes in the biometric images due to changes in pose, illumination and occlusion make the biometric recognition a stochastic problem. In order to better analyze a stochastic process, one needs statistical models. Statistical modeling provides a formulation to understand how a group of random variables are similar to each other or how they are different from other groups of variables. The face images of an individual subject are similar to each other and different from the face images of other subjects. However, face images of an individual are not exactly the same either. Changes in illumination, head pose, facial expressions and cosmetics as well as aging and occlusions, change the appearance of the face. Fig. 1 clearly illustrates these variations for a single subject. The question is how these changes are different from the changes between different subjects.

Defining a deterministic model for the face is not suitable because of the changes mentioned above. These changes create difficulties even for stochastic models. However, normalization can be used in a preprocessing step to reduce the effect of some of these changes.

The well known Active Shape Model (ASM) [1] has been widely used for face modeling. Training the ASM requires a large number of facial images where the feature points are manually labeled. By analyzing the statistics of the locations of the labeled points from a large set of facial images, a point distribution model is built. The model consists of the average locations of the points and some parameters that control the changes in the training set. This approach was later improved by Edwards *et al.* in [2]–[4]. It also inspired Lanitis *et al.* [5] to create an aging function of the model to study aging effects on face images.

In this paper, we present a statistical model for face images in the feature space. Our modeling approach is more computationally efficient than the one in [1], [5] and works in the feature domain. The reduction in the computational complexity comes not only from the use of singular value decomposition (SVD) in the eigendecomposition process, but also from simplifying the modeling equations of [1] and [5]. In order to build a more accurate model, we present a fully automatic preprocessing approach which is capable of removing noise and small occlusions and normalizes for illumination and in-plane rotations of the face. We also study the effect of the variations in pose, facial expression, and illumination on the accuracy of the system.

This paper is organized as follows. Section II describes our face normalization technique as a preprocessing step before face modeling. Section III presents our proposed method for face modeling. The implementation details and experimental results are presented in Section IV. Finally, Section V concludes the paper.

II. Preprocessing for Face Normalization

In this section, we present a preprocessing approach, which is capable of normalizing the facial images for pose and illumination and removing little occlusions.

A. Pose Normalization

Pose variations cause major problems in real-world face recognition systems. Since the human face is approximately symmetric, if it is in frontal pose with no rotations and occlusions, the matrix containing the face image (F) will have the lowest rank. In order to normalize a rotated face, we employ *transform invariant low-rank textures (TILT)* method proposed by Zhang *et al.* in [6].

If we model little occlusion or noise with an error matrix, E, TILT tries to find a transformation (Euclidean, affine, or projective) matrix, τ, such that $\tilde{F} \circ \tau = F + E$, where $\tilde{F}$ is the original face image and F is the corrected low-rank face image, which corresponds to the frontal face. This equation is solved by the following optimization problem:

$$\min_{F,E,\tau} \quad rank(F) + \gamma ||E||_o \quad s.t. \quad \tilde{F} \circ \tau = F + E \tag{1}$$

where $||E||_o$ is the l^0-norm of the error matrix, *i.e.*, number of non-zero elements. It actually looks for the F with the lowest possible rank and the error with the lowest number of non-zero elements, which satisfy the above condition. γ trades off the rank of the matrix with the sparsity of the error.

Optimizing the rank function and the l^0-norm in the above equation is very difficult. Therefore, they are substituted by their convex surrogates. Since the rank of a matrix is equal to the number of its non-zero singular values, we can substitute the $rank(F)$ by its nuclear norm $||F||_*$, which is the sum of its singular values. On the other hand, l^0-norm is substituted by l^1-norm, which is the sum of the absolute values of the elements of the matrix.

Another problem with the above equation is that the constraint, $\tilde{F} \circ \tau = F + E$, is not linear. This problem is solved by linearizing the constraint around its current estimate, which is updated in an iterative process. [7]. With these approximations, the optimization problem turns into:

$$\min_{F,E,\triangle\tau} \quad ||F||_* + \gamma ||E||_1 \quad s.t. \quad \tilde{F} \circ \tau + \nabla \tilde{F} \triangle \tau = F + E \tag{2}$$

where ∇ is the Jacobian.

