IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Vol. 09, Issue 2 (February. 2019), ||V (I) || PP 60-68

The Secured Biometric Protection System Based on Face and Voice Recognition

Addepally Sandeep Kumar[1], KompellaVenkataRamana[2]
[1]M.Tech (CSE), Dept. of Computer Science and Systems Engineering,
[2]Professor, Dept. of Computer Science and Systems Engineering, Andhra University College of Engineering (A), Visakhapatnam, Andhra Pradesh, India.
Corresponding Author: Addepally Sandeep Kumar

Abstract: The recognition accuracy of unimodal biometric systems needs to fight with an assortment of issues, for example, foundation noise, noisy data, non-comprehensiveness, and parody assaults, intra-class varieties, between class likenesses or peculiarity, interoperability issues. This paper depicts another multimodal biometric framework that incorporates various characteristics of a person for recognition, which can reduce the issues looked by unimodal biometric framework while enhancing recognition execution. We have built up a multimodal biometric framework by consolidating iris, face and voice at match score level utilizing straightforward whole guideline. The match scores are standardized by min-max standardization. The personality built up by this framework is considerably more solid and exact than the individual biometric systems. Test assessments are performed on an open dataset showing the accuracy of the proposed framework. The adequacy of proposed framework in regards to FAR (False Accept Rate) and GAR (Genuine Accept Rate) is shown with the assistance of MUBI (Multimodal Biometrics Integration) programming.

Keywords: Multimodal Biometric System, Iris recognition, Face recognition, Voice recognition.

Date of Submission: 21-01-2019 Date of acceptance:05-02-2019

I. INTRODUCTION

The term biometrics is gotten from the Greek word Bio and Metric. The term biometrics identifies with the estimation (metric) of qualities of a living (Bio) thing with the end goal to perceive a man. Biometrics utilizes different physiological or conduct qualities. Regular physiological biometric estimations incorporate fingerprints, iris, confront, hand, retina, and so forth. While basic conduct biometric estimations incorporate mark, speech, mood, and so forth. Single biometric systems have confinements like uniqueness, high mocking rate, high mistake rate, non-all-inclusiveness and noise [1]. Multimodal biometric recognizable proof framework is used for explaining these constraints. Multimodal biometric is the field of example recognition look into perceiving the human personality dependent on physical examples or standards of conduct of human. Biometric method gives the different qualities of a man which is constantly common. Along these lines the advantage to a biometric is that it doesn't change or lose. Single biometric framework may prompt False Acceptance Rate (FAR) and False Rejection Rate (FRR). Biometric recognition systems are intrinsically probabilistic and their execution should be surveyed inside the setting of this principal and real trademark. Biometric recognition includes coordinating, inside a resilience of guess of watched biometric characteristics against natural qualities and practices of a man. The execution of a biometric framework is affected by the unwavering quality of the sensor utilized and the degrees of opportunity offered by the highlights separated from the detected flag. Likewise, if the biometric highlight is identified or estimated is noisy (eg a unique finger impression with a scar or a broken voice by cool), the last score figured by the adjustment module may not be solid. This issue can be comprehended by various biometric attributes. Diverse biometric qualities are utilized by these systems [2]. Biometric systems that utilization in excess of a physiological or social attribute for recognizable proof are called multimodal biometric systems. Multimodal biometric systems ought to be more dependable because of the nearness of various confirmations [3].

