

# A Novel Adaptive Two-phase Multimodal Biometric Recognition System

Venkatramaphanikumar Sistla<sup>1</sup>, Venkata Krishna Kishore Kolli<sup>1</sup>, and Kamakshi Prasad Valurouthu<sup>2</sup>

<sup>1</sup>Department of Computer Science and Engineering, Vignan's Foundation for Science, Technology and Research, India

<sup>2</sup>Department of Computer Science and Engineering, Jawaharlal Nehru Technological University Hyderabad College of Engineering, India

**Abstract:** *Multimodal biometric recognition systems are intended to offer authentication without compromising on security, accuracy and these systems also used to address the limitations of unimodal systems like spoofing, intra class variations, noise and non-universality. In this paper, a novel adaptive two-phase multimodal framework is proposed with face, finger and speech traits. In this work, face trait reduces the search space by retrieving few possible nearest enrolled candidates to the probe using Gabor wavelets, semi-supervised kernel discriminant analysis and two dimensional- dynamic time warping. This nonlinear face classification serves as a search space reducer and affects the True Acceptance Rate (TAR). Later, level-1 and level-2 features of fingerprint trait are fused with Dempster Shafer theory and achieved high TAR. In the second phase, to reduce FAR and to validate the user identity, a text dependent speaker verification with RBFNN classifier is proposed. Classification accuracy of the proposed method is evaluated on own and standard datasets and experimental results clearly evident that proposed technique outperforms existing techniques in terms of search time, space and accuracy.*

**Keywords:** *Gabor filters, radial basis function, discrete wavelet transform, dynamic time warping kernel discriminant analysis.*

*Received April 18, 2017; accepted June 13, 2017*

## 1. Introduction

Traditional security applications with username and password do not offer high-level security to preserve the privacy and integrity of the user. To address these challenges, physiological and behavioural characteristics of a person are used to serve as passwords called biometrics [29]. If these biometric systems are operated with uni-trait data or some specific feature set called uni-modal biometrics. Due to problems in data acquisition devices and discriminative abilities of feature set in uni-modal systems' performance is reduced [2, 24, 27]. To overcome all these limitations, multimodal biometric recognition systems are intended. These multimodal biometric systems strengthen the security and provide reliable performance by making use of two or more biometric traits information or by using multiple discriminating features. Performance of these multimodal systems [6] is hampered by noise held at the data and domination of one feature vector. In addition, these fusion methods have severe constraints like time and space complexity in real time usage.

Most of the multimodal biometric recognition systems are developed using parallel and serial fusion techniques. In serial fusion, recognition accuracy is mostly depended upon initial classifier and parallel fusion is limited with time and space complexities. To overcome all these limitations and to preserve the

efficiency, a novel adaptive framework is proposed with face, finger and speech traits. In the rest of the paper, an extensive overview of different fusion scenarios is presented at section 2; proposed methodology is given in section 3. Experimental results and discussions of the proposed method on benchmark databases, is demonstrated in section 4. Lastly, conclusion is presented in section 5.

## 2. Related Work

A multimodal system offers authentication over fraudulent access by fusing two or more traits data. Generally, fusion is done at three levels; information, feature and decision. In information-level fusion, arithmetic operations are applied among pixels of images to enhance the image quality. This fusion is not apt for inadequate quality images.

Al-Osaimi *et al.* [2] have extracted texture features from 3D face images and used pixel-level fusion to address expression and illumination variations of face recognition. Bharadwaj *et al.* [7] defined quality metrics in data acquisition. They proposed a cascaded framework to demonstrate the efficiency with face, iris and fingerprint traits. Conti *et al.* [9] proposed information-level fusion in frequency based approach using iris and fingerprint traits. They evaluated their approach on FVC-2002, and achieved 5.71% False Rejection Rate (FRR) and 0% FAR. Feature-level

fusion is highly effective for low quality images. In this fusion, multiple methods or approaches are fused to generate single feature vector. Shekhar *et al.* [37] demonstrated a multimodal quality metric to weigh each trait and solved optimization problem.

In decision-level fusion, independent decisions are fused to realize the final decision and it is highly effective for multiple choices. Poh *et al.* [33] proposed SVM-NPS approach to fuse the decisions over face and fingerprint modalities. They offered weighted sum rule for the decision fusion on Bio-secure multimodal dataset. Kim *et al.* [23] described a framework with fusion of face, teeth and speech features. They evaluated their approach on a dataset, which consists of 1000 samples acquired from 50 persons using smart phone. Haghighat *et al.* [14] proposed a framework with finger geometry and finger vein and they fused the similarity scores with score-level fusion. Elmir *et al.* [12] described that fusion at feature-level outperforms score-level with voice and fingerprint modalities. Paul *et al.* [31] presented a decision fusion on face, signature and ear traits. They have extracted discriminating features using FLDA and yielded superior results. In the parallel fusion, all the trait information is captured and used at both acquisition and classification phases. This facilitates lots of inconvenience to the user and increases the system's time complexity.

Baig *et al.* [3] proposed a cascaded multimodal framework with the combination of classifiers. Initially, weak classifiers are used and then strong classifiers are applied to take a final decision. They used mahalanobi's distance to compute possible candidates at each stage. Zhang *et al.* [41] proposed a serial fusion based multimodal system using semi supervised learning. Efficiency of that system is mostly depended on the optimistic order of traits and classifiers used. Let, consider a multimodal system [13, 19] with face and fingerprint modalities, face trait itself able to achieve utmost 96-97% of recognition accuracy but to authenticate remaining 3-4% of users, fingerprint trait features are required.

From the extensive survey on literature, it is clearly evident that different fusion scenarios are proposed with various traits. Most of those methods have the limitations like efficiency, enhanced security, user acceptability etc.

### 3. A Novel Adaptive Two-phase Multimodal Biometric Recognition System

Figure 1, depicts the architecture of the proposed system with face, finger and speech traits to authenticate the user identity. In the first phase, weaker trait like face trait is used to find out top k-correlated samples for reduction of search space and to reduce computational time complexity in the matching of traits. So, face recognition reduces the search space for successor trait and minimizes the FRR.