We modified the above algorithm to be more suitable for face images. The TILT code, kindly provided by the authors [8], requires the user to specify the texture area manually. However, in case of face images, we modified the code so that the reigon of interst is automatically extracted by Viola-Jones face detection method [9]. On the other hand, our experiments show that the forehead hairline misleads the TILT algorithm since the hairline is not necessarily horizontal. Therefore, we trained the Viola-Jones face detector to restrict the face area vertically between the upper side of the eyebrows and the lower lip, and horizontally between the two cheeks. Another advantage of this restriction is that it reduces the background in the extracted region of the face.

Figure 2 shows the results of TILT on some rotated face images from CMU frontal face images database [10]. In this paper, we also restrict the transformation matrix, τ, to be Euclidean. Based on our experiments, affine and projective transformations do not work well in normalizing face images.

Fig. 2. Pose correction using TILT. Top rows: original face images with rotations ($\tilde{F}$); Bottom rows: corrected face images (F).

B. Illumination Normalization

Another major problem that affects face recognition is the illumination variations. To reduce this effect, we normalize the face images using a method based on Weber's Law proposed in [11]. When in a quiet room, one can hear a whispered voice, however, in a noisy place one may not even hear a shouting. The German physician Ernst Heinrich Weber proposed Weber's law based on this phenomenon. He assumed that the ratio between the smallest noticeable change in a stimulus and the original level of the stimulus is constant [12]:

$$\frac{\Delta S_{min}}{S} = k \tag{3}$$

where k is called the *Weber fraction*. Weber's law is almost true for different types of stimuli such as light and sound intensity. It states that stimuli perception is not based on the absolute value of the change in the intensity, but rather is based on the relative change of the intensity level of the stimuli with respect to the original level.

Weber local descriptor (WLD) in a 3×3 neighborhood of the face image is defined as the ratio of the difference in intensity between the current pixel and its neighbors to the intensity of the current pixel [13]:

$$\xi(x,y) = arctan\left(\alpha \sum_{i=-1,0,1} \frac{F(x,y) - F(x-i,y-i)}{F(x,y)}\right) \tag{4}$$

where α is an adjusting parameter for the intensity difference between neighboring pixels, and the arctangent function is used to reduce the effect of noise by preventing the output from being too large.

Based on the Lambertian reflectance model, a face image $F(x,y)$ can be factorized as:

$$F(x,y) = R(x,y)I(x,y) \tag{5}$$

where $R(x,y)$ and $I(x,y)$ are the reflectance and illuminance factors at pixel (x,y). R is the illumination insensitive part, which depends only on the characteristics of the facial surface

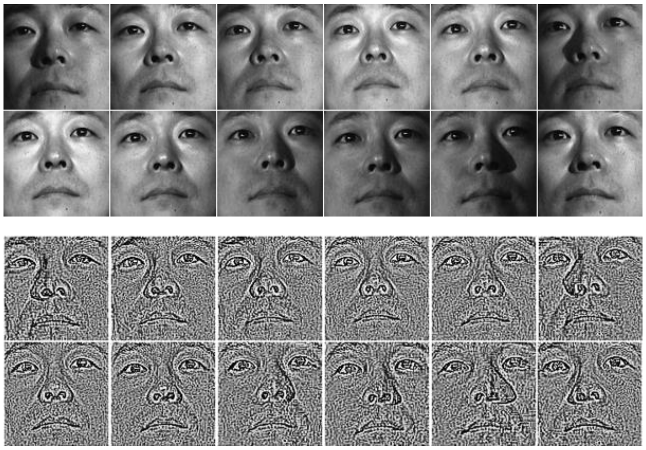

Fig. 3. Illumination normalization using Weber-Face.

including the surface texture (albedo) and the surface normal (i.e., 3D shape).

From equation 5, we have:

$$F(x-i,y-i) = R(x-i,y-i)I(x-i,y-i). \quad (6)$$

The illuminance factor $I(x,y)$ usually varies very slowly between adjacent pixels. Therefore, we assume

$$I(x-i,y-i) \approx I(x,y). \quad (7)$$

By substituting Equations 5, 6, and 7 into 4, we obtain:

$$\xi(x,y) = arctan\left(\alpha \sum_{i=-1,0,1} \frac{R(x,y)-R(x-i,y-i)}{R(x,y)}\right) \quad (8)$$

which shows that $\xi(x,y)$ depends only on the reflectance factor and can be treated as the illumination insensitive representation of the face image. ξ, defined in Equation 4, is called *Weber-face* which is used in our experiments as the normalized face image. Fig. 3 shows some examples of the normalized face images using Weber-face.

