

12-1-2021

Feature Level Fusion of Iris and Voice.

Sally Sameh

Electronics and Communications Engineering Department , Faculty of Engineering, Mansoura University, Mansoura, Egypt, sally__sameh@hotmail.com

Hossam El-Din Moustafa

Electronics and Communications Engineering Department , Faculty of Engineering, Mansoura University, Mansoura, Egypt, hossam_moustafa@hotmail.com

Fayez Zaki

Electronics and Communications Engineering Department, Faculty of Engineering, Mansoura University, Mansoura, Egypt, fwzaki@yahoo.com

Follow this and additional works at: <https://mej.researchcommons.org/home>

Recommended Citation

Sameh, Sally; Moustafa, Hossam El-Din; and Zaki, Fayez (2021) "Feature Level Fusion of Iris and Voice.," *Mansoura Engineering Journal*: Vol. 40 : Iss. 5 , Article 6.

Available at: <https://doi.org/10.21608/bfemu.2020.96269>

This Original Study is brought to you for free and open access by Mansoura Engineering Journal. It has been accepted for inclusion in Mansoura Engineering Journal by an authorized editor of Mansoura Engineering Journal. For more information, please contact mej@mans.edu.eg.

Feature Level Fusion of Iris and Voice

دمج الصوت وبصمة حدقة العين على مستوى السمات

Sally M. Sameh¹, Hossam El-Din Moustafa², Fayez W. Zaki³

¹ Faculty of Engineering, Mansoura University, Mansoura, Egypt,

E-mail: sally_sameh@hotmail.com

² Faculty of Engineering, Mansoura University, Mansoura, Egypt,

E-mail: hossam_moustafa@hotmail.com

³ Faculty of Engineering, Mansoura University, Mansoura, Egypt,

E-mail: fwzaki@yahoo.com

المخلص

تستخدم نظم المقاييس الحيوية بكثرة من قبل العديد من المنظمات للتطبيقات الأمنية المختلفة. حيث ان الاستخدام الرئيسي لهذه الأنظمة هو في تطبيقات التحقق من الهوية وتحديد الهوية التي تتضمن تسجيل الدخول الشبكة الحاسب، مراقبة جوازات السفر، وتحديد هوية الجثث، وأمن البيانات الإلكترونية، والتعرف على الإرهابيين، والدخول إلى الإنترنت. نظم المقاييس الحيوية الأحادية الواسطة تعاني من الكثير من المشاكل مثل نسبة الخطأ العالية، والأداء المنخفض وهجوم المزيفين وبالتالي يتجه الطلب إلى نظم المقاييس الحيوية المتعددة الوسائط بدلا منها نظم المقاييس الحيوية المتعددة الوسائط لديها العديد من المزايا على نظم المقاييس الحيوية الأحادية الواسطة مثل عدم عموميتها، تغطية أكبر للسكان، وانخفاض معدلات الخطأ، والأداء العالي وارتفاع معدلات القبول الحقيقية.

في هذا البحث، يتم تقديم دراسة للاندماج المتعدد الوسائط من نظم المقاييس الحيوية للصوت وقزحية العين على مستوى الخصائص. يتم استخراج ملامح من الإشارات الصوتية باستخدام معامل تطبيع الطاقة الطيفي وصور القزحية المعالجة باستخدام قيمة التحليل الواحدة. وقد تم القيام بالتجربة باستخدام العينات التي تم جمعها من كلية الهندسة جامعة المنصورة لقاعدة بيانات الصوت وقاعدة بيانات القزحية للأكاديمية الصينية للعلوم ومعهد الأتمتة التي أعطت نسبة دقة 98.4% بعد الدمج. وهذه النتيجة مقبولة وتعطى دقة عالية عن استخدام الصوت والقزحية بشكل فردي

Abstract

Biometric systems are vastly used by various organizations for different security applications. The main use of such systems is in authentication and identification applications which including computer network login, passport control, corpse identification, electronic data security, terrorist identification, Internet access. Unimodal biometric systems suffer from a lot of problems like high error rate, low performance and imposters' attack so the demand of Multimodal biometric systems takes place. Multimodal biometric systems have several advantages over unimodal biometric systems such as non-universality, larger population coverage, lower error rates, higher performance and higher genuine acceptance rates. In this paper, a study of multimodal fusion of voice and Iris at feature level is presented. The features are extracted from the voice signals using Power-Normalized Cepstral Coefficients (PNCC) and from the preprocessed iris images using Single Value Decomposition (SVD). The experiment have been done using samples collected from faculty of Engineering , Mansoura university for voice and CASIA iris database for iris which gave accuracy of 98.4% after fusion. This result is acceptable and gives high accuracy than using voice and iris individually.

