ISSN No. 0976-5697

Volume 1, No. 3, Sept-Oct 2010

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

An Efficient Algorithm for Multimodal Biometric Face Recognition using Speech Signal

Nageshkuymar M.*
Department of Electronics and Communication Engg.
J.S.S. Research Foundation
Mysore, India
nageshkumar79m@gmail.com

Shanmukha swamyM. N.
Department of Electronics and Communication Engg.
J.S.S. Research Foundation
Mysore, India
mnsjce@gmail.com

Abstract: Multimodal biometric systems for today's high security applications must meet stringent performance requirements. The fusion of multiple biometrics helps to minimize the system error rates. Fusion methods include processing biometric modalities sequentially until an acceptable match is obtained. More sophisticated methods will combine scores from separate classifiers for each modality. Multi algorithm approach employs a single biometric sample acquired from single sensor. Two or more different algorithms process this acquired sample. The individual results are combined to obtain an overall recognition result. This approach is attractive, both from an application and research point of view. In recent years much advancement have been made in face recognition techniques to cater to the challenges such as pose, expression, illumination, aging and disguise. However, due to advances in technology, there are new emerging challenges for which the performance of face recognition systems degrades and plastic/cosmetic surgery is one of them. In this paper we comment on the effect of plastic surgery face image in multimodal biometric face recognition using text dependent speech signal.

Keywords: Multimodal biometric system; plastic surgery face image; speech signal and matching level fusion.

I. INTRODUCTION

Biometrics refers to the physiological or behavioral characteristics of a person to authenticate his/her identity. The increasing demand of enhanced security systems has led to an unprecedented interest in biometric based person recognition system. Single biometric system may not be able to achieve the desired performance requirement in real world applications. One of the methods to overcome these problems is to make use of multimodal biometric recognition systems, which combines information from multiple modalities to arrive at a decision.

A generic biometric system has sensor module to capture the trait, feature extraction module are used to process the data to extract a feature set that yields compact representation of the trait, classifier module are used to compare the extracted feature set with reference database to generate matching scores and decision module to determine an identity or validate a claimed identity. In multimodal biometric system information reconciliation can occur at the data level or at feature level, at the match score level generated by multiple classifiers pertaining to different modalities and at the decision level.

Multi-algorithmic biometric systems take a single sample from a single sensor and process that sample with two or more different algorithms. The technique could be applied to any modality. Algorithms can be designed to optimize performance under different circumstances.

A multimodal biometric face recognition is a well studied problem in which several approaches have been proposed to address the challenges of illumination [1,2], pose [3, 4, 5], expression [2], aging [6, 7] and disguise [8, 9], the growing popularity of plastic surgery introduces new challenges in designing future face recognition systems. Since these procedures modify both the shape and texture of facial features to varying degrees, it is difficult to find the correlation between pre and post surgery facial geometry. To the best of our knowledge, there is no study that demonstrates any scientific experiment for recognizing faces that have undergone local or global plastic surgery. The major reasons for the problem not being studied are: (i) Due to the sensitive nature of the process and the privacy issues involved, it is extremely difficult to prepare a face database that contains images before and after surgery. (ii) After surgery, the geometric relationship between facial features changes and there is no technique to detect and measure such type of alterations.

The main aim of this paper is to add a new dimension to plastic surgery face recognition by using speech signal and discussing this challenge and systematically evaluating the performance of existing faces recognition algorithms on a database that contains face images before and after surgery.

A. *Related Work*

A number of studies showing the advantages of multimodal biometrics have appeared in the literature. Brunelli and Falavigna [10] used hyperbolic tangent (tanh) for normalization and weighted geometric average for fusion of voice and face biometrics. They also proposed a hierarchical combination scheme for a multimodal identification system. Kittler et al. [11] have experimented with several fusion techniques for face and voice biometrics, including sum, product, minimum, median, and maximum rules and they have found that the sum rule outperformed others. Kittler et al. [11] note that the sum rule is not significantly affected by the probability estimation errors and this explains its superiority.