II. RELATED WORK

A ton of work has been done in the most recent years in the field of multimodal biometrics yielding adult half breed biometric systems. Combination at the match score level has been widely examined in the writing and is the predominant dimension of combination in biometric systems. Luca et al. [5] utilized unique finger impression and face to be intertwined at the match score level. PCA and LDA are utilized for the element extraction and characterization. Mean guideline, item rule and Bayesian standard are utilized as the combination procedures with FAR of 0% and FRR of 0.6% to 1.6%. Kartik et al. [6] joined speech and mark by utilizing entirety rule as combination procedure after the min max standardization is connected. Euclidean separation is utilized as the characterization procedure with 81.25% accuracy execution rate. Rodriguez et al. [7] utilized mark with iris by utilizing entirety guideline and item rule as the combination procedures. Neural Network is utilized as the characterization strategy with EER beneath than 2.0%. Toh et al. [8] joined hand geometry, unique mark and voice by utilizing worldwide and nearby learning choice as combination approach. The accuracy execution is 85% to 95%. Feng et al. [9] joined face and palmprint at highlight level by connecting the highlights extricated by utilizing PCA and ICA with the closest neighbor classifier and bolster vector machine as the classifier. Fierrez-Aguilar and Ortega-Garcia [10] proposed a multimodal approach including face, a details based unique finger impression and online mark with combination at the coordinating score level. The combination approach acquired Equal Error Rate (EER) of 0.5. Viriri and Tapamo [10] presented a multimodal approach including iris and mark biometrics at score level combination with False Reject Rate (FRR) 0.008% on a False Accept Rate (FAR) of 0.01%. Kisku et al. proposed a multibiometric framework including face and Palmprint biometrics at highlight level combination. The framework achieved 98.75% recognition rate with 0% FAR. Meraoumia et al. [3] exhibited a multimodal biometric framework utilizing hand pictures and by coordinating two distinctive biometric qualities palmprint and finger-knuckle-print (FKP) with EER = 0.003 %. Aggithaya et al. [4] proposed an individual validation framework that at the same time abuses 2D and 3D Palmprint highlights. The aggregate guideline classifier accomplishes the best EER of 0.002. Kazi and Rody [5] displayed a multimodal biometric framework utilizing face and mark with score level combination. The outcomes demonstrated that face and mark based bimodal biometric framework can enhance the accuracy rate about 10%, higher than single face/signature based biometric framework.

III. MULTIMODAL BIOMETRIC SYSTEM

Most of the biometric systems deployed in real world applications are unimodal which rely on the evidence of single source of information for authentication e.g. fingerprint, face, voice etc. These systems are vulnerable to variety of problems such as noisy data, intra-class variations, inter-class similarities, nonuniversality and spoofing.

These limitations are overcome by using multimodal biometrics. Multimodal Biometrics are systems that are capable of using more than one physiological or behavioral characteristic for enrollment, verification or identification. The term “multimodal” is used to combine two or more different biometric sources of a person sensed by different sensors. A generic biometric system has sensor module to capture the trait, feature extraction module to process the data to extract a feature set that yields compact representation of the trait, classifier module 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. The benefits of multimodal biometrics is that by using more than one means of identification, your system can retain a high threshold recognition setting and your system administrator can decide the level of security that is needed.

They also effectively deter spoofing because it is near impossible to spoof multiple biometric traits and the system can request the user to present random traits that only a live person can do. This greatly reduces the probability of admitting an imposter. It leads to considerably high false acceptance rate (FAR) and false rejection rate (FRR), limited discrimination capability, upper bound in performance and lack of permanence. For biometric identification to be ultrasecure and provide above average accuracy, more than one type must be used, as only one form of it may not be accurate enough. One example of this inaccuracy is in the area of fingerprints where at least 10% of people have worn, cut or unrecognizable prints. Combining experts, each based on a different modality such as speech, face, fingerprint, etc., increases the performance and robustness of identity authentication systems [10].

Two different properties of the same biometric can also be combined. In orthogonal multimodal biometrics, different biometrics are involved with little or no interaction between the individual biometric whereas independent multimodal biometrics processes individual biometric independently. Orthogonal biometrics are processed independently by necessity but when the biometric source is the same and different properties are sensed, then the processing may be independent, but there is at least the potential for gains in performance through collaborative processing. In collaborative multimodal biometrics the processing of one biometric is influenced by the result of another biometric.

In multimodal biometric system information reconciliation can occur at the data or feature level, at the match score level generated by multiple classifiers pertaining to different modalities and at the decision level. Biometric systems that integrate information at an early stage of processing are believed to be more effective than those which perform integration at a later stage. Since the feature set contains more information about the input biometric data than the matching score or the output decision of a matcher, fusion at the feature level is expected to provide better recognition results [16]. However, fusion at this level is difficult to achieve in practice because the feature sets of the various modalities may not be compatible and most of the commercial biometric systems do not provide access to the feature sets which they use. Fusion at the decision level is considered to be rigid due to the availability of limited information.