Keywords

Multimodal, Feature level fusion, voice recognition, iris recognition

1. Introduction

Biometrics refer to the ability to identify individuals based on their physiological and behavioral characteristics.

Physiological characteristics like fingerprints, hand geometry, retina, iris, facial images etc. While behavioral characteristics include signatures, voice, gait, typing rhythm, etc.

Due to the increasing importance of security, there is a great need to make sure that only authenticated users have access to the system.

Biometrics is a promising solution offering opportunities for a more secure authentication mechanism. However, most of the biometric applications nowadays are unimodal, whereas unimodal biometric system faces a lot of problems and limitations, so to overcome these problems the multimodal biometric system is used [1].

Multimodal biometric systems fuse the results obtained from two or more distinct features of a person for robust biometric authentication task [2]. Multimodal biometric fusion is the process which combines classification results from each biometric trait [3]. According to [4] there are three possible levels of fusion in multimodal biometric systems, which are:

(a) Feature extraction level fusion: The data obtained from each sensor is used to compute a feature vector. As the features extracted from one biometric trait are independent of those extracted from the other; it is reasonable to concatenate the two vectors into a single new vector. The new feature vector has a higher dimensionality and represents a person's identity in a different hyperspace. Feature reduction techniques may be employed to extract useful features from a larger set of features.

(b) Matching score level fusion: Each system provides a matching score indicating the proximity of the feature vector with the template vector. These scores can be combined

to assert the veracity of the claimed identity. Techniques such as logistic regression may be used to combine the scores reported by the two sensors.

(c) Decision level fusion: Each sensor can capture multiple biometric data and the resulting feature vectors are classified individually into the two classes—accept or reject.

A majority vote scheme, such as that employed in [5] can be used to make the final decision according to [6][7] feature level fusion has richer information than match score and decision fusion and it's also able to obtain better performance. In this paper, a feature level fusion between physiological (iris) and behavioral (voice) characteristics is proposed. The rest of the paper is organized into the following sections:

Section 2 – Individual recognizers

Section 3 – Fusion and proposed design

Section 4 – Experimental results

Section 5 - Conclusion

2. Individual recognizers

2.1. Voice recognition

Voice recognition is the process of automatically recognizing individuals by their voice [8], there is no two individuals sound the same, because their vocal tract shapes, larynx sizes, and other parts of their voice production organs are different [9]. This technique makes it possible to use the speaker's voice to verify their identity and to control access to services such as voice dialing, telephone banking, telephone shopping, database access services, information services, voice mail, and security control for confidential information areas, and moreover remote access to computers [10][11].

The important steps involved in voice recognition in this paper are:

1. Noise reduction and Silence removal.
2. Feature extraction (PNCC).
3. Principal component analysis.

2.1.1 Noise Reduction and Silence removal

One of the problems that faces voice recognition is noise in order to overcome this problem speech enhancement is done which refers to the improvement in the quality or intelligibility of a speech signal and the reversal of degradations that have corrupted it. In this paper, speech enhancement is done using Minimum Mean Square Error (MMSE) amplitude estimators [9] [11]. Noise Reduction using (MMSE) can be used where the enhancement of noisy speech signals is essentially an estimation problem in which the clean signal is estimated from a given sample function of the noisy signal. The goal is to minimize the expected value of some distortion measured between the clean and estimated signals. For this approach to be successful, a perceptually meaningful distortion measure must be used, and a reliable statistical model for the signal and noise must be specified [12] [13].