Hong and Jain [12] proposed an identification system based on face and fingerprint, where fingerprint matching is applied after pruning the database via face matching. Ben-Yacoub et al. [13] considered several fusion strategies, such as support vector machines, tree classifiers and multi-layer perceptions, for face and voice biometrics. The Bayes classifier is found to be the best method. Ross and Jain [14] combined face, fingerprint and hand geometry biometrics with sum, decision tree and linear discriminant-based methods. The authors report that sum rule outperforms others.

Kittler [15] evaluated several classifier combination rules on frontal face, face profile, and voice biometrics (using a database of 37 subjects). They found that the "sum of *a posteriori* probabilities" rule outperformed the product, min, max, median, and majority of *a posteriori* probability rules (at EER) due to its resilience to errors in the estimation of the densities.

Ben-Yacoub [16] evaluated five binary classifiers on combinations of three face and voice modalities (database of 295 subjects). They found that (a) a support vector machine and Bayesian classifier achieved almost the same performances; and (b) both outperformed Fisher's linear discriminant, a C4.5 decision tree, and a multilayer perception.

Fierrez-Aguilar [17] found that a support vector machine outperformed (at EER) the sum of normalized scores when fusing face, fingerprint and signature biometrics (database of 100 subjects and 50 chimeras).

The rest of this paper is organized as follows. Section 2 presents the proposed Diagonal PCA and 2DPCA methods for face feature extraction. Section 3 presents the speech feature extraction method using MFCC / VQ. Section 4 presents the fusion at the matching score Level. Section 5 reports on the experimental results. Finally, Section 6 concludes.

II. FACE FEATURE EXTRACTION

Recently a new technique called 2-dimensional principal component analysis (2DPCA) is proposed to solve the face recognition problems [18]. The main idea behind 2DPCA is to find the optimal projective vectors in the row direction of images without the image-to-vector transformation. That is, it will first constructs, the so-called *image covariance matrix* from rows of images and then computes its eigenvectors as the projection vectors. Since the size of the image covariance matrix is equal to the width of images, which is quite small compared with the size of the covariance matrix used in PCA. 2DPCA evaluates the image covariance matrix more accurately and computes the corresponding eigenvectors more efficiently than PCA.

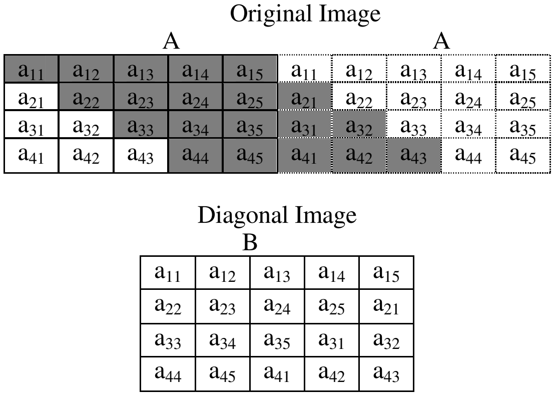
However, the projective vectors of 2DPCA only reflect variations between rows of images by omitting the variations between columns of images. While the omitted variations between columns of images are also useful for recognition purpose. In that case, 2DPCA can hardly obtain improved accuracy. In this paper, a novel method called diagonal principal component analysis (Diagonal PCA) is proposed. Diagonal PCA seeks the optimal projective vectors from *diagonal face images* and therefore the correlations between variations of rows and those of columns of images can be kept.

Our motivation for developing the Diagonal PCA method originates from an essential observation on the recently proposed 2DPCA [28]. That is, 2DPCA can be seen as the row-based PCA, which has been pointed out in [19]. So 2DPCA only reflects the information between rows, which implies some structure information (e.g. regions of a face like eyes, nose, mouth etc.) cannot be uncovered by it. We attempt to solve that problem by transforming the original face images into corresponding *diagonal face images*. Because the rows (columns) in the transformed diagonal face images simultaneously integrate the information of rows and columns in original images, it can reflect both information between rows and those between columns. Through the entanglement of row and column information, it is expected that Diagonal PCA may find some useful block or structure information for recognition in original images.