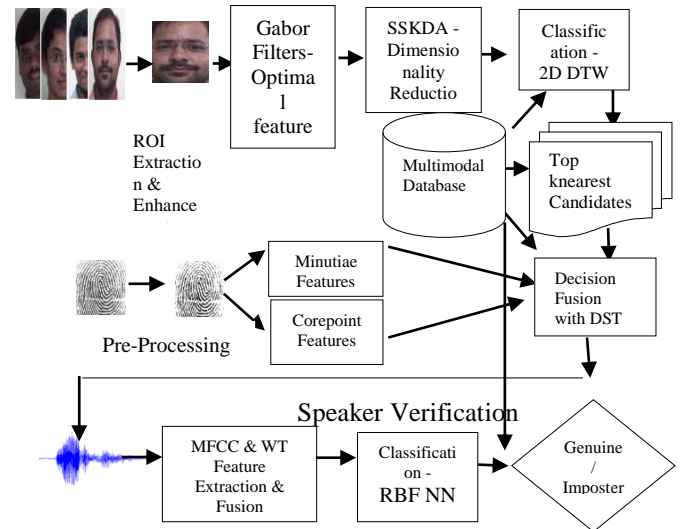


Figure 1. Architecture of proposed adaptive multimodal biometric recognition system.

#### 3.1. Sub-space Reduction with Face Recognition

Face recognition assess the identity of individual by analysing and evaluating discriminating patterns. Firstly, correlative clustering is applied on the face corpus to choose high quality data samples for training purpose. At this step, the dataset is partitioned into clusters using k-means clustering. For example, choose  $k=3$  then, the images {7, 8, 9} are belongs to cluster-1, {5, 6, 12, and 13} belongs to cluster-2 and remaining all 13 samples are in cluster 3 from Figure 2. Let  $C_i$  is set of samples of cluster and  $Cr_i$  is the cluster representative.

$$Cr_i = \frac{1}{n} \sum_{i=1}^n C_i \quad (1)$$

Now, the correlation among the samples of each cluster has been computed using Equation (2) and the sample with highest correlation is being considered as representative of cluster (CR) and all such samples are used for training and remained samples are used to assess the performance of the proposed method.



Figure 2. Sample images of VMB face database.

In this case, samples {4, 8 and 12} are highly correlated compared to rest of them. So, those all above samples are used to build the model.

$$corr(x, y) = \frac{\sum_m \sum_n (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum_m (X - \bar{X})^2} \sqrt{\sum_n (Y - \bar{Y})^2}} \quad (2)$$

### 3.1.1. Pre-Processing

Histogram equalisation enhances the image quality with contrast stretching [36]. Let,  $F$  is a face image of size  $m*n$  and pixel intensities are ranged from 0 to  $g-1$  where  $g$  is the number of gray levels.  $N$  is the enhanced histogram  $E$  and given in Equation (3).

$$E = \frac{P_n}{P} \tag{3}$$

Where  $P$ - number of pixels,  $P_n$ - number of pixels with intensity  $n$ . Histogram equalized image 'HE' of  $E$  is

$$HE_{i,j} = \lfloor ((g - 1) \sum_{n=0}^{M-1} E) \rfloor \tag{4}$$

The floor function rounds off the intensity of the pixel to the next nearest lowest integer value and ' $\gamma$ ' is CDF of  $X$  and is given at Equation (5).

$$\gamma = \left\lfloor \left( (g - 1) \sum_{n=0}^k E \right) \right\rfloor \tag{5}$$

The Probability Density Function (PDF)  $\tau$  of enhanced image is defined as

$$\tau = \gamma(X) = (g - 1) \int_0^x P(x) dx \tag{6}$$

### 3.1.2. Optimal Feature Extraction Using Gabor Filters

Gabor filters [25] operate just like human visual system by preserving optimal resolution in both frequency and spatial domains. Gabor filters are band pass filters and processes texture and discriminative details.

$$\psi_{r,s}(p) = \frac{\|v_{r,s}\|^2}{\sigma^2} e^{\left(\frac{\|v_{r,s}\|^2 \|p\|^2}{2\sigma^2}\right)} (e^{iv_{r,s}p} - e^{-\frac{\sigma^2}{2}}) \|v_{r,s}\| = v_s e^{k\phi_r} \tag{7}$$

In the spatial geometry  $p(i, j)$  is a point,  $\sigma$  is the standard deviation,  $r$  and  $s$  are rotation & scaling kernels respectively. Kernel function is defined by convoluting a face image  $I$  with Gabor filter  $\psi_{ris}$ .

The product of complex plane wave with Gaussian envelope yields a Gaussian kernel. These complex plane waves are represented with both real and imaginary parts.

$$\psi_{f,\theta}(P) = e^{\left(-\frac{1}{2} \left\{ \frac{i_0^2}{\sigma_x^2} + \frac{j_0^2}{\sigma_y^2} \right\}\right)} e^{(2\pi f i_\theta)} \tag{8}$$

Where  $f$  is the wavelength,  $\theta_n$  is the contour angle describes the orientation of the sinusoidal curve and lines specifies the impact of illumination [21].  $\sigma_x$  and  $\sigma_y$  are the deviations of an envelope along the spatial plane that describe bell radius. Power spectrum along real ( $r$ ) and imaginary ( $c$ ) parts is represented in Equation (9).

$$G_{f,\theta} = \sqrt{r^2 + c^2} \tag{9}$$

To address pose variations, an optimistic number of kernels are used in feature generation. To address the curse of dimensionality, a nonlinear feature reduction approach [4] is used to project the features of higher

dimensional space into a lower dimensional space by preserving high within class similarity and low in between class similarity.

### 3.1.3. Feature Reduction with SSKDA

The transformation function  $p: X \rightarrow Y$  [15] transforms the data from higher to lower dimensional space by preserving  $p(X)=Y$ . Let,  $N$  be the number of classes,  $TS$  be the number of samples per each class and  $m$  is the average of samples of a class. Within scatter matrix and between scatter matrices are defined in Equations (10) and (11).

$$S_w = \frac{1}{N} \sum_1^N \frac{1}{TS} \left\| \sum_{i=1}^{TS} \rho * \rho^T \right\| \tag{10}$$

$$S_b = \frac{1}{N} \sum_1^N \left\| (x - m)(x - m)^T \right\| \tag{11}$$

SSKDA yields nontrivial solutions to the eigenvalue problem by computing eigenvalues and eigenvectors and preserves the Equation (12).

$$\lambda S_w x = S_b x \tag{12}$$

Orthonormal vectors  $\lambda$  of (12) are independent and diagonalizes the transformation. Linear sum of the eigenvalues is computed to reduce non-significant eigenvectors using Equation (13).

$$W = \sum_1^N \sum_1^{TS} \gamma \rho(X) \tag{13}$$

### 3.1.4. Top k-candidates Identification Using Two Dimensional Dynamic Time Warping

Dynamic Time Warping (DTW) [35, 40] matches the time series like data patterns and it is used to find an optimal warp by initiating the comparison from (1,1) i.e., top left corner point to bottom right corner point (p, q) in the distance matrix by following Equation (14).

$$w(p, q) = \operatorname{argmin}_{(p,q) \in ((p-1,q-1),(p,q-1),(p-1,q))} w^{(p,q)} \tag{14}$$

An optimal warp is defined using Equation (15) and which preserves boundary, contiguous and monotonicity constraints.

$$w' [p, q] = d(p, q) + w(p, q) \tag{15}$$

Likewise, 2D-DTW is used to find top k-nearest samples to the probe face sample. With all those possible identities, fingerprint template database is generated to find out the authorized user. This reflects that face trait is served as search space reducer for fingerprint trait.

