IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 4, April 2016.

Speech and Palm Recognitions: A Review

Er. Akanksha Sambyal[1], Shailza Thakur[2]

Assistant Professor, Computer Science & Engineering Dept., Sri Sai University Palampur, Himachal Pradesh, India[1]

Student, Computer Science and Engineering Department, Sri Sai University Palampur, Himachal Pradesh, India[2]

Abstract: After years of research and development the accuracy remains one of the most important research challenges .In this paper multimodal biometric system for identify verification using two traits i.e. speech signal and palm, which increase the level of security the multimodal biometric systems, are used. In spite of many research efforts spent in trying to create an intelligent machine that can recognise spoken word and palm image and also it's meaning. Our aim in this paper is to provide an overview of basic concept, types, and approaches and also presents what research has been done around for dealing with the problem of palm and speech recognition so as to provide a framework to understand the concept of palm and speech recognition.

Keywords: Biometrics, Speech recognition, Techniques of Speech recognition, Palm recognition, Techniques of Palm recognition.

I. INTRODUCTION

Biometrics systems automatically determine or verify person's identity. Biometrics is classified into two categories: - (1) Physiological biometrics: consisting of fingerprint, palm print, facial, iris, hand geometry, finger knuckles. (2) Behavioural biometrics: consisting of voice, signature and keystrokes. Multimodal biometrics systems [1] take input from single or multiple sensors measuring two or more different modalities of biometric characteristics. The reason to combine different modalities is to improve recognition rate. Multimodal biometric systems utilize more than one physiological or behavioural characteristic for enrolment, verification and identification.

The aim of multimodal biometrics is to reduce one (or) more of the following

- False Accept Rate [FAR].
- False Reject Rate [FRR].
- Failure to enroll rate [FTE].

A Generic multimodal biometric system has four important modules:

- Sensor level: This fusion is applicable if and only if multiple sources represent samples of the single source of sense obtained either using a single sensor or different compatible sensors.
- Feature level: This fusion is combination of different feature sets extracted from multiple sources of senses. Feature sets may be either homogeneous or heterogeneous.
- Match score level: Match score is a measure of the similarity between the input biometric and template biometric feature vectors. Based on the similarity of feature vector and the template, each subsystem calculates its own match score value. These individual scores are finally combined to obtain a total score, which is then passed to the decision module, after which recognition is performed.
- Rank level: This fusion is generally adopted for the identification of the person rather than verification. Rank level fusion provides less information as compare to score level fusion.
- Decision level: This fusion is carried out at decision level when the decisions output by singular or multiple modes of samples are available.

Speech and palm are important modes of senses in human-machine interactions. Human-machine interaction aims to enhance the interactions between users and computers by making computers more interactive and receptive to user's needs. Speech recognition is a biometric technology used to identify a particular individual voice. Palm recognition is a biometric technology which is used to identify the inner-surface of the hand between the wrist and the fingers.

II. RELATED WORK

Er. Amanpreet Kaur and Er. Tarandeep Singh in October, 2010 had compared the two different segmentation [3] techniques. The conventional method of hand labelling speech based on linguistic interpretation of what was spoken with the Automatic Segmentation technique. Finally, the deviation between manual and automatic segmentation had been calculated for the onset and offset values for the syllable boundaries. According to them, Speech Reorganization system requires segmentation of Speech wave form into fundamental acoustic units.

Proposed method was implemented and analysed for different Punjabi speech signals. Results were shown for signal where the boundaries of syllables are marked automatically. It had also been observed that the difference between two (Automatic and Manual) techniques is very much negligible and the boundaries of syllables marked by automatic technique were very much accurate [4].

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 4, April 2016.

A new method of speech segmentation was suggested by Bartosz Ziolko et al. In the research work, wavelet method of speech segmentation is proposed. It is based on the power fluctuations of the wavelet spectrum for a speech signal. The discrete wavelet transform is applied to analyse speech signals, the resulting power spectrum and its derivatives. This information allows locating the boundaries of phonemes. They have also presented the evaluation by comparing this method with hand segmentation.

Sina Akbari Mistani et al proposed a technique which makes use of the multispectral analysis [7] of the hybrid features to improve the performance of the palm print recognition system. David Zhang et al have proposed an online palm print identification system [8]. This system was developed to make authentication possible in the real time also.