III. Statistical Face Model

In the proposed approach, the statistical face models are generated using the feature vectors extracted from the images of the enrolled subjects. Lets assume that we have n training samples and we extract feature vectors of length d from each sample. We subtract the mean feature vector from each of the feature vectors and put the resulting vectors in a d by n matrix X. The methods presented in [1] and [5] compute the eigenvectors by applying principal component analysis (PCA) to XX^T:

$$(XX^T)_{(d\times d)} V_{(d\times d)} = \lambda_{(d\times d)} V_{(d\times d)}. \quad (9)$$

The eigenvector matrix, V, plays the role of a projection matrix, P, that can reconstruct each training sample, X_i, from a vector of weights, b_i, called *model parameters*:

$$X_{(d\times n)} = P_{(d\times d)} b_{(d\times n)}. \quad (10)$$

Model parameters are calculated by solving the above equation for all training samples:

$$b_{(d\times n)} = P^{-1}_{(d\times d)} X_{(d\times n)}. \quad (11)$$

If the number of features is higher than the number of samples ($d \gg n$), it is computationally easier to calculate the covariance matrix as $(X^TX)_{(n\times n)}$ rather than $(XX^T)_{(d\times d)}$. The eigenvectors, $U_{(n\times n)}$, of this matrix are defined as follows:

$$(X^TX)_{(n\times n)} U_{(n\times n)} = \lambda_{(n\times n)} U_{(n\times n)}. \quad (12)$$

where λ is a diagonal matrix containing the eigenvalues. If we multiply the above equation by X from the left hand side, we obtain:

$$XX^T \underbrace{XU} = \lambda \underbrace{XU}. \quad (13)$$

where XU is the matrix of eigenvectors of the covariance matrix XX^T. If we select the r significant eigenvectors corresponding to the r largest eigenvalues, the projection matrix will be:

$$P_{(d\times r)} = X_{(d\times n)} U_{(n\times r)}, \quad (14)$$

and from Equation 10

$$X_{(d\times n)} = P_{(d\times r)} b_{(r\times n)}. \quad (15)$$

However, since P is not a square matrix, in order to calculate model parameters from Equation 11, we will use the left pseudoinverse of the eigenvectors matrix:

$$b_{(r\times n)} = P^{\dagger}_{(r\times d)} X_{(d\times n)}. \quad (16)$$

Left pseudoinverse of a full rank matrix is defined as:

$$P^{\dagger} = (P^TP)^{-1}P^T. \quad (17)$$

Substituting Equation 14 into 17 and considering Equation 12, the above equation reduces to:

$$P^{\dagger} = (U^T \underbrace{X^TXU}_{\lambda U})^{-1} U^TX^T = U^T\lambda^{-1} \underbrace{UU^T}_{I} X^T. \quad (18)$$

Note that since X^TX is symmetric positive definite, its eigenvectors are orthogonal, *i.e.*, $U^T = U^{-1}$. Therefore:

$$P^{\dagger} = U^T\lambda^{-1}X^T. \quad (19)$$

Substituting in Equation 16, we obtain

$$b_{(r\times n)} = U^T\lambda^{-1}X^TX. \quad (20)$$

If we multiply this equation by an identity matrix, $I_{(n\times n)} = UU^T$, we obtain the following simplified equation for calculating the model parameters:

$$b_{(r\times n)} = U^T\lambda^{-1} \underbrace{X^TXU}_{\lambda U} U^T = U^T_{(r\times n)}. \quad (21)$$

Therefore, the above equation shows that the model parameters can be easily calculated by the transpose of the matrix of the significant eigenvectors.