Thus, fusion at the match score level is usually preferred, as it is relatively easy to access and combine the scores presented by the different modalities. Many of these techniques require the scores for different modalities to be normalized before being fused and develop weights for combining normalized scores. Normalization and developing weights is done using Bayesian Belief Network.

The biometric sensors must be consistent in performance under many environmental operations, have embedded privacy functions, protective solutions, enhance public confidence in biometric technology and must totally safeguard your personal information.

IV. FACE RECOGNITION SYSTEM

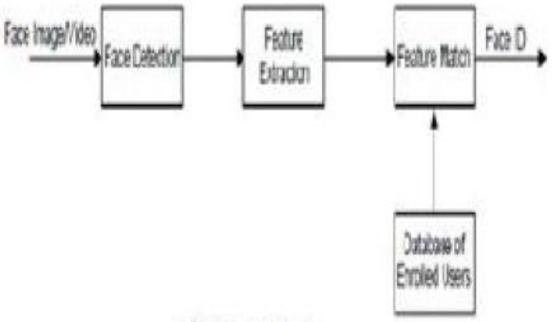
A facial recognition system is a computer application for automatically identifying or verifying a person from a digital image or a video frame from a video source. Face recognition is one of the few biometric methods that possess the merits of both high accuracy and low intrusiveness. It has the accuracy of a physiological approach without being intrusive. For this reason, since the early 70's, face recognition has drawn the attention of researchers in fields from security, psychology, and image processing, to computer vision.

- They are non-intrusive.
- Biometric data of the faces (photos, videos) can be easily taken with available devices like cameras.
- One biometric data is used in many different environments.
- Facial recognition sounds rather interesting in comparison with other biometric technologies.

Therefore, face recognition has been widely used in identification and access management. At the moment, there have been a lot of researches on access control applications and those have been utilized in personal computers and handheld devices authentication.

A. Face Recognition Model

Face recognition systems use a “learning” mechanism to collect data on facial characteristics of users. Hence, the first important point to care about in a face recognition model is the Face Database storing this information.



When the system finishes scanning a video or photo of a user’s face, the digitalized information will go through these following modules one after another:

- Face Detection: locating the face in the photo or video and removing unnecessary details on the background.
- Feature Extraction: extracting facial characteristics needed for recognition.
- Feature Match: comparing scanned information with database to decide if it matches some user’s face. If the face matched, the ID of the corresponding is returned.
-

Numerous algorithms have been proposed for face recognition. One of the ways to do this is by comparing selected facial features from the image and a facial database. Some facial recognition algorithms identify faces by extracting landmarks, or features, from an image of the subject's face. [3] For example, an algorithm may analyze the relative position, size, and/or shape of the eyes, nose, cheekbones, and jaw. These features are then used to search for other images with matching features [3][4][5]. Other algorithms normalize a gallery of face images and then compress the face data, only saving the data in the image that is useful for face detection. A probe image is then compared with the face data. One of the earliest successful systems is based on template matching techniques applied to a set of salient facial features, providing a sort of compressed face representation.

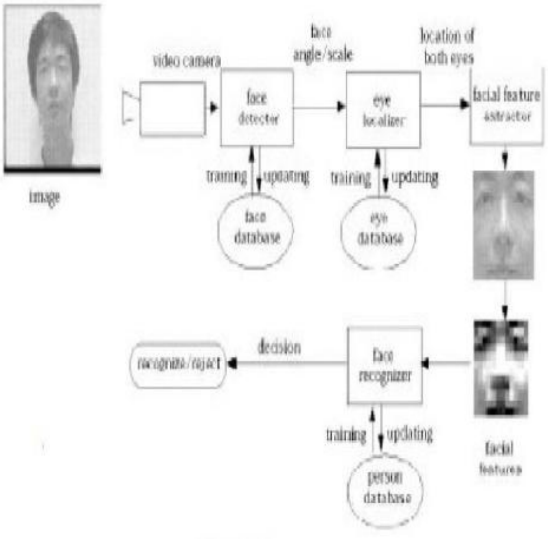
B. Generic Framework

A face recognition algorithm can be divided into the following functional modules:

- Face image detector that finds the locations of human faces from a normal picture against simple or complex background
- Face recognizer that determines who this person is.