Hafiz Imtiaz et all have proposed a novel pre-processing technique for DCT domain palm print recognition [7] in which the task of feature extraction is carried out in local zones using 2 dimensional Discrete Cosine Transform (2D-DCT). A survey of all the palm print recognition systems [2, 8] has also been studied. An automated palm print recognition system [1] evaluated the results in terms of correct recognition rate and verification rate. Correct recognition rate is the percentage of people that can be identified by the system. Verification rate can be calculated by using False Acceptance Rate (FAR), False Rejection Rate (FRR), as well as Equal Error Rate (EER). FAR is the percentage of accepted not genuine claims over the total number of not genuine accesses. FRR is the percentage of rejected genuine claims over the total number of genuine accesses. ERR is the system threshold value when FAR is equal to FRR. For a biometric to work effectively, FAR and FRR must be as low as possible. Total Success Rate (TSR) is the Verification rate of the system. Principal Component Analysis (PCA) [9] is used for dimensionality reduction.

III.OVERVIEW OF SPEECH RECOGNITION

Speech recognition is the ability of a machine or program to identify words and phrases in spoken language and convert them to a machine-readable format. Speech is the most basic, common and efficient form of communication method for people to interact with each other. People are comfortable with speech therefore persons would also like to interact with computers via speech, rather than using primitive interfaces such as keyboards and pointing devices. This can be accomplished by developing an Automatic Speech Recognition (ASR) system which allows a computer to identify the words that a person speaks into a microphone or telephone and convert it into written text.

The process of speech recognition begins with a speaker creating an utterance which consists of the sound waves which is captured by a microphone and converted into electrical signals. These electrical signals are then converted into digital form to make them understandable by the speech-system. Speech signal is then converted into discrete sequence of feature vectors, which is assumed to contain only the relevant information about given utterance that is important for its correct recognition. Finally recognition component finds the best matching the knowledge base, for the incoming feature vectors

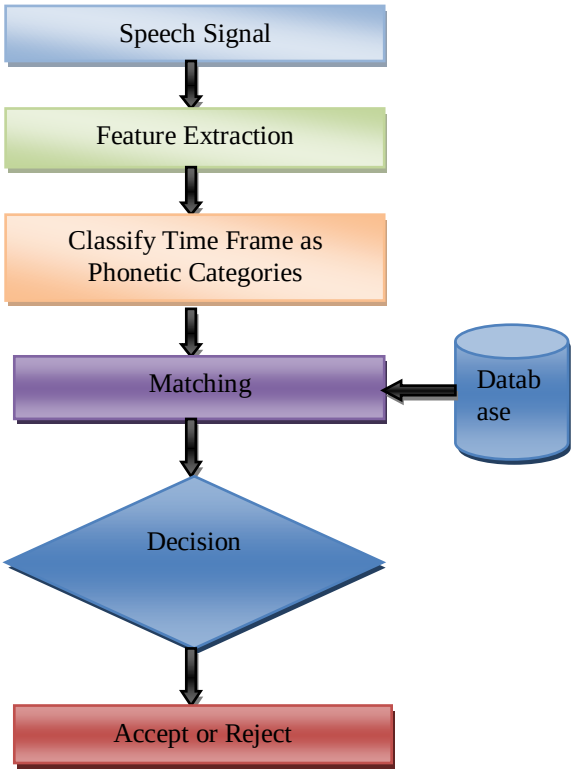

Fig.1: Flow Diagram of Speech Recognition System

Speech can be categorized into three activities:

- Voiced speech: - Vocal cords are tensed and vibrating periodically, resulting in a speech waveform that is quasi-periodic. All the information that is to extract is contained in the voiced part of the speech signal.
- Unvoiced speech: - Vocal cords are not vibrating, resulting in aperiodic or random speech waveform.
- Silence: -When no speech is produced. Voiced and unvoiced region is usually separated by silence region. There is no waveform in the silence region of the speech signal. Hence this part can be easily detected in the signal.

Types of speech utterances:

- Isolated Words:-These systems accept one word at a time.
- Connected Words:-These systems allow speaker to speak slowly and distinctly each word with a short pause i.e. planned speech.
- Continuous Speech:-Continuous speech recognition system allows users to speak almost naturally, while the computer determines its content. Continuous speech recognition system is difficult to develop.
- Spontaneous Speech:-Spontaneous speech is natural that comes suddenly through mouth. Spontaneous speech

may include mispronunciation, false-starts and non words.