Suppose that there are *M* training face images, denoted by *m* by *n* matrices $A_k(k=1,2,....M)$

For each training face image, define the corresponding *diagonal face image* as follows:

- If the height *m* is equal to or smaller than the width *n*, use the method illustrated in Fig.1 to generate the diagonal image *B* for the original image *A*.



Original Image

A A

a_{11}	a_{12}	a_{13}	a_{14}	a_{15}	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}
a_{21}	a_{22}	a_{23}	a_{24}	a_{25}	a_{21}	a_{22}	a_{23}	a_{24}	a_{25}
a_{31}	a_{32}	a_{33}	a_{34}	a_{35}	a_{31}	a_{32}	a_{33}	a_{34}	a_{35}
a_{41}	a_{42}	a_{43}	a_{44}	a_{45}	a_{41}	a_{42}	a_{43}	a_{44}	a_{45}

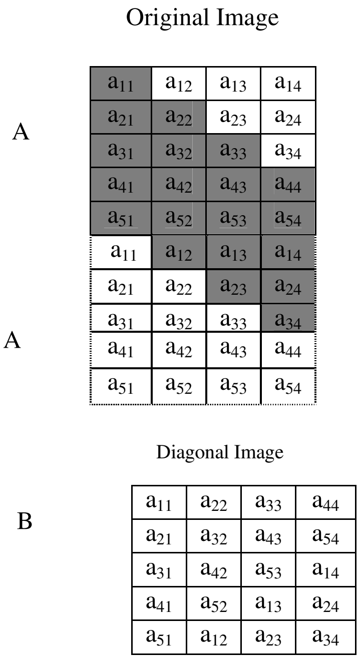
Diagonal Image

B

a_{11}	a_{12}	a_{13}	a_{14}	a_{15}
a_{22}	a_{23}	a_{24}	a_{25}	a_{21}
a_{33}	a_{34}	a_{35}	a_{31}	a_{32}
a_{44}	a_{45}	a_{41}	a_{42}	a_{43}

Figure.1 Illustration for deriving the diagonal face images

- If the height *m* is bigger than the width *n*, use the method illustrated in Fig.2 to generate the diagonal image *B* for the original image *A*.



Original Image

A

A

a_{11}	a_{12}	a_{13}	a_{14}
a_{21}	a_{22}	a_{23}	a_{24}
a_{31}	a_{32}	a_{33}	a_{34}
a_{41}	a_{42}	a_{43}	a_{44}
a_{51}	a_{52}	a_{53}	a_{54}
a_{11}	a_{12}	a_{13}	a_{14}
a_{21}	a_{22}	a_{23}	a_{24}
a_{31}	a_{32}	a_{33}	a_{34}
a_{41}	a_{42}	a_{43}	a_{44}
a_{51}	a_{52}	a_{53}	a_{54}

Diagonal Image

B

a_{11}	a_{22}	a_{33}	a_{44}
a_{21}	a_{32}	a_{43}	a_{54}
a_{31}	a_{42}	a_{53}	a_{14}
a_{41}	a_{52}	a_{13}	a_{24}
a_{51}	a_{12}	a_{23}	a_{34}

Figure. 2 Illustration for deriving the diagonal face images

Without loss of generalization, assume that the width *n* is no smaller than the height *m*. For each training face image A_k, derive the corresponding diagonal face B_k using the method illustrated in Figure.1 Note that B_k *s* is of the same size of A_k *s*. Based on the diagonal faces, define the diagonal covariance matrix as

$$G = \frac{1}{M}\sum_{k=1}^{M}(B_k - \overline{B})(B_k - \overline{B}) \qquad (1)$$

Where $\overline{B} = \frac{1}{M}\sum_{K} B_k$ is the mean diagonal face image, according to eq. (1), the projective vectors X_1.....X_d can be obtained by computing the *d* eigenvectors corresponding to the *d* biggest eigenvalues of G. Since the size of G is only *n* by *n*, computing its eigenvectors can be efficient.

