ISSN 2349-6193
Impact Factor (PIF): 2.243

IJESMR
International Journal OF Engineering Sciences & Management Research

SMARTPHONE BASEDFACE AND VOICE AUTHENTICATION IN MOBILE DEVICESBY USING MFCC ALGORITHM

***Mr.T. S. Arulananth, Dr.T.Jayasingh, Dr. N.Gangatharan, Mr.M.Baskar**
*Asst. Professor/ ECE,RMKCET,Thiruvallur-206. (Research Scholar, Faculty of ECE, Dr. M.G.R. University, Chennai)
Dean, CSI Institute of Technology, Thovalai, Nagercoil
Professor and Head,Dept.ofECE,RMKCET,Thiruvallur-206,
Associate. Professor/ CSE,RMKCET,Thiruvallur-206

Keywords: -Haar Cascade Classifier, Principle Component Analysis (PCA),Mel-Frequency Cepstral Coefficients, Face recognition,Sphinx-4 speech recognition system.

ABSTRACT
The Smart phones are capable of doing new kind of protection mechanism for sensitive data stored in the mobile. Smartphones with camera and microphone can be used for biometric authentication mechanism to protect data which is stored in the phone memory. All human voice is unique in nature. Using the integrated camera on the mobile device we can verify that users are who they claim to be from their facial biometrics. This work based on making rapid authentication using face variations and voice biometrics.

INTRODUCTION
This paper is fully based on rapid face and voice verification to secure sensitive data from unauthorized person. To make an efficient storing feature of image and voice, we use phone memory and retrieve the data for authentication. While capturing face and voice biometric, for authentication using Smartphone camera and microphone, this proposed system matches these biometric with the existing data.If data matches, the user is authorized to access information and this system protects the mobile device from unauthorized user.

SYSTEM FOR FACE VERIFICATION
A. Face Detection using Haar Cascade Classifier
The object detector described below has been initially proposed by Paul Viola [Viola01]&Jones and improved by Rainer Lienhart [Lienhart02].
First, a classifier (namely a *cascade of boosted classifiers working with haar-like features*) is trained with a few hundred sample views of a particular object
(i.e., a face or a car), called positive examples, that are scaled to the same size (say, 20x20), and negative examples - arbitrary images of the same size. After a classifier is trained, it can be applied to a region of interest (of the same size as used during the training) in an input image. The classifier outputs a "1" if the region is likely to show the object (i.e., face/car), and "0" otherwise. To search for the object in the whole image one can move the search window across the image and check every location using the classifier. The classifier is designed so that it can be easily "resized" in order to be able to find the objects of interest at different sizes, which is more efficient than resizing the image itself. So, to find an object of an unknown size in the image the scan procedure should be done several times at different scales. The word "cascade" in the classifier name means that the resultant classifier consists of several simpler classifiers (*stages*) that are applied subsequently to a region of interest until at some stage the candidate is rejected or all the stages are passed. The word "boosted" means that the classifiers at every stage of the cascade are complex themselves and they are built out of basic classifiers using one of four different boosting techniques (weighted voting). Currently Discrete Adaboost, Real Adaboost, Gentle Adaboost and Logit boost are supported.

The basic classifiers are decision-tree classifiers with at least 2 leaves. Haar-like features are the input to the basic classifiers, and are calculated as described below. The current algorithm uses the following Haar-like features as shown in figure 1

ISSN 2349-6193
Impact Factor (PIF): 2.243

IJESMR
International Journal OF Engineering Sciences & Management Research

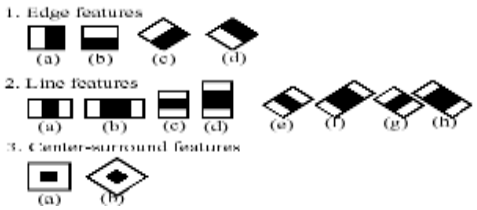

Fig. 1. Haar based features detection.

The feature used in a particular classifier is specified by its shape (1a, 2b etc.), position within the region of interest and the scale (this scale is not the same as the scale used at the detection stage, though these two scales are multiplied). For example, in the case of the third line feature (2c) the response is calculated as the difference between the sum of image pixels under the rectangle covering the whole feature (including the two white stripes and the black stripe in the middle) and the sum of the image pixels under the black stripe multiplied by 3 in order to compensate for the differences in the size of areas. The sums of pixel values over a rectangular region are calculated rapidly using integral images.