## 3.2. Validation of User Identity with Fingerprint Identification

Every human can have distinct fingerprint pattern due to variety of ridges and valleys. Termination points are the points held at the end of the ridge and a point where a ridge has divided into two parts is called bifurcation point [20, 26]. Combination of both

bifurcation and termination points is called minutiae points and level-1 features [39]. Firstly, ridge patterns are enhanced with Gabor filters & fuzzy adaptive histogram equalization method then local adaptive binarization method [13,17] transforms the gray to binary where furrows and ridges are represented with 1 and 0 respectively. Thinning of ridges erodes the width until they become one pixel wide.

### 3.2.1. Level 1 and Level 2 Feature Extraction

Cross number approach is applied to find out the minutiae from thinned fingerprint. A cross number of center pixel in a 3\*3 neighborhood is defined in Equation (16);

$$CN(p) = \frac{\sum_{k=1}^8 |p_{k \bmod 8} - p_k|}{2} \quad (16)$$

$$CN(p) = \begin{cases} 1 & \text{Termination} \\ 3 & \text{Bifurcation} \end{cases}$$

Morphological operators enhance the fingerprint quality by avoiding cavities and background noise that are created during the pre-processing phase. Generally, a quality fingerprint may consists of 50-60 minutiae points, but a low quality fingerprint may have more minutiae but which reduces the recognition accuracy and those are referred as spurious minutiae [5].

Removal of such spurious minutiae is done by using inter ridge distance. With low quality fingerprints, minutiae features alone does not yield high accuracy. So, corepoints i.e., level-2 features are used to authenticate the user identity in the proposed work. Ridge orientation is computed by moving the 3\*3 window through the pixels in 16 similar directions; projections were recorded through y axes. Orientation of the pixel is having maximum variance along projection and it is referred as the corepoint [32, 34]. The point matching method computes the matching score of probe sample with all gallery images. Matching scores of both level-1 and level-2 features are integrated with Dempster Shafer theory.

### 3.2.2. Score-Level Fusion with Dempster Shafer Theory

Dempster Shafer [11, 30] proposed statistical evidence based reasoning method to assign vague values to disjoint subsets of hypothesis referred as theory of belief functions. Consider the similarity scores of level-1 and level-2 features as  $S_M$  and  $S_C$  and those are normalized into the range of [0, 1].

$$S^*_i = \frac{S_i}{\sum_{i=1}^n S_i} \quad (17)$$

Some basic belief value is assigned to each individual intension called mass. Let,  $m$  is a mass function defined as  $m: 2^{S^*} \rightarrow [0, 1]$ , where  $S^*$  is the power set.

$$m(\emptyset) = 0 \text{ and } \sum_{s \in S^*} m(s) = 1 \quad (18)$$

Mass function assigns a value, to every subset  $s$  of  $S^*$  in the range of [0, 1] is called as degree of belief and if it

is nonnegative then it is referred as a focal element and belief function is expressed in Equation (19).

$$bel(s) = \sum_{s \subseteq S^*} m(s) \quad (19)$$

Mass functions  $m_1$  and  $m_2$  are applied to the focal elements  $x$  and  $y$  respectively and then  $m$  is the joined function of mass with focal element  $z$ .

$$m(z) = m_1(x) \Phi m_2(y) = \frac{\sum_{A \cap B = C} m_1(x)m_2(y)}{1 - \sum_{x \cap y = \emptyset} m_1(x)m_2(y)} \quad (20)$$

Denominator of Equation (20) is the normalizing factor and which defines the conflict of masses  $x$  and  $y$ . The pignistic and plausibility functions are defined in Equations (21) and (22) to obtain a decision.

$$pgn(p) = \sum_{q \subseteq S, q \neq \emptyset} m(q) \frac{|p \cap q|}{|q|} \quad (21)$$

The plausibility of  $p$  defines up to which extent failed on doubt about  $q$ .

$$pls(p) = \sum_{p \cap q \neq \emptyset} m(q) \quad (22)$$

Let, probe fingerprint sample  $X$  has the highest matching score with two of the candidates  $\{Y, Z\}$  as 0.6 and 0.7 using minutia and corepoint features. The frame of discernment is  $\{X, Y, Z\}$ . To address the vagueness held among the user identity validation, DST is used.

Masses	(X, Y)	(X, Y, Z)
m1	0.6	0.4
m2	0.7	0.3

Intermediary evidences are represented as

	(X)	(X, Y)	(X, Z)	(X, Y, Z)
m3	0.42	0.18	0.28	0.12

The belief, pignistic and plausibility values on focal elements over  $\{X, Y$  and  $Z\}$  are  $\{0.42, 0, 0\}$ ,  $\{0.69, 0.144, 0.13\}$  and  $\{1, 0.4, 0.3\}$  respectively. These values illustrate that the probe sample is a valid identity. In this work, initially face trait reduced the search space, and then fingerprint yields better accuracy over such reduced subspace. To reduce the False Acceptance Rate (FAR), in the second phase user identity is verified with text dependent speaker recognition.

## 3.3. Ensembled Text-dependent Speaker Verification with RBF Neural Network

In text-dependent speaker verification [8, 16], user is highly supportive and required to be recognised even under variant environmental noisy conditions and user emotions, which affects the accuracy.

### 3.3.1. MFCC Feature Extraction

For each sound frequency, a specific subjective pitch is defined on a mel-scale. Mel-Frequency Cepstral Coefficient (MFCC) [1, 10] is a trustworthy analytical technique acts just like a human ear in speaker recognition. Computation of MFCC features is more

affected with the size of frame. Procedure of MFCC features computation is as follows:

1. Denoise the speech signal with high pass filter, and attenuate the noise part by boosting high frequency signal details.
2. To enhance the quality of the speech utterance and to compute the magnitude, FFT is applied.

$$f_n = \sum_{k=0}^{N-1} f_k e^{-\frac{2\pi j n k}{N}} \quad (23)$$

Where  $n = 0, 1, 2, \dots, N-1$

3. Calculate log of mel frequencies by correlating speech signal to the nonlinear mel-scale by using power spectrum.

$$f_{mel} = 2595 \log_{10} \left( 1 + \frac{f_n}{700} \right) \quad (24)$$

4. Local spectral properties are hold in cepstral features. Log mel-frequencies are de-correlated and well compressed. DCT is applied to transform into time spectrum called MFCC features by using Equation (25).

$$mc_n = \sum_{i=1}^n \log(c_i) * \cos\left(\left(k - \frac{1}{2}\right)\pi\right) \quad (25)$$

$c_i$  is the mel-cepstral features,  $m$  is the cepstral coefficients and  $n$  is the filter banks. In the power spectrum, derivatives of  $c_i$  are defined as delta features and derivatives of delta features are referred as second order derivatives and those are also called delta-delta features.

$$dd_t = \frac{\sum_{i=1}^n i(mc_{n+i} - mc_{n-i})}{2 \sum_{i=1}^n i^2} \quad (26)$$

Performance is increased by combining these delta and delta-delta features to the MFCC features.