Both the face detector and the face recognizer have a feature extractor that transforms the pixels of the facial image into a useful vector representation, and a pattern recognizer that searches the database to find the best match to the incoming face image [6]. The difference between the two is the following. In the face detection scenario, the pattern recognizer categorizes the incoming feature vector to one of the two image classes: “face” images and “non-face images. In the face recognition scenario, on the other hand, the recognizer classifies the feature vector as some person’s face that is already registered in the database. Figure III.B.1 depicts the generic framework:

Feature extraction refers to a process whereby the given data is compressed to features or patterns fewer in number than the given data. Thus, the data space is transformed into a feature space and undergoes a dimensionality reduction. Evidently, this data compression is lossy.



C. Eiegnface based Face Detection

The information theory approach of encoding and decoding face images extracts the relevant information in a face image, encode it as efficiently as possible and compare it with database of similarly encoded faces. The encoding is done using features which may be different or independent than the distinctly perceived features like eyes, ears, nose, lips, and hair. There are patterns which occur in any input signal or image data, which can be observed in all signals, in the domain of facial recognition - the presence of some objects (eyes, nose, mouth) in any face as well as relative distances between these objects. These characteristic features are called eigenfaces in the facial recognition domain or principal components generally. They can be extracted out of original image data by means of a mathematical tool called Principal Component Analysis (PCA).

By means of PCA one can transform each original image of the training set into a corresponding eigenface [3][4][5]. An important feature of PCA is that one can reconstruct any original image from the training set by combining the eigenfaces. Eigenfaces are nothing less than characteristic features of the faces. Therefore the original face image can be reconstructed from eigenfaces if one adds up all the eigenfaces (features) in the right proportion. Each eigenface represents only certain features of the face, which may or may not be present in the original image. If the feature is present in the original image to a higher degree, the share of the corresponding eigenface in the”sum” of the eigenfaces should be greater. If, contrary, the particular feature is not or almost not present in the original image, then the corresponding eigenface should contribute a smaller or not at all part to the sum of eigenfaces. So, in order to reconstruct the original image from the eigenfaces, one has to build a kind of weighted sum of all eigenfaces. That is, the reconstructed original image is equal to a sum of all eigenfaces, with each eigenface having a certain weight. This weight specifies, to what degree the specific feature (eigen face) is present in the original image. If all the eigenfaces extracted from original images are used, one can reconstruct the original images from the eigenfaces exactly. But only a part of the eigenfaces can also be used. Then the reconstructed image is an approximation of the original image. However, losses due to omitting some of the eigenfaces can be minimized. This happens by choosing only the most important features or eigenfaces.

Omission of eigenfaces is necessary due to scarcity of computational resources. It is possible not only to extract the face from eigenfaces, given a set of weights, but also to go the opposite way. This opposite way would be to extract the weights from eigenfaces and the face to be recognized. These weights tell the amount by which the face in question differs from”typical” faces represented by the eigenfaces. Therefore, using these weights two important things can be determined:

- Determine, if the image in question is a face at all. In case the weights of the image differ too much from the weights of face images, the image probably is not a face.
- Similar faces (images) possess similar features to similar degrees (weights). If weights from all the images available are extracted, the images could be grouped to clusters. That is, all images having similar weights are likely to be similar faces.

V. VOICE RECOGNITION SYSTEM

Voice has been cited as the most acceptable of biometrics. The applications of speech signal processing such as speech recognition and speaker identification have been drastically increased in recent years, because of its noncontact characteristic and speaker identification system that can be utilized to suspect identification. General overviews of speaker recognition have been given by Atal, Doddington, Furui, O’Shaughnessy, Rosenberg, Soong, Sutherland, and Jack.