Let $X=[X_1......X_d]$ denote the projective matrix, projecting training faces Ak s onto X, yielding *m* by *d* feature matrices

$$C_k = A_k X \qquad (2)$$

Given a test face image A, first use eq. (2) to get the feature matrix *C=AX*, then a nearest neighbor classifier can be used for classification. Here the distance between C and Ck is defined as

$$d\left(C,C_k\right) = \left\|C - C_k\right\| = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{d}\left(C^{(i,j)} - C^{(i,j)}{}_K\right)^2} \qquad (3)$$

<u>Combining Diagonal PCA and 2DPCA</u>:

Suppose the *n* by *d* matrix *X = [X₁,......X_d]* is the projective matrix of Diagonal PCA.

Let $\overline{A} = \frac{1}{M}\sum_{k} A_k$ denote the mean training face, the projective matrix $Y = [Y_1,......Y_q]$ 2DPCA is computed as follows. When the height *m* is equal to the width *n*, Y is gotten by computing the *q* eigenvectors corresponding to the *q* biggest eigenvalues of the image covariance matrix;

$$\frac{1}{M}\sum_{k=1}^{M}\left(A_k - \overline{A}\right)^T\left(A_k - \overline{A}\right) \qquad (4)$$

On the other hand, when the height *m* is not equal to the width *n*, *Y* is gotten by computing the *q* eigenvectors corresponding to the *q* biggest eigenvalues of the alternative image covariance matrix;

$$\frac{1}{M}\sum_{k=1}^{M}\left(A_k - \overline{A}\right)\left(A_k - \overline{A}\right)^T \qquad (5)$$

Projecting training faces A_k s onto X and *Y* together, yielding the *q* by *d* feature matrices

$$D_k = Y^T A_k X \qquad (6)$$

Given a test face image *A*, first use eq. (6) to get the feature matrix, then a nearest neighbor classifier can be used for classification.

III. SPEECH FEATURE EXTRACTION

Mel Frequency Cepstral Coefficients (MFCC) is chosen because of the sensitivity of the low order cepstral coefficients to overall spectral slope and the sensitivity properties of the high-order cepstral coefficient. Currently it is the most popular feature extraction method. MFCC is produced after the recorded signal is pre-emphasized, framed and hamming windowed. Then the signal is normalized and filtered using lowpass. Lowpass filter is used to remove the potential artificial high frequencies appearing in their modulation spectrum due to transmission errors.

Before identifying or training a command that should be identified by the system, the voice signal must be processed to extract important characteristics of speech. Pitch frequency and formants are most important features of voice signal. Pitch is fundamental frequency of speech signal. The pitch frequency corresponds to the fundamental frequency of vocal cord vibrations. Pitch is a characteristic of excitation source. Formants are resonance frequencies of vocal tract and so they are characteristics of vocal tract.

The vocal tract is a non-uniform acoustic tube. For a uniform tube, the resonance frequencies are obtained as follows:

$$F_i = \frac{C}{4L}\left(2i - 1\right) \qquad for\, i = 1,2,3,.... \qquad (7)$$

Where length of tube, L=17.5 cm (almost equal to an adult human vocal tract length) and C= speed of sound. Therefore we obtain different resonance frequency for this tube (in this case 500Hz, 1500Hz, 2500Hz…).

According to this model, speech signal, *S(n),* is composed of a convolved combination of excitation signal, with the vocal tract impulse response.

We have access only to the output signal, *S(n),* but we need separated e(n) and θ(n) for recognizing the command. Because individual parts are not combined linearity, the cepstral analysis is used to separate *e(n)* and *θ(n).* In order to feature extraction, calculation of cepstral coefficients in Mel frequency scale is required.