B. Face Recognition

Face recognition is one of the nonintrusive biometric techniques commonly used for verification and authentication. PCA is a mathematical procedure that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components.

PRINCIPLE COMPONENTANALYSIS(PCA)

Biometrics is the automated technique of measuring the physical characteristics and comparing those characteristics to a database for example finger recognition, retinal scanning, face recognition etc. Face reorganization is the most popular technique. It is used for identifying or Verifying a person especially for security and identification purposes. Here for better performance the SDCard i.e. the phone memory is used for the storage purpose instead of a database. Physical contact with scanner is not needed in this methodology. Implementing face recognition with PCA algorithm. PCA is a best dimensionality reduction tool which helps to reduce large dataset to a smaller one. In this we preserve as much information as possible in mean square sense. Training images are taken; covariance matrix of the average images is computed. Eigen vectors are also constructed and these are stored in cache files. When an image is given as an input then Euclidean distance is calculated and the image with minimum Euclidean distance is retrieved .Not only for 2D images but this algorithm can be implemented on 3D images for better performance. When a 2D image is given into the Dataset the 3Dportions of that image with varying angle are constructed and stored. This kind of storage will increase the efficiency. At the time of authentication of the user, we capture the image from the camera and the captured image is verified with the stored binary information.

PROPOSED SYSTEM FOR VOICE VERIFICATION

The goal of this project is to develop a speaker verification system in the platform independent software language Java, and since the system will be using **Mel Frequency cepstral coefficients** (MFCC)and it is classified as a text-independent speaker verification system since its task is to verify the person who speaks regardless of what this person is saying.

All speaker recognition/verification system contains of two basic building blocks Speech feature extraction and feature matching.

A. Speech Feature Extraction

The each speech utterance is between one and two seconds in length. We separate each utterance into 60 ms segments with 10 ms time shifts. Since human speech consists of audible and inaudible segments, we only analyze the acoustic features for the audible segments, and we ignore the inaudible ones. A segment is selected to be an audible segment if a certain percentage of the samples' absolute amplitudes in that segment are above a certain threshold. Mel-Frequency Cepstral Coefficients (MFCC) feature extraction process contains following steps.

1) Pre-processing: In the pre-processing stage first each signal is de-noised by soft-thresholding the wavelet coefficients, and since the silent parts of the signals do not carry any useful information, those parts including

ISSN 2349-6193
Impact Factor (PIF): 2.243

IJESMR
International Journal OF Engineering Sciences & Management Research

the leading and trailing edges are eliminated by thresholding the energy of the signal. The signals are divided into frames using a Hamming window of length 23 ms.

2)Framing: It is a process of segmenting the speech samples obtained from the analog to digital conversion (ADC), into the small frames with the time length within the range of 20-40 Ms. Framing enables the non stationary and random based speech signal to be segmented into quasi-stationary frames, and enables Fourier Transformation of the speech signal. It is because, speech signal is known to exhibit quasi-stationary behaviour within the short time period of 20-40 ms.

3)Windowing*:* Windowing step is meant to window each individual frame, in order to minimize the signal discontinuities at the beginning and the end of each frame. A typical window utilized for speaker verification is the Hamming window.

4)Fast Fourier Transform (FFT): This algorithm is ideally used for evaluating the frequency spectrum of speech. FFT converts each frame of N samples from the time domain into the frequency domain. The FFT is a fast algorithm to implement the Discrete Fourier Transform (DFT).

5) Mel Filterbank and Frequency wrapping*:* The Mel filter bank consists of overlapping triangular filters with the cutoff frequencies determined by the center frequencies of the two adjacent filters. The filters have linearly spaced centre frequencies and fixed bandwidth on the Mel scale.

6) Take Logarithm: The logarithm has the effect of changing multiplication into addition. Therefore, this step simply converts the multiplication of the magnitude in the Fourier transform into addition.

7) Take Discrete Cosine Transform (DCT):

It is used to orthogonalise the filter energy vectors. Because of this orthogonalization step, the information of the filter energy vector is compacted into the first number of components and shortens the vector to number of components. The process of calculating MFCC is shown.