### 3.3.2. Feature Extraction with Discrete Wavelet Transform

Wavelets can analyse distinct parts of a signal at various scales and their characteristics are explored to handle user emotions and environmental noise conditions. Wavelets are more useful to find discontinuities, short-time phenomena, and abrupt variations held in the utterances. The basis of wavelet is shown in Equation (27).

$$w(a, b) = \int f(t) \cdot \varphi_{a,b}^*(t) dt \quad (27)$$

Based on all these characteristics MFCC features are fused with wavelet features. To achieve better recognition accuracy, feature vectors of MFCC and DWT are fused and then classified with RBFNN.

### 3.3.3. Verification with Radial Basis Function Neural Network

The first layer of RBF maps the input data to hidden layer ( $X_i \rightarrow X_j$ ). The hidden layer maps the activations to  $k$  output nodes ( $X_j \rightarrow X_k$ ). Gaussian [28] is the activation

function used to update the weights. The activation function is defined as

$$\varphi_i(x) = \exp\left(-\frac{1}{2}(x - m_i)^T C^{-1}(x - m_i)\right) \quad (28)$$

Where  $\varphi_i(x)$  the response of  $i$ th hidden layer node and  $x$  is the  $i$  dimensional input vector and  $m$  is the  $i$ <sup>th</sup> dimensional hidden layer node weight and  $C$  is the Covariance matrix. The weights of the nodes in output layer are demonstrated as linear combination from the RBF activations to the output nodes.

$$y_j = w_{0j} + \sum_{i=1}^k w_{ij} \varphi_i(x) \quad (29)$$

Where  $y_j$  is the  $j$ <sup>th</sup> output node,  $w_{ij}$  is the weight from the  $i$ <sup>th</sup> centre to the  $j$ <sup>th</sup> output node and  $w_{0j}$  is the bias. Architectural steps of RBF are:

1. Initialize input node weights.
2. Compute weights of the hidden units
3. Find the weights of output layer units.
4. Compare the target class with output units.
5. Compute errors at output nodes.
6. Gradient descent method is used to train the output nodes.
7. Evaluate error for each hidden unit.
8. Gradient descent is used to train hidden units.

The learning is continued with the parameters set with learning rate=0.001; epochs=6000; goal=1.0e-5; Least mean square method is used to compute output weights. Inherent scale invariance of RBF classifier guarantees an optimized linear solution in the classification with high learning rates.

## 4. Experimental Results and Discussions

In this work, experiments are carried on standard and on own developed Vignan's Multimodal Biometric Database (VMB) in four phases. Firstly, performance of the proposed non-linear face recognition is presented. In the second phase, multi-feature based fingerprint recognition with DST fusion is given. Further, performance of text dependent speaker recognition with RBF neural network is defined in phase-3. Finally, performance of the proposed novel adaptive multimodal biometric system is compared with the other state of art fusion methods in phase-4.

### 4.1. Performance Evaluation of Nonlinear Face Recognition with 2D-DTW

In Experiment 1: TAR of the proposed nonlinear face recognition on ORL database is 98.4% with normalisation. TAR of the proposed method is 99.1%, 99.7% and 96.34% on Yale, Grimace and VMB face databases respectively.

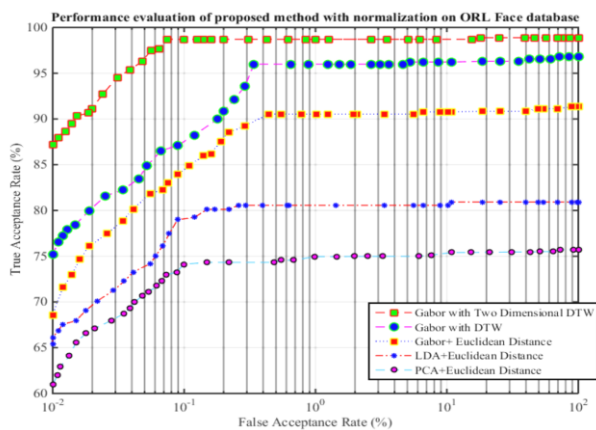


Figure 3. ROC Curve of proposed method with normalization on ORL face database.

ROC of the proposed method on ORL [6] is shown in Figure 3. In the second Experiment, TAR of the proposed method with pose variations is evaluated by adding  $2^0$  &  $5^0$  of rotational noise to the standard databases. TAR of proposed method outperformed in comparison with existing methods on these synthesized databases. Proposed method has yielded 96.3%, 98.3%, 95.1% and 93.7% on ORL, Grimace, Yale and VMB databases respectively. In the Experiment 3, performance of the proposed approach on Grimace database with variant number of kernels and shown in Table 1.

Table 1. TAR of proposed method with different no. of kernels.

Method	40 (8*5)	35 (7*5)	32 (8*4)	24 (8*3)	24 (6*4)	20 (5*4)
Gabor + ED	98.5	98.1	98.2	97.9	97.6	92.7
Gabor + DTW	99.7	99.7	99.3	99.1	98.5	96.2
Gabor+2D-DTW	99.7	99.3	99.3	99.1	98.8	97.9

Equal Error Rate (EER) and Average Total Error Rate (ATER) of the proposed face recognition approach is presented in Figure 4.

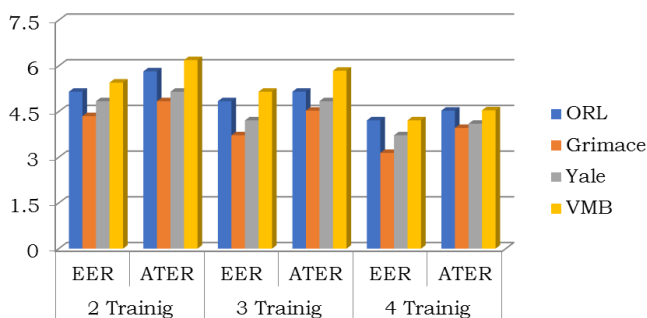


Figure 4. TAR of proposed method on benchmark databases.