2.1.2 Feature Extraction

Feature extraction is a special form of dimensionality reduction which transforms the input data into a set of features. In this paper a new feature extraction algorithm is used, Power-Normalized Cepstral Coefficients [14] (PNCC) which is based on auditory processing. Major new features of PNCC processing include the usage of a power-law nonlinearity replaces the traditional log nonlinearity used in MFCC coefficients, a noise-suppression algorithm based on asymmetric filtering that suppress background excitation, a module that accomplishes temporal masking and medium-time power analysis, in which environmental parameters are estimated over a longer duration than commonly used for speech, as well as frequency smoothing. PNCC processing provides better recognition accuracy in noisy environments than techniques such as Vector Taylor Series (VTS) and the ETSI Advanced Front End (AFE) while requiring much less computation.

The next Figure (1) illustrates the main steps for PNCC:

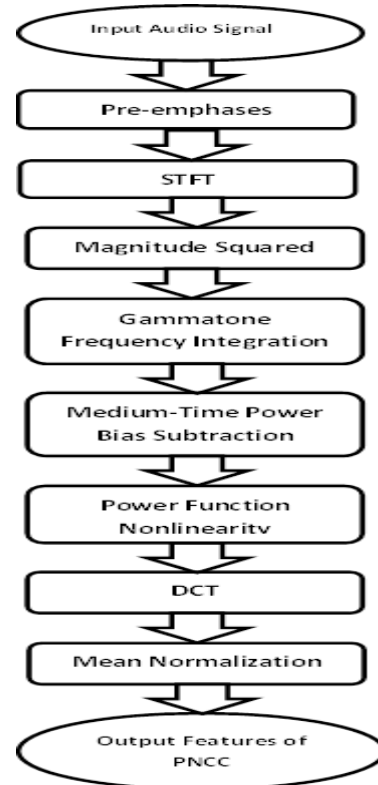


Figure (1) illustrates the main steps for PNCC

2.1.3 Principal component analysis

Although Feature extraction was achieved in the previous part using PNCC, the number of the features is still large, so PCA is used. Principal Component Analysis (PCA) [15] is a statistical procedure that uses orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal component

The number of principal components is less than or equal to the number of original variables. This transformation is defined in such a way that the first principal component has the largest possible variance (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it be

orthogonal to (i.e., uncorrelated with) the preceding components. Principal components are guaranteed to be independent if the data set is jointly normally distributed. PCA is sensitive to the relative scaling of the original variables [16].

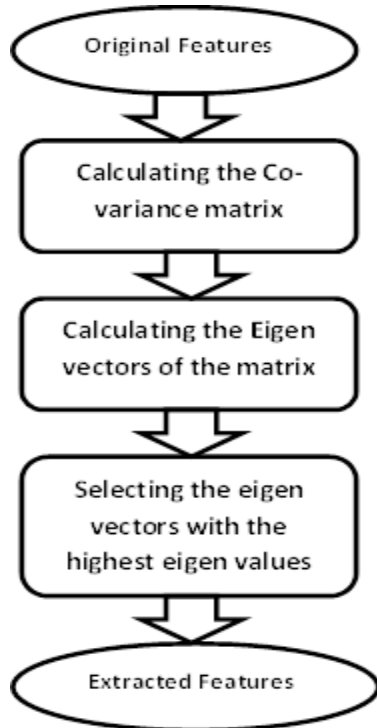


Figure (2) shows the major steps for PCA:

2.2. Iris Recognition

Iris recognition provides one of the most stable biometric signals for identification, with a distinctive texture is formed before age one and remains constant throughout life, unless there is an injury to the eye [17]. Compared with other biometric features such as face and fingerprint, iris Patterns are more stable and reliable[18]. Iris recognition is done through the following steps:

1. Iris segmentation & localization.
 - a. pupil detection
 - b. iris detection.
2. Feature reduction.
3. Feature extraction (SVD).

2.2.1 Iris Segmentation

Iris segmentation & localization is the master key of iris recognition [19]. Segmentation of iris includes separation of iris which locates between the pupil and the sclera of iris images. The purpose of iris localization is to accurately localize the inner and outer boundaries of an iris and to remove the non-useful information like the sclera and the pupil information, and to extract the region of interest iris localization is accomplished through two steps: the detection of inner boundary (iris/pupil) and outer boundary (iris/sclera).