Types of speech Speaker Model:
Speech recognition system is classified into three main categories according to their voice as follows:

- Speaker Dependent Models:-Systems that require a user to train the system are according to his or her voice. These systems are usually cheaper, easier to develop and more accurate .But these systems are not flexible as speaker independent systems.
- Speaker Adaptive Models:-Speaker adaptive speech recognition system uses speaker dependent as well as speaker independent models.
- Speaker Independent Models: - Systems that don't require a user to train the system i.e. They are develop to operate any speaker. Implementations of speaker independent systems are difficult and costly. Its Accuracy is less than speaker dependent system.

IV. TECHNIQUES OF SPEECH RECOGNITION

There are three types of Techniques. They are:

- Acoustic phonetic Technique.
- Pattern Recognition Technique.
- Template Matching Technique.
- Statistic based Technique.
- Artificial intelligence Technique.
- Dynamic Time Warping (DTW) Based Technique.
- Neural Network-Based Technique.

A. Acoustic Phonetic Technique

Acoustic phonetic Technique [3] is also known as rule-based or knowledge about variation in speech is hand-coded into a system. It takes set of features from the speech and then train the system to generate set of production rules automatically from the samples. These rules are resulted from the parameters that provide useful information about a classification. It has poor performance due to difficulty to express rules, to improve the system. This Technique identifies individual phonemes, words, sentence structure and/or meaning.

B. Pattern Recognition Technique

This method has two steps i.e. training of speech patterns and recognition of pattern by way of pattern comparison. In the parameter measurement phase (filter bank, LFC, DFT), a sequence of measurements is made on the input signal to define the "test pattern". The unknown test pattern is then compared with each sound reference pattern and a measure of similarity between the test pattern & reference pattern best matches the unknown test pattern based on the similarity scores from the pattern classification phase (dynamic time warping).

C. Template Matching Technique

Template based Technique has a collection of prototypical speech patterns. These patterns are stored as reference patterns representing the dictionary of words. Speech is recognized by matching an unknown spoken utterance with each of these reference templates and selecting the category of the best matching pattern.

This Technique has the benefit of using perfectly accurate word models; but this has the drawback that the pre-recorded templates are fixed Template training and matching become prohibitively expensive or impractical as vocabulary size increases beyond a few hundred words. This method is rather inefficient in terms of both required storage and processing power needed to perform the matching.

D. Statistic based Technique

It can be seen as extension of template based Technique, using some powerful and statistical tools and sometimes seen as anti-linguistic Technique. It collects a large corpus of transcribed speech recording and train the computer to learn the correspondences. At run time, statistical processes are applied to search for all the possible solutions & pick the best one.

E. Artificial Intelligence Recognition Technique

This Technique is a combination of the acoustic phonetic Technique & the pattern recognition Technique. In the AI, an expert system implemented by neural networks is used to classify sounds. The basic idea is to compile and incorporate knowledge from a variety of knowledge sources with the problem at hand.

F. Dynamic Time Warping (DTW) Based Technique

Dynamic Time Warping is an algorithm for measuring similarity between two sequences which may vary in time or speed. DTW is a Technique that allows a computer to find an optimal match between two given sequences with certain limitations. This Technique is useful for isolated word recognition and can be modified to recognize connected words also [3].

G. Neural Network-Based Technique

Neural Network Technique is capable of solving more complicated recognition tasks, but could not perform as excellent as Hidden Markov Model (HMM) when it comes to large vocabularies [4]. They can grip low quality, noisy data and speaker independency. This type of systems can achieve more accuracy than HMM based systems when there is training data and the vocabulary size is limited.

V. OVERVIEW OF PALM RECOGNITION

Palm recognition systems [1] scan the entire palm, while others allow the palm image to be segmented in order to improve performance and reliability. In general terms, reliability and accuracy is improved by searching smaller data sets. Palm systems categories data based upon the location of a friction ridge area.

A palm is the inner-surface of the hand between the wrist and the fingers. A palm print is defined as the prints on a palm, which are mainly composed of the palm lines and ridges. In Palm print recognition, a palm print image obtains by using a scanner or a CCD camera.

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 4, April 2016.