The voice biometric is a physical measure. The vibration of the vocal chords and the patterns created by the physical components resulting in human speech are as distinctive as fingerprints. In the case of voice authentication, there is both a physiological biometric component such as voice tone and pitch and a behavioural component such as accent. This makes it very useful for biometric authentication. Speaker recognition is the process of automatically recognizing who is speaking by using the speakerspecific information included in speech waves to verify identities being claimedby people accessing systems; that is, it enables access control of various services by voice .It is the identification of the person who is speaking by characteristics of their voices, also called "voice recognition". There is a difference between speaker recognition and speech recognition. Speech recognition is used to identify words in spoken language. Voice recognition is a biometric technology used to identify a particular individual's voice.

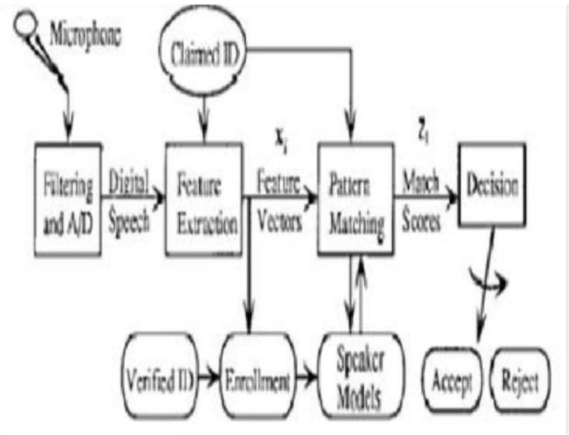
Each speaker recognition system has two phases:

- Enrolment
- Verification.

During enrolment, the speaker's voice is recorded and typically a number of features are extracted to form a voice print, template, or model. In the verification phase, a speech sample or "utterance" is compared against a previously created voice print.

A. Speaker Authentication Model

Automatic speaker verification (ASV) is the use of a machine to verify a person’s claimed identity from his voice. Speaker verification is defined as deciding if a speaker is whom he claims to be. The figure IV.A.1 below shows the block diagram of ASV.



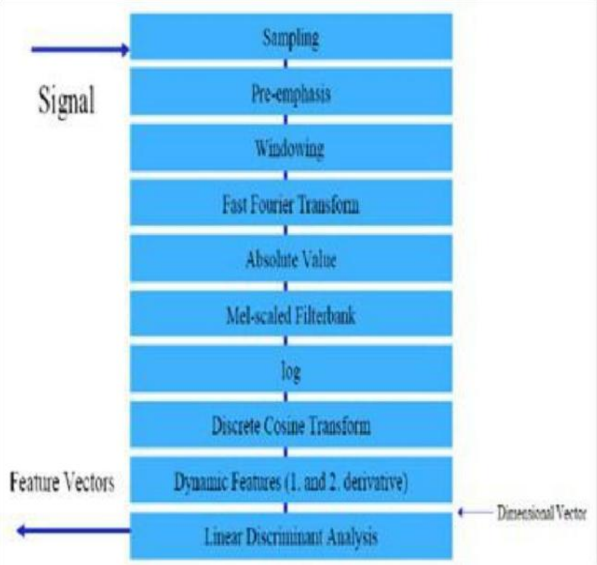
The general approach to voice authentication consists of five steps: Digital speech data acquisition, feature extraction, pattern matching, making an accept/reject decision, and enrolment to generate speaker reference models. Feature extraction maps each interval of speech to a multidimensional feature space. A speech interval typically spans 10–30 ms of the speech waveform and is referred to as a frame of speech. This sequence of feature vectors xi is then compared to speaker models by pattern matching. This results in a match score zi for each vector or sequence of vectors. The match score measures the similarity of the computed input feature vectors to models of the claimed speaker or feature vector patterns for the claimed speaker. Last, a decision is made to either accept or reject the claimant according to the match score or sequence of match scores, which is a hypothesis testing problem.

B. Speech Feature Extraction

The speech feature extraction in a categorization problem is about reducing the dimensionality of the input-vector while maintaining the discriminating power of the signal. The number of training and test vector needed for the classification problem grows exponential with the dimension of the given input vector, hence feature extraction is needed. But extracted feature should meet some criteria while dealing with the speech signal, such as:

- Easy to measure extracted Speech features.
- Distinguish between speakers while being lenient of intra speaker variability’s.
- It should not be susceptible to mimicry
- It should show little fluctuation from one speaking environment to another.
- It should be stable over time.
- It should occur frequently and naturally in speech.