A. Cepstral Analysis

Cepstral is a time domain analysis that its main idea is separation of two convolved signals.

The output signal of speech production system *S(n),* is as follows:

$$s(n) = e(n) * \theta\,(n) \qquad (8)$$

Using Fourier transform we have:

$$s(w) = E(w)\theta\,(w) \qquad (9)$$

With taking logarithm, following equation is obtained:

$$\log s(w) = \log E(w) + \log\theta\,(w) \qquad (10)$$

This equation is shown as follows:

$$cs(w) = ce(w) + c\theta\,(w) \qquad (11)$$

Using IDFT, the cepstral coefficients are obtained.

$$cs(n) = ce(n) + c\theta\,(n) \qquad (12)$$

In other word, cepstral coefficients are computed in the form of:

$$cs(n) = f - 1(\log[f(s(n))] \qquad (13)$$

B. Mel-frequency Scaling

Physiological studies have shown that human auditory system does not follow a linear scale. Thus for each tone with an actual frequency, *f*, measured in *Hz*, a subjective pitch is mapped on a scale called the Mel scale. The mel-frequency scale is linear frequency spacing below 1000 Hz and a logarithmic spacing above 1000 Hz. The main advantage of using Mel frequency scaling is that Mel frequency scaling is very approximate to the frequency response of human auditory systems and can be used to capture the phonetically important characteristics of speech. One approach for simulating the subjective spectrum is to use a filter bank, spaced uniformly on the Mel scale as shown in the fig.3. That filter bank has a triangular band pass frequency response, and the spacing as well as the bandwidth is determined by a constant Mel frequency interval.

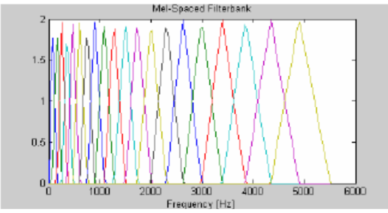


Figure.3 Mel spaced filter bank

The relation between linear frequency and Mel frequency is as follows:

$$Mel(f)=2595* \log 10\,(1+ f / 700) \qquad (14)$$

C. MFCC Computation

In first step, the continuous speech signal is blocked into frames of N samples, with adjacent frames being separated by M (M < N).

Typical values for N and M are N = 256 and M = 100.

The next step in the processing is to window each individual frame so as to minimize the signal discontinuities at the beginning and end of each frame. Typically the Hamming window is used.

The next processing step is the Fast Fourier Transform, which converts each frame of N samples from the time domain into the frequency domain. After that the scale of frequency is converted from linear to Mel scale. Then logarithm is taken from the results. In final step, the log Mel spectrum is converted back to time domain. The result is called the Mel frequency cepstrum coefficients (MFCC). The cepstral representation of the speech cepstrum provides a good representation of the local spectral properties of the signal. Using triangular filter bank, we obtain significant decrease in amount of data. But for more simplicity in next computations, more decreasing in amount of data is needed. For this purpose vector quantization algorithm is used.

D. Vector Quantization

Vector quantization (VQ) is used for command identification in our system. VQ is a process of mapping vectors of a large vector space to a finite number of regions in that space. Each region is called a cluster and is represented by its center (called a centroid). A collection of all the centroids make up a codebook. The amount of data is significantly less, since the number of centroids is at least ten times smaller than the number of vectors in the original sample. This will reduce the amount of computations needed when comparing in later stages.

E. Command Matching

In the recognition phase the features of unknown command are extracted and represented by a sequence of feature vectors *{x_1... x_n}*.

Each feature vector in the sequence *X* is compared with all the stored codewords in codebook, and the codeword with the minimum distance from the feature vectors is selected as proposed command. For each codebook a distance measure is computed, and the command with the lowest distance is chosen.