Finally, performance of the face trait is evaluated on benchmark databases by making use of 3 training samples and computed top 5-nearest candidates to the probe sample shown in Figure 5. In the retrieval of top 5-candidates to the probe sample, genuine identity is found at 100% on all benchmark databases and it is presented in Figure 5. In the second phase, accuracy of the proposed fingerprint recognition method is

evaluated on standard fingerprint datasets like FVC-2000, FVC-2002 and VMB.

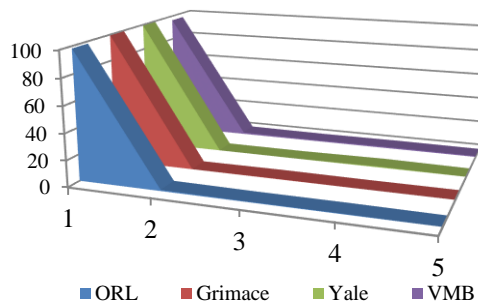


Figure 5. Recognition of top 5-candidates over bench mark databases.

### 4.2. Performance Evaluation of Multi-feature Based Fingerprint Recognition with Score Fusion

Firstly, TAR of minutiae and corepoint approaches on FVC-2000 [7] with enhancement in frequency, spatial and in both the domains are presented in Figure 6. Minutiae approach has achieved 93.27% of recognition accuracy and 92.4% with corepoint approach on FVC-2002. In the second experiment: Similarity scores of both level-1 and level-2 approaches are fused with DST and their performance is compared with weighted score-level fusion. The proposed multi-feature DST fusion approach has yielded 96.77% of recognition accuracy on FVC 2000 and it is 1.45% more than weighted score-level fusion approach. Likewise, performance evaluation is carried on FVC 2002, the proposed approach has given 95.6% of recognition accuracy and it is 1.2% more than weighted score-level fusion.

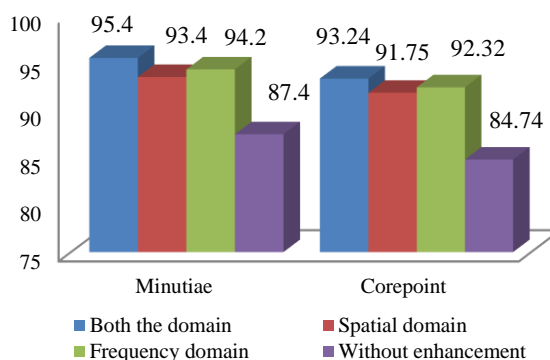


Figure 6. Performance evaluation of Minutiae and Core point approaches with enhancement.

In experiment 3: Robustness of the proposed approach is computed with distorted fingerprints. Gaussian white noise and  $+2^0$  rotational noise is added to FVC 2000 and generated synthesized fingerprint database. TAR of proposed approach is 96.2% and it is 1.7% more than minutiae approach and 2.9% more than corepoint approach. Performance of the proposed method on FVC 2000 is presented in Figure 7. On

VMB fingerprint dataset, proposed multi-feature approach has achieved TAR of 98.1%, which is 2.4% more than level-1 feature approach, and 3.1% more than level-2 approach.

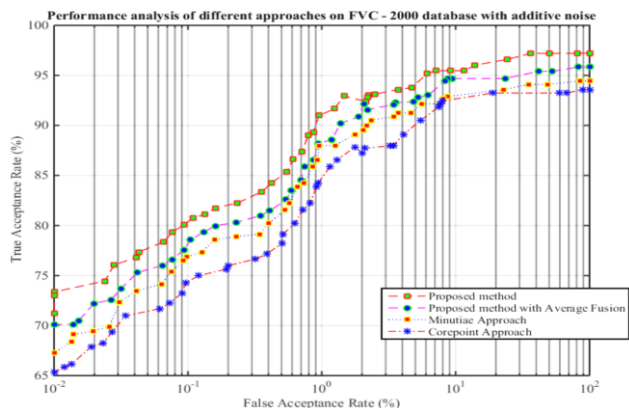


Figure 7. ROC curve of various methods with additive noise on FVC 2000.

ROC curve illustrating the performance evaluation of proposed method with other approaches on VMB fingerprint database with additive noise is depicted in Figure 8. When the experiments are carried with the proposed approach on top-5 candidates identified by the face trait, then the fingerprint has yielded 100% of recognition accuracy on all benchmark databases.

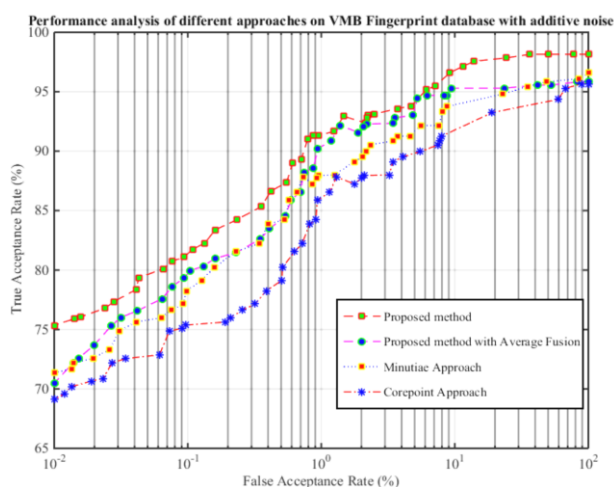


Figure 8. ROC curves of different methods with additive noise on VMB.

### 4.3. Performance Evaluation of Text Dependent Speaker Verification with RBFNN

In the third phase, as a first experiment; the proposed speaker recognition approach MFCC+DWT+RBF is compared with DTW and GMM classifiers. Experimentation is carried on VMB database by making use of two and three utterances for training. TAR of MFCC+DWT +DTW is 96.8% on original corpus, it is 0.9% more than babble noisy corpus and 3.5% more than white noise added speech corpus. DET curve illustrating the FAR & FRR of MFCC+DWT+DTW with two gallery samples is given in Figure 9. With three training, TAR of DTW is 98.9% on original

speech corpus and it is reduced to 98.1% and 97.2% with addition of babble and white noise respectively.

Likewise, performance of MFCC+DWT+ GMM with two training samples is 98.1% and it is 0.5% and 2.4% more than speech corpus with babble and white noise respectively. When the experimentation is carried with three training there is a slight 0.3% growth is observed.

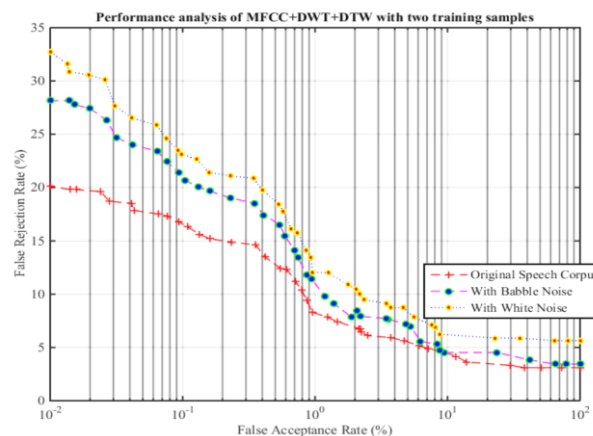


Figure 9. DET curve of MFCC +DWT +DTW with two training samples.