We use singular value decomposition (SVD) to calculate U^T. Using SVD is numerically more efficient than eigendecomposition using PCA. It does not require the formation of X^TX matrix, which is not only computationally expensive but also can cause loss of precision [14], [15]. As Equation

22 shows, U^T is extracted easily from the singular value decomposition of matrix X:

$$X_{(d\times n)} = V_{(d\times d)} \Sigma_{(d\times n)} U^T_{(n\times n)} \,. \tag{22}$$

Now, we have model parameters, b_i, for all the enrolled samples. Since each enrolled subject might have multiple samples, the statistical model of a subject is obtained by computing the centroid of the model parameters of the training samples of that subject.

Given a query sample, we first normalize the face image and compute its statistical model parameters using Equation 16:

$$b_{(r\times 1)} = P^{\dagger}_{(r\times d)} X_{(d\times 1)} \,, \tag{23}$$

where $P^{\dagger}$ is calculated using Equation 19. The query sample is then classified as the nearest neighbor based on the distance between the query's model and the models of the subjects in the gallery.

The algorithm in [1] and [5] has four expensive steps: 1) calculating the covariance matrix XX^T, 2) applying the eigendecomposition using PCA, 3) inverting the eigenvector matrix, and 4) calculating the model parameters using Equation 11. However, our algorithm just applies the singular value decomposition on matrix X, which is significantly faster to calculate. Furthermore, since our algorithm does not require the matrix inversion step, it avoids the problems that arise with singular matrices in the so-called *small sample size* (SSS) scenarios, where the number of available samples is smaller than the dimensionality of the feature space.

Note that the selection of r is a degree of freedom for the designer and it can be defined, considering the *reconstruction error*, as [16]:

$$e_{rec} = 1 - \frac{\sum_{i=1}^{r} \lambda_i}{\sum_{i=1}^{t} \lambda_i} \,. \tag{24}$$

where λ_is are sorted in a descending order and t is the total number of non-zero eigevalues, *i.e.*, the rank of matrix X.

IV. Experimental Results

The performance of our proposed model was evaluated using the Facial Recognition Technology (FERET) database [17]. There are only 200 subjects in FERET database who have images with pose, expression, and illumination variations. In our experiments, 1400 face images for these 200 subjects were selected, *i.e.*, seven images per subject. Three of the images are frontal faces with different facial expressions and illumination. These images are letter coded as *ba*, *bj*, and *bk*. The other four images are faces in different poses with $+25°$, $+15°$, $-15°$, and $-25°$ degrees of rotation. These images are letter coded as *bd*, *be*, *bf*, and *bg*, respectively.

Viola-Jones face detection method [9] was used to extract the face regions from the images. The first row of Fig. 4 shows the face detection results for a sample set of images of one subject from the database. The results of pose and illumination normalization of these face images are shown in the second and third rows, respectively. The size of the face images used in our experiments is 120×120 pixels.

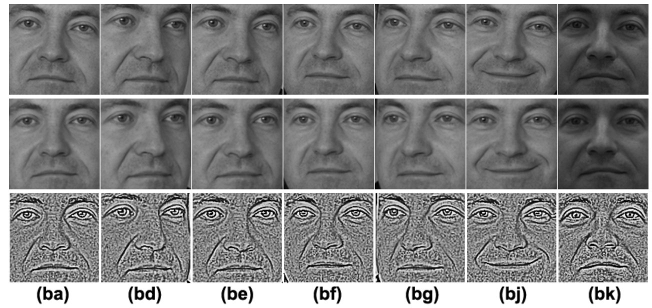

Fig. 4. Face samples of a subject from FERET database (first row). Results of the pose correction (second row). Illumination normalization (third row).

TABLE I. Recognition rates of the proposed method in confrontation with different face distortions.

Code	Description	Recognition Rate
bd	$+25°$ pose change	93.50 %
be	$+15°$ pose change	99.00 %
bf	$-15°$ pose change	99.50 %
bg	$-25°$ pose change	96.50 %
bj	Alternative expression	95.50 %
bk	Different illumination	99.50 %

After normalizing the face images, we extract the Gabor wavelet features to construct the feature vectors [18], [19]. Here we used forty Gabor filters in five scales and eight orientations. Since the adjacent pixels in an image are usually highly correlated, we can reduce this information redundancy by downsampling the feature images that result from Gabor filters [20]. In our experiments, the feature images were downsampled by a factor of four. That is, the length of each feature vector, X_i, is $d = (120\times 120)\times(5\times 8)/(4\times 4) = 36000$. After applying whitening transform on all the feature vectors, to have zero mean and unit variance, the model parameters are calculated using the proposed technique.