Here, the Mel Frequency Cepstral Coefficients (MFCC) technique is used to extract features from the speech signal and compare the unknown speaker with the exist speaker in the database. The complete pipeline of Mel Frequency Cepstral.



Mel Frequency Cepstral Coefficients (MFCC) algorithm is feature-extraction type speaker recognition method.

Some of the limitations of the speech recognition algorithms are

- Ambient noise levels can impede both collections of the initial and subsequent voice samples. Noise reduction algorithms can be employed to improve accuracy, but incorrect application can have the opposite effect. Performance degradation can result from changes in behavioural attributes of the voice and from enrolment using one telephone and verification on another telephone.
- Voice changes due to ageing may impact system performance over time. Some systems adapt the speaker models after each

successful verification to capture such long term changes in the voice, though there is debate regarding the overall security impact imposed by automated adaptation Digitally recorded audio voice identification and analogue recorded voice identification uses electronic measurements as well as critical listening skills that must be applied by a forensic expert in order for the identification to be accurate.

C. K Means Clustering Algorithm

The K-means algorithm is a way to cluster the training vectors to get feature vectors. In this algorithm clustered the vectors based on attributes into k partitions. It use the k means of data generated from Gaussian distributions to cluster the vectors. The objective of the k-means is to minimize total intracluster variance, V. The process of k-means algorithm used least divides the input vectors into k initial sets. It then calculates the mean point, or centroid, of each set. It constructs a new partition by associating each point with the closest centroid. Then the centroids are recalculated for the new clusters, and algorithm repeated until when the vectors no longer switch clusters or alternatively centroids are no long changed.

VI. ARCHITECTURE OF PROPOSED SYSTEM

The structural design of proposed multimodal biometric recognition system integrating iris, face and voice is shown in Fig.In the operational phase, the three biometric sensors capture the images individually from the person to be identified and converts them to a raw digital format, which is further processed by the feature extraction modules individually to produce a compact representation that is of the same format as the templates stored in the corresponding databases taken during the enrollment phase. The three resulting representations are then fed to the three corresponding matchers. Here, they are matched with templates in the corresponding databases to find the similarity between the two feature sets. The match scores generated from the individual biometrics are then passed to the fusion module to perform fusion at match score level using simple sum rule.

1) Fusion: The first step involved in fusion is score normalization. Since the match scores output by the three biometric traits (face and voice) are heterogeneous because they are not on the same numerical range, so score normalization is done to transform these scores into a common domain prior to combining them. Here, min-max normalization is used to transform all these scores into a common range [0, 1]. The three normalized scores are fused using sum rule to generate final match score. Finally, fused matching score is passed to the decision module where a person is declared as genuine or an imposter. The normalized scores are obtained by following min-max equation:

$$S_i^{'} = \frac{S_i - S_{\min}}{S_{\max} - S_{\min}} \qquad (2)$$

where S' i is the normalized matching score, Si is the matching score, Smin is the minimum match score and Smax is the maximum match score for ith biometric trait. In order to combine the match scores output by the three individual matchers (iris, face and voice), simple sum rule is used and its equation is given below;

$$Sum = \sum_{i-1}^{n} S_i \qquad (3)$$

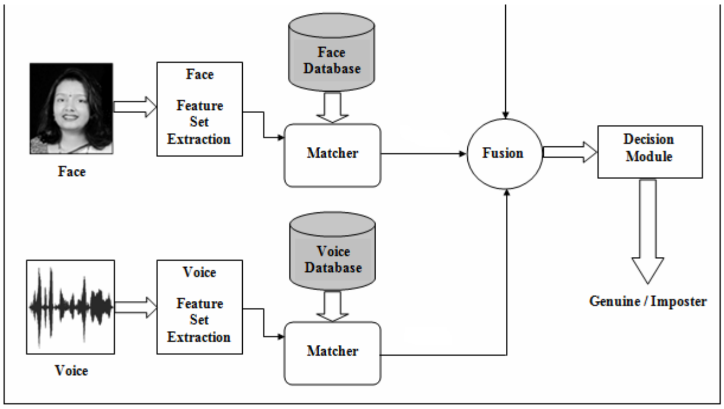


Fig. Architecture of proposed multimodal biometric recognition system integrating face and voice