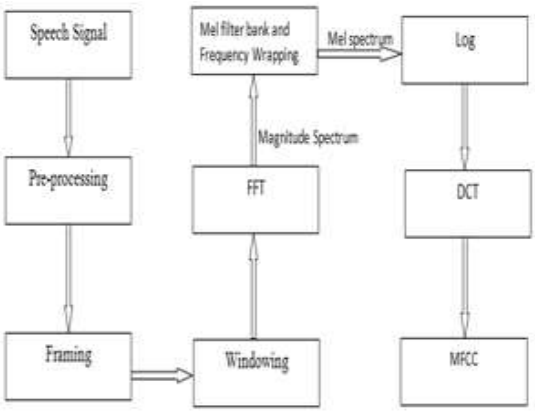

Fig. 2.Feature extraction using MFCC processor.

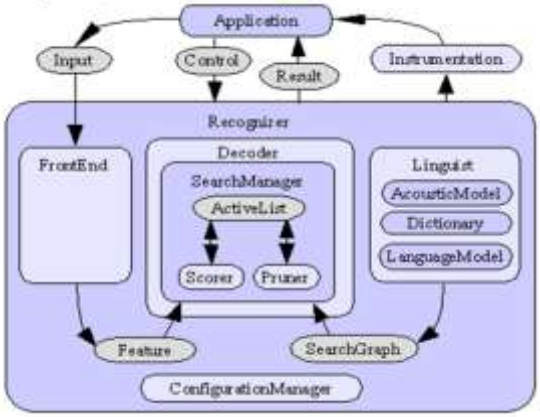

Fig. 3. Sphinx-4 speech recognition system.

ISSN 2349-6193
Impact Factor (PIF): 2.243

IJESMR
International Journal OF Engineering Sciences & Management Research

B. Sphinx-4 Speech Recognition System

We describe the main components of Sphinx-4, and how they work together during the recognition process.
First of all, let's look at the architecture diagram of Sphinx-4showninFig. 3.
When the recognizer starts up, it constructs the front end (which generates features from speech), the decoder, and the linguist (which generates the search graph) according to the configuration specified by the user. These components will in turn construct their own subcomponents. For example, the linguist will construct the acoustic model, the dictionary, and the language model. It will use the knowledge from these three components to construct a search graph that is appropriate for the task. The decoder will construct the search manager, which in turn constructs the scorer, the pruner, and the active list.
Most of these components represent interfaces. The search manager, linguist, acoustic model, dictionary, language model, active list, scorer, pruner, and search graph are all Java interfaces. When the application asks the recognizer to perform recognition, the search manager will ask the scorer to score each token in the active list against the next feature vector obtained from the front end. This gives a new score for each of the active paths. The pruner will then prune the tokens (i.e., active paths) using certain heuristics. Each surviving paths will then be expanded to the next states, where a new token will be created for each next state. The process repeats itself until no more feature vectors can be obtained from the front end for scoring. This usually means that there is no more input speech data. At that point, we look at all paths that have reached the final exit state, and return the highest scoring path as the result to the application.

SYSTEM IMPLEMENTATION

The face verification system is divided into two basic modules: face detection and face recognition. Face detection is a technology to determine the locations and size of a human being face in a digital image. It only detects facial expression and rest all in the image is treated as background and is subtracted from the image. It is a special case of object-class detection or in more general case as face localizer. OpenCV (Open Source Computer Vision) is an android library for real time computer vision. It focuses mainly on real-time image processing. We detect face via camera built in the android based smartphones. User face is detected and captured then the features are extracted via feature extractor and this proposed system generate templates of the detected image, finally this binary information is stored in the SD Card. This process takes place at enrolment time and this stored image is binary processed. This trained image is compared with the new image detected during user authentication process.
Face recognition is a technique to identify a person face from a still image or moving pictures with a given image database of face images.