1) Detecting the Pupil (inner boundary)

Precise pupil detection is the most important step for an iris recognition system as known that, the contrast between the iris and the pupil is usually stronger than the contrast between the sclera and the iris, and that the pupil is always the largest dark area, so it's edges can be detected by using suitable threshold on the intensity image.

Pupil detection is done through two steps: First: we apply a proper threshold to the iris image:

$$TH(x) = \begin{cases} F(x) > 70: 1 \\ F(x) \leq 70: 0 \end{cases}$$

where $F(x)$ the original is image and $TH(x)$ is the threshold. If a pixel in the image has an intensity value less than the threshold value, the corresponding pixel in the resultant image is set to black (0). Otherwise, if the pixel intensity value is greater than or equal to the threshold intensity, the resulting pixel is set to white (1).

Second: we apply Freeman's chain code[20][21] to find regions of 8-connected pixels that are with value (0) so we found that eyelashes also fulfill the threshold condition, but as we know that pupil has larger area than eyelashes so we perform a cycle through all regions and apply the following condition:

For each Region M

If $AREA(M) < 2500$

Set all pixels of M to 0

Thus the pupil is localized and extracted as shown in Figure (3).

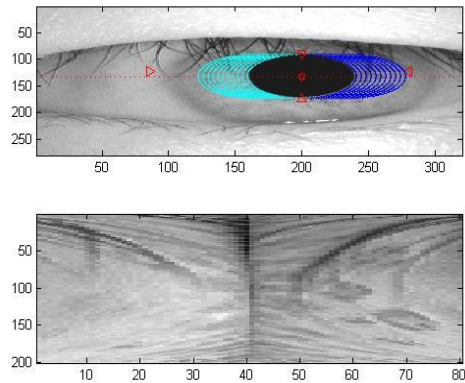


Figure (3) shows the pupil localization and the horizontal and vertical radiuses of the pupil

2) Iris edge detection (outer boundary)

After detecting the pupil location and with the previous knowledge that the pupil located in the center of the iris, the next step is to find the edge of the iris. First we pass a horizontal line that crosses the center of the pupil and go through the iris image starting from the edges of the pupil, we analyze the signal composed by pixel intensity from the center of the image towards the border then we try to detect sudden increases of intensity level. Knowing that the edge between the iris and the sclera has greater intensity than iris pixels this difference is emphasized by applying a linear contrast filter. It is possible that some pixels inside the iris disk are very bright, causing a sudden rise in intensity. That could mislead the algorithm to detect it as an iris edge at that point. To avoid this, we take the average intensity of small windows when the sudden rises occur from these intervals Figure (4) shows iris detection steps.

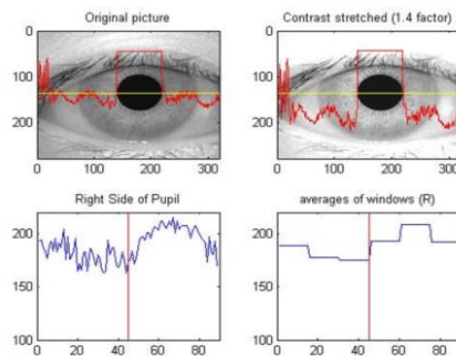


Figure4 image segmentation and finding the edge points of the iris

2.2.2 Feature reduction and extraction

First, Iris Model technique is used to reduce the dimension by eliminating the irrelevant data and focus only on the data that effectively identify the individual. Also, we restrict the mapping of the iris to areas known to have less influence of eyelashes and eyelids, the sides of the iris.

Assuming that intra-class rotation of iris is practically void, we propose an approach to extract pixels of either side of the pupil and forming one reduced image of the iris Figure(5).

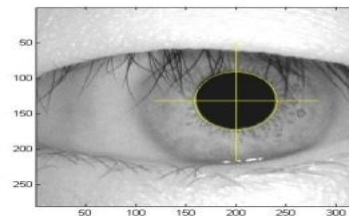


Figure (5) Extraction of the Biometric Template(iris model)

Second, Single Value Decomposition [22] is used to extract the features of the iris and as a dimension reduction tool also. SVD is an algebraic extension of classical vector model, which discovers significant properties and represents the images as linear combinations of the base vectors. Moreover, the base vectors are ordered according to their

significance for the reconstructed image, which allows us to consider only the first k base vectors as important (the remaining ones are interpreted as "noise" and discarded). Furthermore, SVD is more successful in recalling when compared to querying whole image vectors, we decompose the matrix of images compared to querying whole image vectors we decompose the matrix of images A by singular value decomposition (SVD), calculating singular values and singular vectors of A .