TAR is reduced to 97.9% and 96.5% with the addition of babble and white noises respectively. Similarly, performance evaluation is carried with RBFNN by using two and three training samples. With two training, TAR is 98.4% and it is 0.7% more than speech with babble noise and 1.8% more with white noise. DET curve to show performance evaluation of MFCC+DWT+RBF is shown in Figure 10.

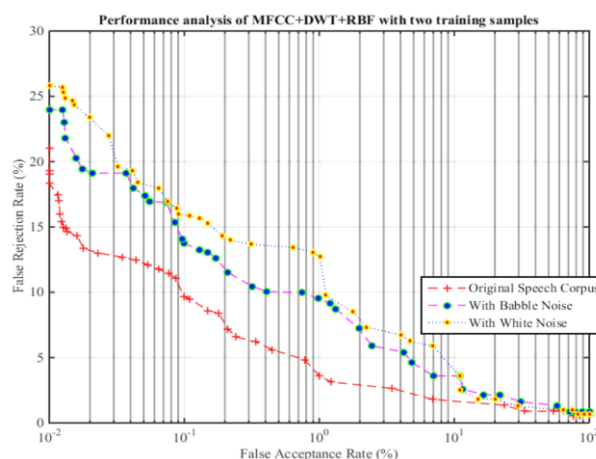


Figure 10. DET curve presenting performance of MFCC +DWT +RBF with two training.

RBF has achieved 100% of classification accuracy with three training samples and it is reduced by 0.9% and 1.2% with the addition of babble and white noise. In the verification phase, the proposed method has yielded 100% of TAR even with all types of noise like babble and white. These results are clearly evidencing that MFCC+DWT+RBF is robust.

### 4.4. Experiments and Result Analysis of Proposed Novel Framework for Adaptive Multimodal Biometric System

Finally, performance of the proposed multimodal approach is compared with state of art feature, decision and score-level multimodal fusion techniques [38]. At feature-level fusion, features that are extracted from face, finger and speech traits are concatenated and further reduced with SSKDA. TAR of face and fingerprint with feature-level fusion is 99.2% and it is 0.7%, 1.9% more than feature level fusion of fingerprint + speech and face+speech respectively. TAR and Equal error rate of this approach on VMB database is 0.7% and 99.2% respectively. Performance evaluation of feature-level fusion is illustrated in Figure 11. Further, decision-level fusion experiments are carried. TAR of face with fingerprint trait is 98.87%, and which is 0.23% lower than feature-level fusion. TAR of fingerprint + speech with Decision-level fusion is 97.41% and it is 0.51% more than face + speech trait. TAR of decision-level fusion of all three traits is 98.9% and which is 0.25% inferior to the feature-level fusion. EER of decision-level fusion is 0.62%. Performance of decision-level fusion is given in Figure 12. Likewise, further experimentation is done with score-level fusion. TAR of face and finger is 97.8%, finger and speech is 98.1%, face and speech is 97.1%. TAR of all three traits with score-level fusion is 98.1% and it is reduced by 1% and 0.2% with feature-level and decision-level fusion techniques respectively and EER is 0.84%. Performance evaluation of face+ fingerprint+ speech traits with score-level fusion is shown in Figure 13. Recognition rate of the proposed method is better than the proposed method in [22].

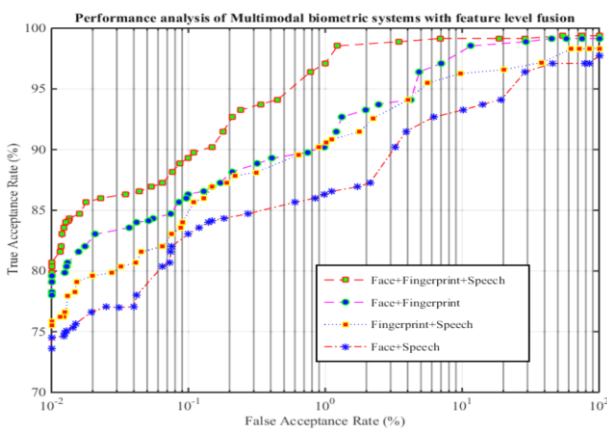


Figure 11. ROC curve presenting TAR of the Multimodal system with feature-level fusion.

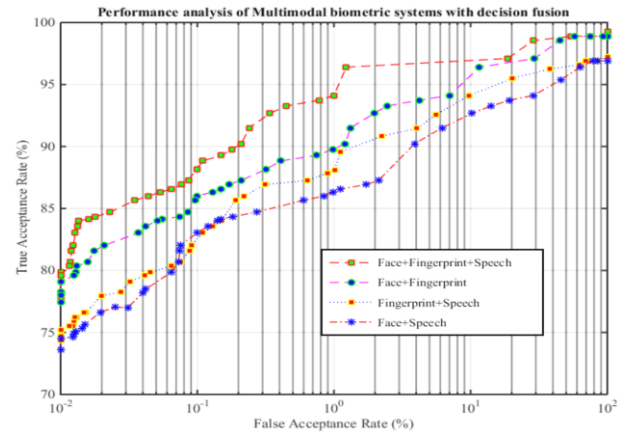


Figure 12. ROC curve presenting the TAR of Multimodal system with decision-level fusion.

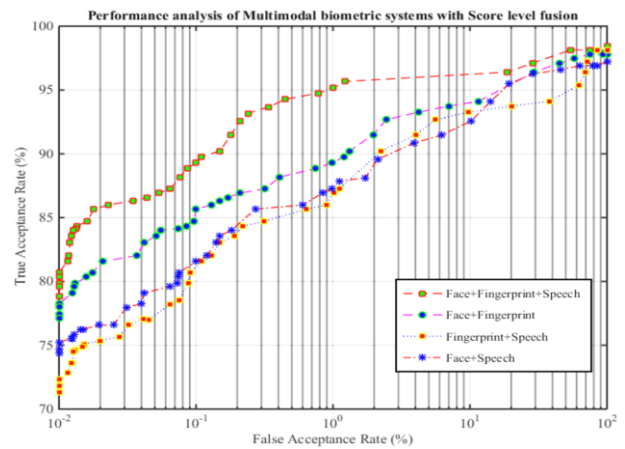


Figure 13. ROC curve defining the TAR of Multimodal system with score-level fusion.