Two different experiments were performed on the database. In the first one, we obtained the face models from only one frontal face image of each subject. The frontal face image with neutral expression, labeled *ba*, was used for enrollment and obtaining the model parameters. The remaining six images with different poses, expressions, and illumination were used for testing. Fig. 5 shows the recognition accuracy based on the proposed model for different number of features. Each of the plots corresponds to one of the letter coded sets of images, as shown in Fig. 4 in FERET database, used for testing.

The maximum recognition accuracy, over the number of features, for each set is shown in Table I. It is obvious that the proposed face model copes very well with small pose changes and variations in illumination. However, the accuracy decreases in cases of large pose changes and facial expression variations. For the purpose of comparison, the maximum accuracy of the algorithm without any preprocessing is shown in Table II.

In the second experiment, we use four samples to enroll each subject and compute the model parameters, then evaluate the recognition results using the remaining three samples. In order to have more accurate results, a seven-fold cross validation is applied. Fig. 6 and Table III show the recognition rate of the proposed approach in case of different prepro-

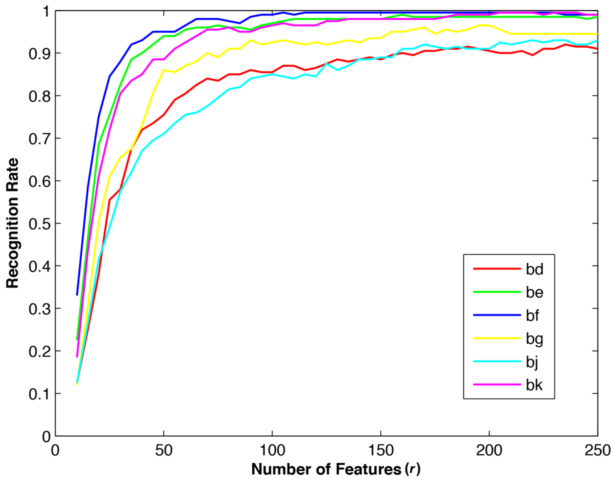

Fig. 5. Face recognition using different face poses, expresions and illumination. Legends are the letter codes of the database described in Table I and Fig. 4.

TABLE II. RECOGNITION RATES OF THE PROPOSED METHOD USING GABOR FEATURES OF THE RAW FACE IMAGES WITH NO PREPROCESSING.

Code	Description	Recognition Rate
bd	$+25°$ pose change	88.00 %
be	$+15°$ pose change	98.00 %
bf	$-15°$ pose change	98.50 %
bg	$-25°$ pose change	86.00 %
bj	Alternative expression	95.00 %
bk	Different illumination	97.50 %

cessing methods. It is clear that both illumination and pose normalizations increase the recognition accuracy of the system when applied separately. The results are further improved when applying them together.

We compared the processing time of our modeling algorithm with that of the algorithm in [1] and [5]. For an X matrix of size 1400×36000, the algorithm in [1] and [5] takes 54.85 seconds, while our proposed algorithm, takes 20.08 seconds, averaged over multiple runs.

V. Conclusion

In this paper, we presented a computationally efficient statistical face model by applying mathematical simplifications to the core formula of a previously proposed face modeling approach. The proposed method works in the feature domain and has a proven computational improvement in comparison with a previously proposed statistical face modeling approach. Before modeling, we preprocess the face images to normalize for pose and illumination. Experimental results demonstrated the computational efficiency and robust performance of our proposed model on FERET database. In the future, we will address the issue of how to enhance the performance of the model in cases of larger pose variations and changes in facial expressions.

TABLE III. MAXIMUM FACE RECOGNITION RATES IN THE SECOND EXPERIMENT.