The SVD equation (1) for an $(m \times n)$ singular matrix A is

$$A = U\Sigma V^T \quad (1)$$

Where U is an $(m \times m)$ orthogonal matrix, V is an $(n \times n)$ orthogonal matrix and Σ is an $(m \times n)$ diagonal matrix containing the singular values of A arranged in decreasing order of magnitude.

3. The Proposed Fusion Algorithm

The purpose of the fusion is to achieve more security and higher performance than using a single biometric alone. Figure (6) shows the proposed multimodal system. The steps involved in feature fusion are:

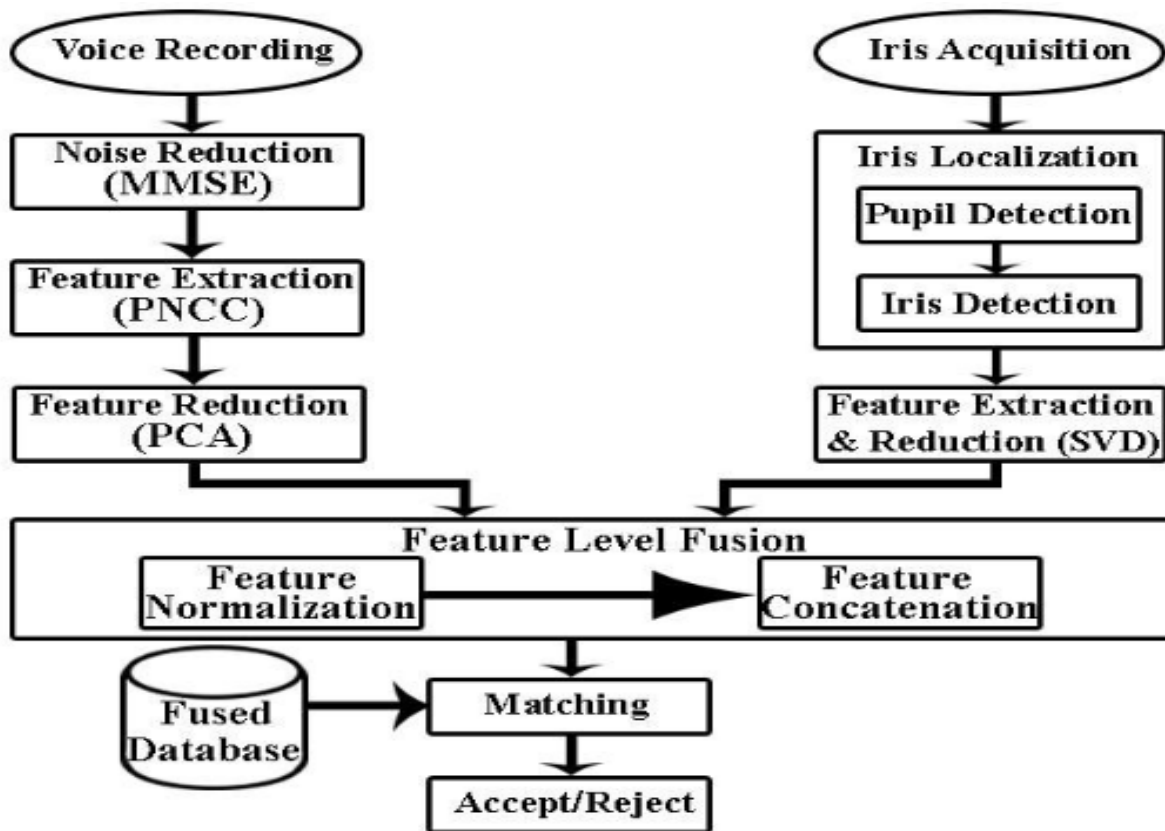


Figure 6 representing the proposed multimodal fusion system

1-Normalization

Although the incompatibility between the features of voice and iris, so in order to fix this problem, we normalized the feature

vectors using the min-max normalization technique [23][24] which normalize the feature vector by scaling between 0 and 1. Let $M = [m_1 m_2 m_3 m_4 \dots m_n]$ represent the feature

vector and M' represent the normalized feature vector

$$M' = \frac{m_i - M_{min}}{M_{max} - M_{min}}$$

Where;