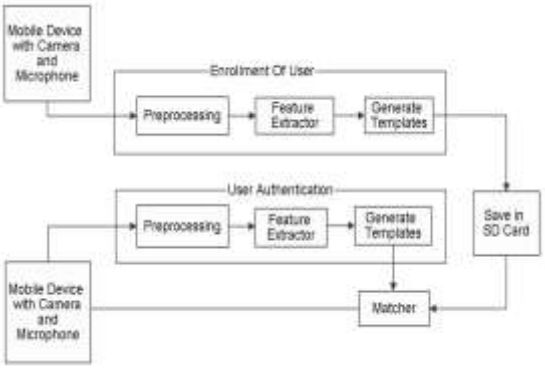

Fig. 4. Block diagram for Implementation of facial and voice verification system.

Face recognition is biometric information of a person. However, face is subject to lots of changes and is more sensitive to environmental changes. Thus, the recognition rate of the face is low than the other biometric information of a person such as fingerprint, voice, iris, ear, palm geometry, retina, etc. There are many methods for face recognition and to increase the recognition rate.
Speaker verification and speaker recognition are complex areas and still hot research subjects, with many different interesting ways of doing the feature extraction. An automatic speaker verification system works based on the premise that a person's speech exhibits characteristics that are unique to the speaker.

ISSN 2349-6193
Impact Factor (PIF): 2.243

IJESMR
International Journal OF Engineering Sciences & Management Research

For our verification system, The 12 acoustic features employed are Pitch, Energy, Pitch difference and Energy difference, Formant (Frequency and bandwidth for first four formants, thus eight features).A threshold is also computed from these12 features for each 60 ms segment of the speech sample. we find the threshold values by calculating the mean, the maximum, the minimum, the range, and the standard deviation for each feature, resulting in 12 × 5 = 60 attributes that are sent to the classifier. Later, during the verification phase, the input speech is matched with the earlier stored models and a decision calculation is made, deciding if the speaker is the claimed or not.

Table 1. OPEN CV Based Video Acquisition and its Description

Title	Description
Object detection Identify colored objects	**Procedure:** (a) The user is presented with a menu, (i) Capture from camera (ii) Load Existing Image (iii) Exit
	(i) When the user selects option 1, the camera device is opened using cv Capture From CAM (0) where 0 is index of camera. **(a)** cv Query Frame function is used to get the latest frames and again the user is provided with 2 options, - Press 'c' to capture image and identifies object from camera frames. - Press 'Esc' to exit. If user presses 'c' button, the frame is captured from the camera and the function Detect Object () is called passing frame as an argument.

A. Sphinx-4 Configuration System

The performance of Sphinx-4 critically depends on how Sphinx-4 is configured to suit with the task. For example, a large vocabulary task needs a different linguist than a small vocabulary task. System has to be configured differently for the two tasks. The configuration mechanism of Sphinx-4,is done with XML-based configuration file.

B. Managing the Sphinx Configuration

The Sphinx-4 configuration manager system has two primary purposes:

1) Determining which components are to be used in the system: The Sphinx-4 system is designed to be extremely flexible. At runtime, just about any component can be replaced with another. For example, in Sphinx-4 the *FrontEnd* component provides acoustic features that are used scored against the acoustic model. Typically, Sphinx-4 is configured with a *FrontEnd* that produces Mel frequency cepstral coefficients (MFCCs),however it is possible to reconfigure Sphinx-4 to use a different *FrontEnd* that, for instance, produces *Perceptual Linear Prediction coefficients* (PLP). The Sphinx-4 configuration manager is used to configure the system in this fashion.

2*)* Determining the detailed configuration of each of these components: The Sphinx-4 system is like most speech recognition systems in that it has a large number of parameters that control how the system functions. For instance, a *beam width* is sometimes used to control the number of active search paths maintained during the speech decoding. A larger value for this beam width can sometimes yield higher recognition accuracy at the expense of longer decodes times. The Sphinx-4 configuration manager is used to configure such parameters.

RESULTS AND DISCUSSIONS

The system proposed is a real-time system. It takes input image through smartphone camera. The captured image is then cropped by the face detection module and saves only the facial information. The images are saved

ISSN 2349-6193
Impact Factor (PIF): 2.243

IJESMR
International Journal OF Engineering Sciences & Management Research

in a sequence of their occurrence time. That is, the face which is detected first is saved first in the SD Card and the next is saved at the next place in the SD Card .The name of the face image is the name given at the time of enrolment. At the time of training the system sequentially takes the training dataset of face images. After creating the dataset the system is trained itself by calculating the face space. This is done by using the principal component analysis algorithm.
The various output of the proposed system are shown below one by one. In figure 5, the face detection is shown and in figure 6, the face recognition is shown

Fig. 5. Single user face detection using haar cascade classifier via camera integrated smartphone.