Pre-processing	Recognition Rate
No Pre-processing	96.00 %
WeberFace	97.36 %
TILT	97.50 %
TILT+WeberFace	98.02 %

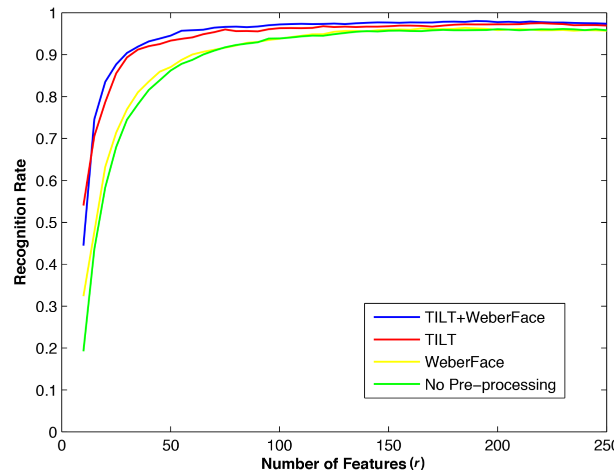

Fig. 6. Face recognition rates in the second experiment.

References

[1] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models-their training and application," *Computer vision and image understanding*, vol. 61, no. 1, pp. 38–59, 1995.

[2] G. J. Edwards, A. Lanitis, C. J. Taylor, and T. F. Cootes, "Statistical models of face imagesimproving specificity," *Image and Vision Computing*, vol. 16, no. 3, pp. 203–211, 1998.

[3] G. J. Edwards, C. J. Taylor, and T. F. Cootes, "Learning to identify and track faces in image sequences," in *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on.* IEEE, 1998, pp. 260–265.

[4] ——, "Interpreting face images using active appearance models," in *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on.* IEEE, 1998, pp. 300–305.

[5] A. Lanitis, C. J. Taylor, and T. F. Cootes, "Toward automatic simulation of aging effects on face images," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 4, pp. 442–455, 2002.

[6] Z. Zhang, A. Ganesh, X. Liang, and Y. Ma, "TILT: Transform invariant low-rank textures," *International Journal of Computer Vision*, vol. 99, no. 1, pp. 1–24, 2012.

[7] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma, "RASL: Robust alignment by sparse and low-rank decomposition for linearly correlated images," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 11, pp. 2233–2246, 2012.

[8] "TILT: Transform invariant low-rank textures," http://perception.csl.illinois.edu/matrix-rank/tilt.html, accessed: 2014-03-18.

[9] P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.

[10] H. A. Rowley, S. Baluja, and T. Kanade, "CMU Frontal Face Images Database," http://vasc.ri.cmu.edu/idb/html/face/frontal_images/, accessed: 2014-03-18.

[11] B. Wang, W. Li, W. Yang, and Q. Liao, "Illumination normalization based on weber's law with application to face recognition," *Signal Processing Letters, IEEE*, vol. 18, no. 8, pp. 462–465, 2011.

[12] A. K. Jain, "Fundamentals of digital signal processing," *Fundamentals of Digital Signal Processing*, 1989.

[13] J. Chen, S. Shan, C. He, G. Zhao, M. Pietikainen, X. Chen, and W. Gao, "WLD: A robust local image descriptor," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 9, pp. 1705–1720, 2010.

[14] L. N. Trefethen and D. Bau III, *Numerical linear algebra.* SIAM, 1997, vol. 50.

[15] P. Läuchli, “Jordan-elimination und ausgleichung nach kleinsten quadraten,” *Numerische Mathematik*, vol. 3, no. 1, pp. 226–240, 1961.

[16] M. B. A. Haghighat and E. Namjoo, “Evaluating the informativity of features in dimensionality reduction methods,” in *Application of Information and Communication Technologies (AICT), 2011 5th International Conference on*. IEEE, 2011, pp. 1–5.

[17] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, “The FERET evaluation methodology for face-recognition algorithms,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 10, pp. 1090–1104, 2000.

[18] B. S. Manjunath and W.-Y. Ma, “Texture features for browsing and retrieval of image data,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 18, no. 8, pp. 837–842, 1996.

[19] M. Haghighat, S. Zonouz, and M. Abdel-Mottaleb, “Identification using encrypted biometrics,” in *Computer Analysis of Images and Patterns*. Springer, 2013, pp. 440–448.

[20] C. Liu and H. Wechsler, “Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition,” *Image processing, IEEE Transactions on*, vol. 11, no. 4, pp. 467–476, 2002.