One way to define the distance measure is to use the Euclidean distances:

$$D = \left(\sum (x_i - y_i)^2\right)^{1/2} \qquad (15)$$

IV. Fusion at the Matching Score Level

In the context of verification, there are two approaches for consolidating the scores obtained from different matchers. One approach is to formulate it as a classification problem, where a feature vector is constructed using the matching scores output by the individual matchers; this feature vector is then classified into one of the two classes: "Genuine user" or "Impostor". In the combination approach, the individual matching scores are combined to generate a single scalar score, which is then used to make the final decision.

Now consider a multimodal biometric verification system that utilizes the combination approach to fusion at the match score

level. The theoretical framework developed by Kittler in [25] can be applied to this system only if the output of each modality is of the form *P (genuine|Z)* i.e., the posteriori probability of user being "genuine" given the input biometric sample *Z*. In practice, most biometric systems output a matching score *s*, and Verlinde et al. [26] have proposed that the matching score *s* is related to *P (genuine|Z)* as follows:

$$S = f\{P(genuine \mid Z)\} + \eta(Z) \qquad (16)$$

where *f* is a monotonic function and η is the error made by the biometric system that depends on the input biometric sample *Z*. This error could be due to the noise introduced by the sensor during the acquisition of the biometric signal and the errors made by the feature extraction and matching processes.

If we assume that is zero, it is reasonable to approximate *P(genuine|Z)* by *P(genuine|s)*. In this case, the problem reduces to computing *P(genuine|s)* and this requires estimating the conditional densities *P(s|genuine)* and *P(s|impostor)*. The probability of the score being that of a genuine user was then computed as,

$$P(genuine \mid S) = \frac{p(s \mid genuine)}{p(s \mid genuine) + p(s \mid imposter)} \qquad (17)$$

V. Experimental Results

Face recognition algorithms cannot handle global facial plastic surgery such as skin resurfacing and full face lift. In most of the test cases, for global surgery, differences between pre and post surgery images of the same individual is very large. In other words, facial feature and texture is drastically altered after surgery and hence the algorithms do not yield good performance. For few test cases of skin resurfacing that have relatively closer resemblance in pre and post surgery images, most of the recognition algorithms are able to perform correct classification.

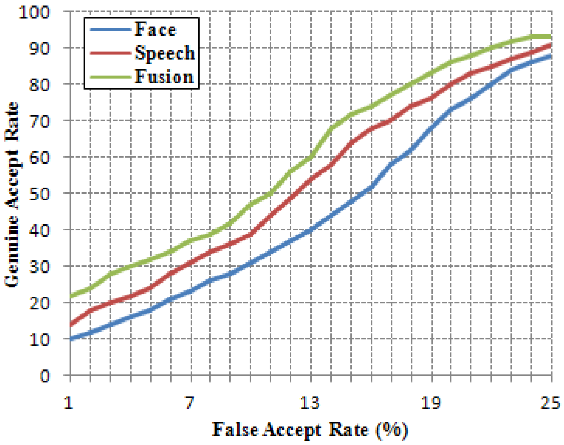


Figure 4. Matching Performance of Accuracy rates.

To evaluate the performance of face recognition algorithms in such an application scenario, the plastic surgery database is partitioned into two groups: training database and testing database. This partition ensures that the verification is performed on unseen images. The train-test partitioning is repeated again and again by computing the false rejection rates (FRR) over these trials at different false accept rate (FAR). The verification accuracy is computed at 6.26% FAR. The experimental result for the recognition rate using the proposed method is summarized in Table 2. In this case, the FAR can accept a person out of 120. Table 2 shows the result of the recognition rate and FAR for the proposed method.