Fig.6. Displays user name at the time offace recognition.

A voice verification application has been developed and implemented using Sphinx-4 Java API. Although it is difficult to get accurate result, we can show the variations that occur when emotion changes. By using MFCC algorithm feature is extracted from which we can observe how changes occur in pitch, frequency and other features when emotion changes. We have done Frame blocking and windowing steps of MFCC algorithm for a same voice and a same sentence in two different emotions and showed difference in pitch with change in emotion. For our verification system, a threshold is also computed from the training samples. Later, during the verification phase, the input speech is matched with the earlier stored models and a decision calculation is made, deciding if the speaker is the claimed or not.

CONCLUSION

We have proposed a new system with algorithm for identity verification of human faces on a mobile platform that use face and voice characteristics. To ensure the system is robust, we adapt models to the estimated capture conditions and fuse signal modalities, all within the constraints of a consumer grade mobile device. The objective of this paper is to develop a robust and secure verification system for sensitive data that are stored in phone memory. Mobile internet is an obvious example where biometric verification may complement traditional access methods such as passwords. Other potential applications include using biometrics to lock and unlock the phone, and mobile banking transactions.

REFERENCES

1. J. Xiao, S. Baker, I. Matthews, and T. Kanade, “Real-time combined 2D+3D active appearance models,” in Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 535–542, 2004.
2. T.F. Cootes, G.J. Edwards, and C.J. Taylor, “Active appearance models,” IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 23, No. 6, pp. 681–685, Jun. 2001.

ISSN 2349-6193
Impact Factor (PIF): 2.243

IJESMR
International Journal OF Engineering Sciences & Management Research

3. Bhoomika Panda, Debananda Padhi, Kshamamayee Dash, and Prof. Sanghamitra Mohanty, "Use of SVM Classifier & MFCC in Speech Emotion Recognition System,'' International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 2, No. 3, pp. 225-230, March 2012.
4. Zhao Guoshuai , Ye Yanchao , Liu Huqiu & Liu Guoli , 2011, 'The Design and Implementation of Automatic Access Control System Based on the Face Recognition ', International Conference on Mechanic Automation and Control Engineering , vol.5,pp. 4525 - 4528
5. Zhengming Li, LijieXue&Fei Tan, 2010, 'Face Detection in Complex Background Based on Skin Color Features and Improved AdaBoost Algorithms', IEEE International Conference on Progressing Informatics and Computing ,vol. 2 pp. 723 - 727
6. Zhifeng Liu, Jinfeng Jiang, Wentong Yang, Aiping Zhang&Jianhua Wang , 2009, 'Improved Motion Estimation Algorithm Based on ME-Skip ',Computer-Aided Industrial Design & Conceptual Design, IEEE 10th International Conference, vol .6, pp. 2061 - 2064
7. Ziyin Li&Qi Yang , 2012, 'Fast Adaptive Motion Estimation Algorithm', Computer Science and Electronics Engineering, International Conference,vol.3,no.1,pp.656-660
8. Phil Tresadern and Timothy F. Cootes, "Mobile Biometrics: Combined Face and Voice Verification for a Mobile Platform" Pervasive Computing, IEEE Vol.12, No.1, pp 79-87 January 2013.
9. Guo,J.-M. ; Chen-Chi Lin ; Min-Feng Wu ; Che-Hao Chang ;Lee, H.," Complexity Reduced Face Detection Using Probability-Based Face Mask Prefiltering and Pixel-Based Hierarchical-Feature Adaboosting "IEEE Signal Processing Letters, Vol.18,No.8, pp 447 – 450 August 2011.
10. K. Susheel Kumar, Shitala Prasad, Vijay Bhaskar Semwal, R C Tripathi, "Real Time Face Recognition using Adaboost Improved Fast PCA Algorithm," International Journal of Artificial Intelligence & Applications (IJAIA), Vol.2, No.3, pp.45-58,July 2011.