Finally, the proposed novel adaptive multimodal biometric recognition is evaluated on VMB and achieved 99.89% of TAR and performance is illustrated in Figure 14. EER of proposed approach is compared with other fusion methods and given in Figure 15. The time taken for user identity validation is 1.2 sec, and it is 0.6 sec, 0.34 sec less compared with parallel and decision-level fusion approaches. Experimental results illustrate that the proposed method achieved high TAR and low FAR.

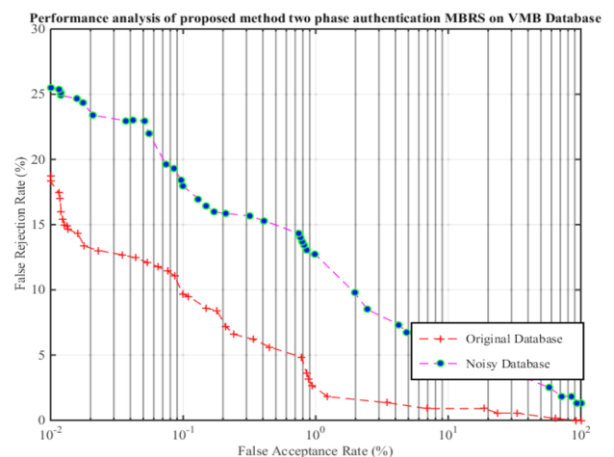


Figure 14. DET curve of proposed novel adaptive multimodal biometric system.



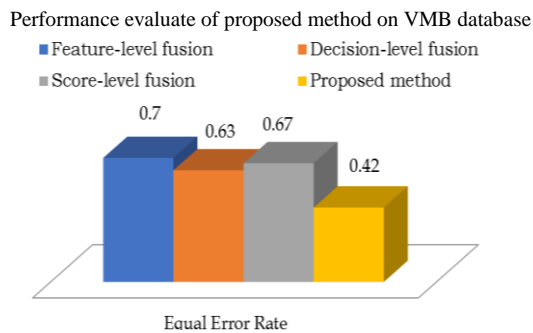


Figure 15. EER of various multimodal biometric security systems.

## 5. Conclusions

In this work, challenges of present multimodal biometric systems such as search time, accuracy, user convenience, and space have addressed. Based on the characteristics and uniqueness, face, finger and voice traits are considered for the proposed novel adaptive two-phase multimodal system to overcome the limitation like spoofing, intra class variations and noise. In this work, initially, face trait used as a search space reducer for fingerprint trait and then fingerprint realizes 100% of recognition accuracy over a closed set of identities. In the fingerprint recognition, Dempster Shafer theory is used to fuse the both level-1 & level-2 features. Finally, to minimize the FAR, text dependent speaker recognition with RBFNN to verify the user identity. Extensive experiments are carried on benchmark databases and own VMB database, the proposed framework outperforms existing state of fusion methods in all the respects of TAR, FAR and search space and time.

## References

- [1] Almaadeed N., Aggoun A., and Amira A., "Speaker Identification Using Multimodal Neural Networks and Wavelet Analysis," *IET Biometrics*, vol. 4, no. 1, pp. 18-28, 2014.
- [2] Al-Osaimi F., Bennamoun M., and Mian A., "Spatially Optimized Data-Level Fusion of Texture and Shape for Face Recognition," *IEEE Transactions on Image Processing*, vol. 21, no. 2, pp. 859-872, 2012.
- [3] Baig A., Bouridane A., Kurugollu F., and Albeshir B., "Cascaded Multimodal Biometric Recognition Framework," *The Institution of Engineering and Technology Biometrics*, vol. 3, no. 1, pp. 16-28, 2012.
- [4] Baudat G. and Anouar F., "Generalized Discriminant Analysis Using a Kernel Approach," *Neural Computation*, vol. 12, no. 10, pp. 2385-2404, 2000.
- [5] Bazen A. and Gerez S., "Systematic Methods for the Computation of the Directional Fields and Singular Points of Fingerprints," *IEEE Transactions on PAMI*, vol. 24, no. 7, pp. 905-919, 2002.
- [6] Belhumeur P., Hespanha J., and Kriegman D., "Eigenfaces Vs. Fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711-720, 1997.
- [7] Bharadwaj S., Vats M., and Singh R., "Biometric Quality: A Review of Fingerprint, Iris, and Face," *EURASIP Journal on Image and Video Processing*, vol. 34, pp. 1-28, 2014.
- [8] Campbell J., Shen W., Campbell W., Schwartz R., Bonastre J., and Matrouf D., "Forensic Speaker Recognition," *IEEE Signal Processing Magazine*, vol. 26, no. 2, pp. 95-103, 2009.
- [9] Conti V., Militello C., Sorbello F., and Vitabile S., "A Frequency-Based Approach for Features Fusion in Fingerprint and Iris Multimodal Biometric Identification Systems," *IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews*, vol. 40, no. 4, pp. 384-395, 2010.
- [10] Cutajar M., Gatt E., Grech I., Casha O., and Micallef J., "Comparative Study of Automatic Speech Recognition Techniques," *IET Signal Processing*, vol. 7, no. 1, pp. 25-46, 2013.
- [11] Denoeux T., "A K-Nearest Neighbor Classification Rule Based on Dempster-Shafer Theory," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 25, no. 5, pp. 804-813, 1995.
- [12] Elmir Y., Elberrichi Z., and Adjoudj R., "Score-Level Fusion Based Multimodal Biometric Identification (Fingerprint and Voice)," in *Proceedings of 6<sup>th</sup> International Conference on Sciences of Electronics, Technologies of Information and Telecommunications*, Sousse, pp. 146-150, 2012.
- [13] Fridman A., Stoleran A., Acharya S., Brennan P., Juola P., Greenstadt R., and Kam M., "Decision Fusion for Multimodal Active Authentication," *IT Professional*, vol. 15, no. 4, pp. 29-33, 2013.
- [14] Haghghat M., Abdel-Mottaleb M., and Alhalabi W., "Discriminant Correlation Analysis: Real-Time Feature Level Fusion for Multimodal Biometric Recognition," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 9, pp. 1984-1996, 2016.
- [15] Haghghat M., Zonouz S., and Abdel-Mottaleb M., "Cloudid: Trustworthy Cloud-Based And Cross-Enterprise Biometric Identification," *Expert Systems with Applications*, vol. 42, no. 21, pp. 7905-7916, 2015.
- [16] Hébert M., *Text-Dependent Speaker Recognition*, Springer Handbook of Speech Processing, 2008.