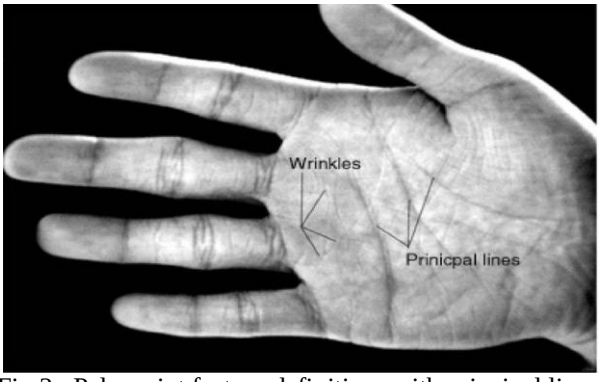

Fig 2: Palm print feature definitions with principal lines and wrinkles.

In Palm recognition, Biometric devices consist of 3 elements:

- Scanner- which scans user's palm print
- Software-converts the palm prints into digital form and compare it with the sampled data.
- System database- stores the palm print (biometric data).

Feature Vectors for all images in the database have been calculated in the feature extraction module, and stored in the form of a text file, called the system database. In the matching module feature vector has been calculated from the query image and compared with the system database. A decision for verification or recognition is taken as per the problem targeted.

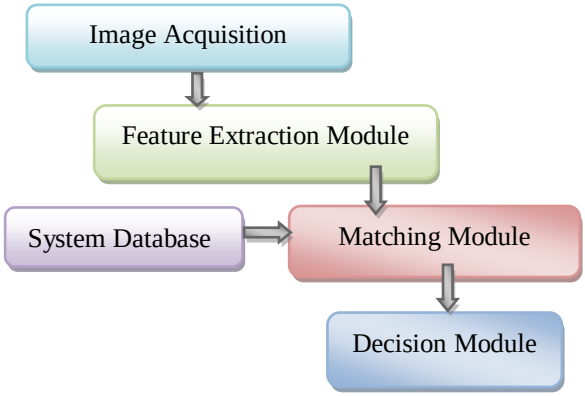

Fig 3: Block Diagram of Palm Recognition.

H. Image Acquisition

In the Image Acquisition, a person placed his/her hand in front of a webcam that captured the image of hands. This setup is relatively costly as compared to the fingerprint sensors. The fingers must be clearly separated from each other in the image in order to obtain a complete hand shape. Background should be a dark one.

I. *Feature Extraction Module*

• *Feature selection*:-A careful study of anatomy of human hand reveals a large number of features useful in palm biometrics. The overall palm area, the elongation indices of the fingers the perimeter, area and circularity of the palm are known to be very distinctive in nature. The ratio of the length and area of the features, also preserve a good amount of information. All these features when combined together form a comprehensive set of 19 features. The extracted features are

- Finger Length
- Finger Width
- Area of the Distal Phalanx
- Length to Width Ratio
- Area of the Palm
- Perimeter of the Palm
- Differences with Centre Ratio

• *Pre-processing*:- The goal of digital image pre-processing is to increase both the accuracy and the interpretability of the digital data during the image processing phase.

In pre-processing the image in jpeg format is converted to a grayscale image of the palm without any deformity such as a missing finger the system expresses its inability to fingers are process the image. It is also critical that the separated from each other. The hand should be placed in a relaxed state with fingers separated from each other. Since features such as length and width which are dependent on the image size and resolution are being used, it is critical that to have uniform size of images.

• *Boundary extraction*: - The image obtained after elimination of noise contains regions of black and white pixels. In order to extract geometric features of the palm it is required that the image contains only edges. It is required to convert regions of white space to an image containing only the boundary of the white pixels. This is achieved by using an edge detection algorithm. The algorithm converts all pixels excluding those at the boundary of black and white regions to black.

• *Locating a set of Reference Points*: - Feature extraction primarily requires that the image be segmented into some relevant regions for reference, which would require knowledge of some special reference points like valleys between the fingers, tips of the fingers etc.

• *Computing the features*: - Once the special reference points are located in the image, a comprehensive set of palm geometry features such as finger length. Finger width, Area of Distal Phalanx, Length to Width ratio etc. is computed.

J. System Database

Feature Vectors are calculated in the feature extraction module, and stored in the form of a text file in the system database. A feature with good discriminating ability should exhibit a large variance between individuals and small variance between samples from the same person.

K. Matching Module

Matching algorithm can be designed to simply return the closest match. In matching, feature is extracted from the given hand to person and compared with the feature vector stored in the database associated with the claimed identity.

DOI 10.17148/IJARCCE.2016.54180

L. Decision Module

The person, whose image is found nearest to the database, would be the recognized person. Decision is made on the bases of distance value obtained from the matching module.. The users are authenticated by the palm print recognition systems. If the user is accepted, then whose palm print match with a palm print present in the database. If the user is not accepted, then the user is rejected. This process of accepting and rejecting the user is done on the basis of extracted features by using the matching algorithm.