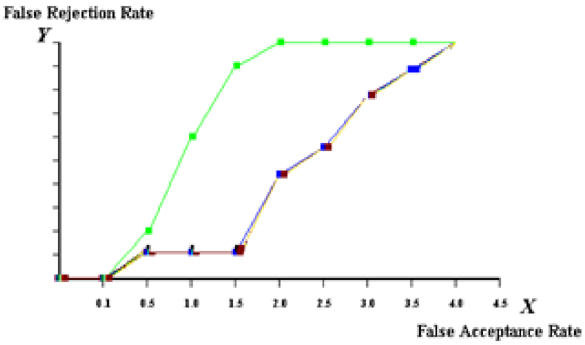


Figure 5. Matching Performance of Genuine acceptance

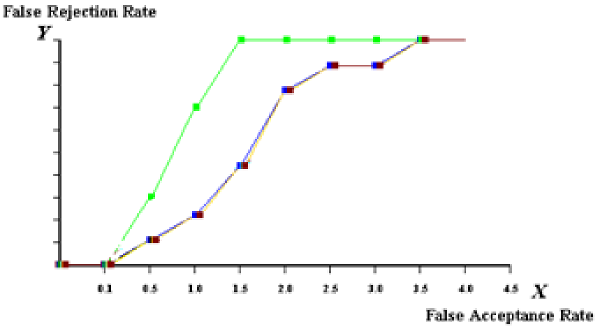


Figure 6. Matching Performance of Imposter Acceptance

Table.1 Verification rates of face and speech

Test Database	Verification Rates (%)	FAR (%)
Face	88.52	11.48
Speech	91.37	8.63

Table 2. Verification rate of the proposed method

Test Database	Verification Rates (%)	FAR (%)
Fusion of Face & Speech	93.74	6.26

VI. Conclusion

In this paper, we present a multimodal biometric human identification method using combined plastic surgery face image and speech information in order to improve the problem of multimodal biometric face recognition system. Current face recognition algorithms mainly focus on handling pose, expression, illumination, aging and disguise. This paper formally introduces plastic surgery as another major challenge for face recognition algorithms using speech signal. Based on the results, we believe that more research is required to design optimal face recognition algorithms that can account for the challenges due to plastic surgery. The procedures can significantly change the facial regions both locally and globally, altering the appearance, facial features and texture. Existing face recognition algorithms generally rely on this information and any variation can affect the multimodal biometric recognition performance.

VII. References

[1] S. Li, R. Chu, S. Liao, and L. Zhang, "Illumination invariant face recognition using near-infrared images", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 4, pp. 627–639, 2007.

[2] R. Singh, M. Vatsa, and A. Noore, “Improving verification accuracy by synthesis of locally enhanced biometric images and deformable model”, Signal Processing, vol. 87, no. 11, pp. 2746–2764, 2007.

[3] V. Blanz, S. Romdhami, and T. Vetter, “Face identification across different poses and illuminations with a 3d morphable model”, in Proceedings of International Conference on Automatic Face and Gesture Recognition, 2002, pp. 202–207.

[4] X. Liu and T. Chen, “Pose-robust face recognition using geometry assisted probabilistic modeling”, in Proceedings of International Conference on Computer Vision and Pattern Recognition, 2005, vol. 1, pp. 502– 509.

[5] R. Singh, M. Vatsa, A. Ross, and A. Noore, “A mosaicing scheme for pose-invariant face recognition”,IEEE Transactions on Systems, Man and Cybernetics - Part B, vol. 37, no. 5, pp. 1212–1225, 2007.

[6] A. Lanitis, C.J. Taylor, and T.F. Cootes, “Toward automatic simulation of aging effects on face images”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 4, pp. 442–450, 2002.

[7] N. Ramanathan and R. Chellappa, “Face verification across age progression”, IEEE Transactions on Image Processing, vol. 15, no. 11, pp. 3349–3362, 2006.

[8] N. Ramanathan, A.R. Chowdhury, and R. Chellappa, “Facial similarity across age, disguise, illumination and pose”, in Proceedings of International Conference on Image Processing, 2004, vol. 3, pp. 1999–2002.