- [17] Hong L., Wan Y., and Jain A., "Fingerprint Image Enhancement: Algorithms and Performance Evaluation," *IEEE Transactions on PAMI*, vol. 20, no. 8, pp.777-789, 1998.
- [18] Jain A., Hong L., and Bolle R., "On-Line Fingerprint Verification," *IEEE Transactions on PAMI*, vol. 19, no. 4, pp. 302-314, 1997.
- [19] Jain A., Hong L., and Pankanti S., "Biometric Identification," *Communications of the ACM*, vol. 43, no. 3, pp. 90-98, 2000.
- [20] Jain A., Prabhakar S., Hong L., and Pankanti S., "Filter Bank-Based Fingerprint Matching," *IEEE Transactions on Image Processing*, vol. 9, no. 5, pp. 846-859, 2000.
- [21] Kamarainen J., Kyrki V., and Kalviainen H., "Invariance Properties of Gabor Filter-Based Features-Overview and Applications," *IEEE Transactions on Image Processing*, vol. 15, no. 5, pp. 1088-1099, 2006.
- [22] Kang B. and Park K., "Multimodal Biometric Method Based on Vein and Geometry of a Single Finger," *Institution of Engineering and Technology Computer Vision*, vol. 4, no. 3, pp. 209-217, 2010.
- [23] Kim D., Chung K., and Hong K., "Person Authentication Using Face, Teeth and Voice Modalities For Mobile Device Security," *IEEE Transactions on Consumer Electronics*, vol. 56, no. 4, pp. 2678-2685, 2010.
- [24] Kumar A., Kanhangad V., and Zhang D., "A New Framework for Adaptive Multimodal Biometrics Management," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 1, pp. 92-102, 2010.
- [25] Liu C. and Wechsler H., "Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model for Face Recognition," *IEEE Transactions on Image Processing*, vol. 11, no. 4, pp. 467-476, 2002.
- [26] Maio D. and Maltoni D., "Direct Gray-Scale Minutiae Detection in Fingerprints," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 1, pp. 27-40, 1997.
- [27] Maltoni D., Maio D., Jain A., and Prabhakar S., *Handbook of Fingerprint Recognition*, Springer Science and Business Media, 2009.
- [28] Manchala S., Prasad V., and Janaki V., "GMM Based Language Identification System Using Robust Features," *International Journal of Speech Technology*, vol. 17, no. 2, pp. 99-105, 2014.
- [29] Nagar A., Nandakumar K., and Jain A., "Multi biometric Cryptosystems Based on feature-Level Fusion," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 1, pp. 255-268, 2012.
- [30] Nguyen K., Denman S., Sridharan S., and Fookes C., "Score-Level Multibiometric Fusion Based on Dempster-Shafer Theory Incorporating Uncertainty Factors," *IEEE Transactions on Human-Machine Systems*, vol. 45, no. 1, pp. 132-140, 2015.
- [31] Paul P., Gavrilova M., and Alhaji R., "Decision Fusion for Multimodal Biometrics using Social Network Analysis," *IEEE Transactions on Systems, Man and Cybernetics: Systems*, vol. 44, no. 11, pp. 1522-1533, 2014.
- [32] Perumal E. and Ramachandran S., "A Multimodal Biometric System Based on Palmprint and Finger Knuckle Print Recognition Methods," *The International Arab Journal of Information Technology*, vol. 12, no. 2, pp. 118-128, 2015.
- [33] Poh N., Kittler J., and Bourlai T., "Quality-Based Score Normalization with Device Qualitative Information for Multimodal Biometric Fusion," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 40, no. 3, pp. 539-554, 2010.
- [34] Ross A. and Jain A., "Information Fusion in Biometrics," *Pattern Recognition Letters*, vol. 24, no. 13, pp. 2115-2125, 2003.
- [35] Sakoe H., "Two-Level DP-Matching--A Dynamic Programming-Based Pattern Matching Algorithm for Connected Word Recognition," *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. 27, no. 6, pp. 588-595, 1979.
- [36] Samaria F. and Harter A., "Parameterization of A Stochastic Model for Human Face Identification," in *Proceedings of IEEE Workshop on Applications of Computer Vision*, Sarasota, pp. 138-142, 1994.
- [37] Shekhar S., Patel V., Nasrabadi N., and Chellappa R., "Joint Sparse Representation for Robust Multimodal Biometrics Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 1, pp. 113-126, 2014.
- [38] Sultana M., Paul P., and Gavrilova M., "Social Behavioral Information Fusion in Multimodal Biometrics," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 12, pp. 2176-2187, 2017.
- [39] Venkatramaphanikumar S. and Prasad V., "Multi Feature Based Fingerprint Recognition with Score Fusion," *International Journal of Applied Engineering Research*, vol. 10, no. 8, pp. 20393-20402, 2015.
- [40] Venkatramaphanikumar S. and Prasad V., "Gabor Based Face Recognition with Dynamic Time Warping," in *Proceedings of 6<sup>th</sup> International Conference on Contemporary Computing*, Noida, pp. 349-353, 2013.
- [41] Zhang Q., Yin Y., Zhan D., and Peng J., "A Novel Serial Multimodal Biometrics Framework

Based on Semi Supervised Learning Techniques,”  
*IEEE Transactions on Information Forensics and Security*, vol. 9, no. 10, pp. 1681-1694, 2014.



**Venkatramaphanikumar Sistla** is currently working as an Assistant Professor in the department of Computer Science and Engineering at VFSTR University. He is working towards Ph.D. degree at department of Computer Science and Engineering, JNTU Hyderabad, Hyderabad. His research interests include Biometrics, Image Processing, Data Mining and Pattern recognition. He has published more than 15 research articles in reputed Conferences and Journals. He served as an Organizing Chair for one IEEE Conference and two National Conferences.



**Venkata Krishna Kishore Kolli** received his Ph.D degree from Acharya Nagarjuna University, M.Tech in Computer Science and Engineering from Andhra University, India, and B.E from University of Madras, India. He is currently working as Professor and Dean, IT Services, VFSTR University. He got more than 25 years of teaching and research experience in the field of Computer Science and Engineering. His current research interests include Digital image processing, Biometrics, Neural Networks, Machine learning, Pattern Recognition and Data Science. He has published more than 30 journals and conference papers in these areas. He has received best teacher award for four times. He has chaired one IEEE Conference and three National conferences.



**Kamakshi Prasad Valurouthu** received his Ph.D. from the Indian Institute of Technology (IIT) Madras in 2002, M. Tech from the Andhra University, Vizag in 1992. He is currently working as professor & Head of Department of Computer Science and Engineering, JNTU Hyderabad College of Engineering, India. His research interests include Image Processing, Speech Processing, Data Mining and Ad Hoc Networks etc. He has guided 12 Ph.D., 2 M.S and currently guiding 11 Ph.D. scholars. He is an author of three books and published more than 95 publications in reputed Journals & Conferences. He has organized several refresher courses, workshops and conferences.