VI. TECHNIQUES OF PALM RECOGNITION

There are five techniques. They are:-

M. Line Based Technique

This Technique develops edge detectors or use existing one to extract palm lines [9]. Palm line are matched directly or represented in other format for matching. Euclidian distance is used for matching. First using sober masks magnitude of palm lines are computed. The magnitude are projected in x and y directions to form histograms. They computed first and second order derivative of palm images. These derivatives can be obtained by rotating the masks. The zero crossing of first order derivative is used to identify the edge points and corresponding directions. Second order derivative is used to identify the magnitude of the lines. The weighted sum of local directional magnitude is an element in the feature vector.

N. Statistical Technique

There are two types local and global. The local v transforms images into another domain and divides the transform into several regions such as mean and variance of each small region. The global features like moments centre of gravity and density directly from the whole transformed images.

O. Coding Technique

Palm code uses a single Gabor filter to extract the local phase information of palm print [1].Kong et al. introduced a fusion code method to encode the phase of the filter responses from a bank of Gabor filters with different orientations. A practical palm print recognition algorithm using 2D phase information (i) reduces the registered data size by registering quantized phase information and (ii) deals with nonlinear distortion between palm print images by local block matching using Phase-Only Correlation.

P. Subspace Based Technique

This involves Principal Component Analysis (PCA) [6], Linear-Discriminate Analysis (LDA) and Independent Component Analysis (ICA). The subspace coefficients are considered as features. With these researchers also employs wavelets, DCT and kernel.

Q. Other Technique

Some Technique are difficult to classify because they combine several image processing methods to extract Palm print features such as neural network to make final decision, two dimensional dual-tree complexes transform on pre-processed palm print to decompose the images, phase only correlations etc.

VII. CONCLUSION

In this paper, we have presented a brief discussion of the basic concepts and several existing methods used for palm print recognition system and speech recognition. In single biometric system there is a problem during matching the query data with the sampled data. If data is not matched, then we can't access the system. Multimodal system provides two modes to access the system and overcome the problem of matching. A single biometric system provides less accuracy and speed than multimodal biometric systems. There will be lot of work to do with palm print and speech signal feature extraction methods and matching algorithms for increasing speed and accuracy.

REFERENCES

[1] Jugal Kishore Patel and Sanjay Kumar Dubey," Deployment of Palm Recognition Approach using Image Processing Technique**",** International Journal of Computer Science Issues, Vol. 10, Issue 2, No 2, March 2013

[2] Jobin J., Jiji Joseph, Sandhya Y.A, Soni P. Saji, Deepa P.L.," Palm Biometrics Recognition and Verification System" International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering Vol. 1, Issue 2, ISSN 2278 – 8875, August 2012 .

[3] Ibiyemi ,T.S Aliu S.A, Akintola A.G.," Face and Speech Recognition Fusion in Personal Identification", International Journal of Computer Applications (0975 – 8887) Volume 47– No.23, June 2012.

[4] S. Nishi, and S. Parminder, "Automatic Segmentation of Wave File", International Journal of Computer Science & Communication Vol. 1, No. 2, pp. 267-270, July-December 2010.

[5] Santosh K.Gaikwad, Bharti W.Gawali, Pravin Yannawar, "A Review on Speech Recognition Technique" International Journal of Computer Applications, Vol 10– No.3, November 2010.

[6] N.Uma Maheswari, A.P.Kabilan3, R.Venkatesh," Speech Recognition System Based On Phonemes Using Neural Networks", IJCSNS International Journal of Computer Science and Network Security, VOL.9 No.7, July 2009.

[7] Saurabh Parashar, Anand Vardha, C.Patvardhan, Prem Kumar Kalra, "Design and Implementation of a Robust Palm Biometrics Recognition and Verification System." Sixth Indian Conference on Computer Vision, Graphics & Image Processing, pp. 543-550, 2008.

[8] T. Andrew, T. Connie, N. David, "Remarks on BioHash and Its Mathematical Foundation," Information Processing Letters, vol. 100(4),pp.145-150,2006.

[9] Connie, T. Andrew, K.O. Goh, "An Automated Palm print Recognition System," Image and Vision Computing, vol. 23 (5), pp. 501-515, 2005.