VII. CONCLUSION

This paper provides results obtained on a multi-modal biometric system that uses face and voice features for recognition purposes. We showed that the resulting system (multi-modal) considered here provide better performance than the individual biometrics. For the near future we are collecting data corresponding to threebiometric indicators - fingerprint, face and voice in order to conceive a better multi-modal recognition system.

REFERENCES

[1]. C. Sanderson and K. K. Paliwal, Information Fusion and Person Verification Using Speech and Face, Information. Research Paper IDIAP-RR 02-33, IDIAP, September 2002.

[2]. A. Ross, K. Nandakumar, and A. K. Jain, Handbook of Multibiometrics, New York: Springer, 2006.

[3]. A. Ross and R. Govindarajan, Feature Level Fusion Using Hand and Face Biometrics, In Proceedings of SPIE Conference on Biometric Technology for Human Identification II, volume 5779, pages 196–204, Orlando, USA, March 2005.

[4]. A.K. Jain, A. Ross, Multibiometric systems, Communications of the ACM, Special Issue on Multimodal Interfaces, Vol. 47, January 2004, 34-40.

[5]. Gian Luca Marcialis and Fabio Roli, “Serial Fusion of Fingerprint and Face Matchers”, M. Haindl, MCS 2007, LNCS volume 4472, pp. 151-160, © Springer-Verlag Berlin Heidelberg 2007.

[6]. Kartik.P, S.R. MahadevaPrasanna and Vara.R.P, “Multimodal biometric person authentication system using speech and signature features,” in TENCON 2008 - 2008 IEEE Region 10 Conference, pp. 1-6, Ed, 2008.

[7]. Rodriguez.L.P, Crespo.A.G, Lara.M and Mezcua.M.R, “Study of Different Fusion Techniques for Multimodal Biometric Authentication,” in Networking and Communications. IEEE International Conference on Wireless and Mobile Computing, 2008.

[8]. Toh.K.A, J. Xudong and Y. Wei-Yun, “Exploiting global and local decisions for multimodal biometrics verification,” Signal Processing, IEEE Transactions on Signal Processing, vol. 52, pp. 3059-3072, 2004.

[9]. G. Feng, K. Dong, D. Hu and D. Zhang, “When Faces Are Combined with Palmprints: A Novel Biometric Fusion Strategy,” in Biometric Authentication. vol. 307, 2004.

[10]. J. Fierrez-Aguilar, J. Ortega-Garcia, D. Garcia-Romero, and J. Gonzalez Rodriguez, “ A comparative evaluation of fusion strategies for multimodal biometric verification,” in Proc. 4th Int, Conf,Audio-video-based Biometric Person Authentication , J. Kittler and M. Nixon, Eds., vol. LNCS 2688, pp. 830–837, 2003.

Authors

ADDEPALLY SANDEEP KUMAR Holds a B.Tech certificate from AvanthiCollege of engineering, Narsipatnam, AP, India. He is currently pursuing M.Tech degree in the department of computer science and System Engineering, Andhra University College of engineering.

KOMPELLA VENKATA RAMANA is a professor in the department of Computer Science and System Engineering at Andhra University College of engineering, Visakhapatnam, AP, India. He received the BE degree in Electronics and Communication Engineering and ME degree in Computer Science and Engineering from Andhra University, Vishakapatnam, AP, India. He received Ph.D in the department of computer science and System Engineering, Andhra University, Visakhapatnam, AP, India. His Specialization in Image Processing, Compilers, System Software.

IOSR Journal of Engineering (IOSRJEN) is UGC approved Journal with Sl. No. 3240, Journal no. 48995.

Addepally Sandeep Kumar. “The Secured Biometric Protection System Based on Face and Voice Recognition.” IOSR Journal of Engineering (IOSRJEN), vol. 09, no. 02, 2019, pp. 60-68.