M_{min} = the minimum value of the feature vector

M_{max} = the maximum value of the feature vector

2-Concatination

The feature fusion is obtained by concatenating the normalized feature vectors of voice, iris and performing feature selection using Sequential Forward Selection technique [25] on the concatenated vector in order to overcome the problem of curse of dimensionality [26]. Let $S'_i = \{s'_1 s'_2 s'_3 \dots s'_n\}$ and $W'_i = \{w'_1 w'_2 w'_3 \dots w'_n\}$ represent the normalized feature vectors of voice and iris, respectively, and $Z'_i = \{s'_1 s'_2 s'_3 \dots s'_n w'_1 w'_2 w'_3 \dots w'_n\}$ represent the fused vector.

4-Results and Discussion

The experiments reported in this paper have been conducted on samples collected from faculty of Engineering, Mansoura University for voice and CASIA iris image database version 4.0 for iris. The database consisted of 16 individuals, each providing 7 images for iris and 7 samples for voice. For each individual 4 images and 4 samples have been used in the training phase and 3 images and 3 samples for testing phase. The experiments were conducted using MATLAB & Neural Network toolbox in MATLAB and done on windows 7, core i5 processor and 4 GB RAM computer to obtain accuracy. Two levels of experiments have been done. At the first level iris and voice were tested individually, in this level results and accuracy were computed individually for each of them and was found 96.3% for iris and 94.7% for voice. At the second level, iris and voice were combined at feature level using the proposed model and the overall performance of the

system has increased showing an accuracy of 98.4%. Four performance metrics have been computed: False acceptance rate (FAR), false rejection rate (FRR), Accuracy and Receiver Operating Characteristic curve (ROC). Figure (7) shows the roc curve plotted Genuine Acceptance Rate (GAR) versus False Acceptance Rate (FAR).

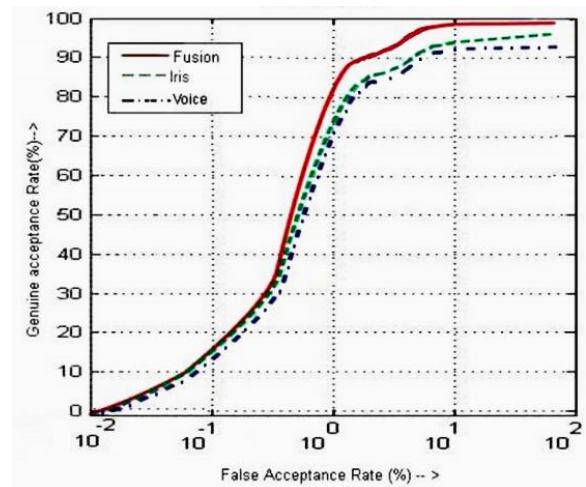


Figure 7 The ROC curves of the unimodal and multimodal

Table (1) shows a performance comparison between the proposed multimodal fusion system and each individual.

Table (1) showing unimodal and multimodal accuracy				
Me tric	V oice	I ris	Fuse d model	
FA	4	3		
R	.23%	.6%		1.4%
FR	1	5		
R	0.3%	.7%		4.6%
Ac curacy	9	9		98.4
	4.7%	6.3%		%

5-Conclousion

This paper discusses the work done on feature level fusion in multimodal system, as it is proved in many aspects, it offers a better fusion method for reliability and quality increasing of biometric systems required for sensitive authentication and verification for both known and unknown individuals. This

work might still get improved in the future in its performance rate with another more advanced suitable feature extraction method.