[9] R. Singh, M. Vatsa, and A. Noore, “Face recognition with disguise and single gallery images”, Image and Vision Computing, vol. 27, no. 3, pp. 245–257, 2009.

[10] R. Brunelli and D. Falavigna, “Person Identification Using Multiple Cues”, *IEEE Trans. PAMI*, vol. 17, no. 10, pp. 955-966, 1995.

[11] J. Kittler, M. Hatef, R.P.W. Duin, and J. Matas, “On Off Combining Classifiers”, *IEEE Trans. PAMI*, vol. 20, no. 3, pp. and 226-239, 1998.

[12] L. Hong and A.K. Jain, “Integrating Faces and fingerprint for Personal Identification”, *IEEE Trans. PAMI*, vol. 20, no.12, pp 1295-1307, 1998.

[13] S. Ben-Yacoub, Y. Abdeljaoued, and E. Mayoraz, “Fusion of Face and Speech Data for Person Identity Verification”, *IEEE Trans. Neural Networks*, vol. 10, no. 5, pp. 1065-1075, 1999.

[14] A. Ross and A.K. Jain, “Information Fusion in Biometrics”, *Pattern Recognition Letters*, vol. 24, no. 13, pp. 2115-2125, 20-2003.

[15] J. Kittler, M. Hatef, R. Duin, and J. Matas; "On Combining Classifiers"; *IEEE Trans on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 3, March 1998.

[16] Ben-Yacoub, Abdeljaoued, Mayoraz; “Fusion of Face and Speech Data for Person Identity Verification”; 1999.

[17] J. Fierrez-Aguilar, J. Ortega-Garcia, D. Garcia-Romero and J. Gonzalez-Rodriguez; “A comparative evaluation of fusion strategies for multimodal biometric verification”; *Proc. 4th IAPR Intl. Conf. on Audio- and Video-based Biometric Person Authentication, AVBPA*, Springer LNCS-2688, pp. 830-837; 2003.

[18] J. Yang, D. Zhang, A.F. Frangi, J.Y. Yang, Two-dimensional PCA: a new approach to appearance-based face representation and recognition, IEEE Trans. on Pattern Analysis and Machine Intelligence 26 (1) (2004) 131-137.

[19] D. Q. Zhang, S.C. Chen, J. Liu, Representing image matrices: Eigenimages vs. Eigenvectors, In: Proceedings of the 2nd International Symposium on Neural Networks (ISNN'05), Chongqing, China, LNCS 3497 (2005) 659-664.

[20] Ramasubramanian V., Das A. and Kumar V. P. (2006).“Text-dependent speaker recognition using one-pass dynamic programming algorithm”, in Proc. ICASSP 2006, vol. 1, pp. 901-904.

[21] Hebert, M. (2008), “Text-Dependent Speaker Recognition”, chapter 37 in Benesty, Sondhi and Huang (Eds.) “Handbook of Speech Processing”, Springer.

[22] Matsui T. and Furui S. (1993). “Concatenated phoneme models for text-variable speaker recognition”, in Proc. ICASSP 1993, vol. 2, pp. 391-394.

[23] Che C.-W., Lin Q. and Yuk D.-S. (1996). “An HMM approach to text-prompted speaker verification”, in Proc. ICASSP 1996, vol. 2, pp. 673-676.

[24] Bimbot F., Hutter H. P., et al. (1997). “Speaker verification in the telephone network: research activities in the CAVE project”, in Proc. Eurospeech 1997, pp. 971-974.

[25] J. Kittler, M. Hatef, R.P. Duin, J.G. Matas, On combining classifiers, IEEE Trans. Pattern Anal. Mach. Intell. (1998) 226–239.

[26] P. Verlinde, P. Druyts, G. Cholet, M. Acheroy, Applying Bayes based classifiers for decision fusion in a multi-modal identity verification system, in: Proceedings of International Symposium on Pattern Recognition “In Memoriam Pierre Devijver”, Brussels, Belgium, 1999