References:

- [1] S. K. Bhardwaj, "An Algorithm for Feature Level Fusion in Multimodal Biometric System"-Volume 3 Issue 10, (Oct.2014).
- [2] Ashraf Aboshosha, Kamal A. El Dahshan, Eman A. Karam "Score Level Fusion for Fingerprint, Iris and Face Biometrics", (International Journal of Computer Applications) Vol. 111, No.4, (2015)
- [3] K.Sasidhar, Vijaya L Kakulapati, Kolikipogu Ramakrishna & K. Kailasa Rao, "Multimodal Biometric Systems – Study to Improve Accuracy and Performance"-International Journal of Computer Science & Engineering Survey (IJCSES) Vol.1, No.2,(Nov. 2010).
- [4] Ross and A.K. Jain, "Information fusion in biometrics", Pattern Recognition Letters- vol. 24, pp.2115– 2125 (Sep. 2003).
- [5] Y. Zuev and S. Ivanon, "The voting as a way to increase the decision reliability" Foundations of Information Decision fusion with applications to engineering problems, Washington, DC, USA, pp. 206-210 (1996).
- [6] R.Gayathri, P. Ramamoorthy, "Feature Level Fusion of Palmprint and Iris"-Vol. 9, Issue 4, No 1,(Jul 2012).
- [7] Walia, E., Vaidya, M."Feature level fusion of palm print and fingerprint modalities using Discrete Cosine Transform" (ICAETR), International Conference (2014)
- [8] Springer Handbook of Speech Processing, (2008). pp 743-762.
- [9] Tomi Kinnunen, Haizhou Li, "An Overview of Text-Independent Speaker Recognition: from Features to Super vectors".
- [10] Scheffer, N., Ferrer, L., McLaren, M."A novel scheme for speaker recognition using a phonetically-aware deep neural network", (ICASSP), IEEE International Conference (2014)
- [11] http://www.ifp.illinois.edu/~minhdo/teaching/speaker_recognition/
- [12] <http://www.vocal.com/noise-reduction/minimum-mean-square-estimators/>
- [13] R.C. Hendriks, J.S. Erkelens, J. Jensen, and R. Heusdens "Minimum Mean-Square Error Amplitude Estimators for Speech Enhancement under the Generalized Gamma Distribution", IWAENC Paris(Sep 12-14, 2006).
- [14] Chanwoo Kim and Richard M. Stern, "Power-Normalized Cepstral Coefficients (PNCC) for Robust Speech Recognition Member"-IEEE (2013).
- [15] http://en.wikipedia.org/wiki/Principal_component_analysis
- [16] Dong Hyun Jeong, Caroline Ziemkiewicz, William Ribarsky, Remco Chang," Understanding Principal Component Analysis Using a Visual Analytics Tool", (2008)
- [17] Y. Du, R. W. Ives, and D. M. Etter, "Iris Recognition, the Electrical Engineering Handbook"-3rd Edition, Boca Raton, FL: CRC Press (2004).
- [18] K. Saminathan, T.Chakravarthy, M.Chithra Devi,"Iris Recognition based on Kernels of support vector machine "ICTACT Journal on soft computing Vol.5 No.2 (2015)
- [19] Shideh Homayon, "Iris recognition for personal identification using Lamstar neural network", (IJCSIT) Vol. 7, No. 1
- [20] Pulipati Annapurna, Sriraman Kothuri, Srikanth Lukka," Digit Recognition Using Freeman Chain Code" (IJAIEM) Vol.2 No.8 (2013)
- [21] Nor Amizam Jusoh, Jasni Mohamad Zain," Application of Freeman Chain

- Codes: An Alternative Recognition Technique for Malaysian Car Plates”(IJCSNS) Vol.9 No.11 (2009).
- [22] Kirk Baker,” Singular Value Decomposition Tutorial”, (Mar 29, 2005).
- [23] S.Gopal Krishna Patro1, Kishore Kumar sahu,” Normalization: A Preprocessing Stage”,(2015)
- [24] Aarohi Vora, Chirag Paunwala, Mita Paunwala,” Review on Fusion Algorithms for Multimodal Authentication System”,(IJEDR), Vol.3 No.2 (2015)
- [25] Gomez-Barrero,M.,BuschC.”Towards cancelable multi-biometrics based on bloom filters: a case study on feature level fusion of face and iris” IEEE(IWBF), International Workshop (2015)
- [26] Girish Chandrashekar, Ferat Sahin, “A survey on feature selection methods”, Elsevier ,(2014)