International Journal of Scientific Research & Engineering Trends
Volume 6, Issue 3, July-Aug-2020, ISSN (Online): 2395-566X

A Review on Person Recognition using Support Vector Machine and Local Binary Pattern

Priyanka Bhate
M.Tech. Scholar
Dept.of CSE
Vindhya Institute of Technology & Science
Indore (M.P)
priyankabhate1617@gmail.com

Ashish Tiwari
Assistant Professor & Head
Dept.of CSE
Vindhya Institute of Technology & Science
Indore (M.P)
ashishtiwari205@gmail.com

Abstract – **This paper discusses the concept of speech recognition with deep learning methods. Introduction of speech recognition, deep learning and deep learning methods is discussed in this review paper. Models of deep learning that are used in speech recognition is also described in this paper. This paper defines the related work on speech recognition using deep learning methods and about the sphinx, software allow the implementation of speech recognition in java language. The main motive of this review is to define the use of sphinx and eclipse to recognize speech. we will proposed We proposed new technique for human identification using fusion of both face and speech which can substantially improve the rate of recognition as compared to the single biometric identification for security system development. Our system using Viola Jones Algorithm for face detection. The proposed system uses Local Binary Pattern (LBP) as feature extraction techniques which calculate the local features. The Extracted features given as input SVM classifier to used to recognize the person and then display the result. This new system can be applied in various different fields such as identity verification and other potential commercial applications.**

Keywords– **SVM, Speech Recognition. Feature Extraction, LBP.**

I. INTRODUCTION

From the early part of the previous century, there has been curiosity in making computers do what only humans could perceive, like recognizing speech, understanding natural language, processing images, etc. Speech being the primary, most efficient mode of communication between human beings, research in speech recognition has received much enthusiasm for the past five decades right from the advent of artificial intelligence.

Many reasons can be attributed to this enthusiasm ranging from mere technological curiosity to the desire of automating tasks using machines and providing more natural machine interfaces. The study on speech analysis dates back to the beginning of the nineteenth century, when Homer Dudley of Bell Laboratories made the first proposal for a speech analysis and synthesis system in 1930s [1,2]. In 1952, an isolated digit recognizer for a single speaker was built by Davis et al. of Bell Laboratories [3], followed by a system that could recognize 10 syllables of a single speaker proposed by Olson and Belar et al. [4]. A significant achievement occurred in the year 1959 when a phoneme recognizer was developed to recognize four vowels and nine consonants utilizing statistical information about phoneme sequences in English [5]. This marked the first use of statistical syntax in speech recognition. A precious technique that becomes popular in the 1970s is dynamic programming for automatic speech recognition (ASR). A technique generally known as dynamic time warping was first suggested by Vintsyuk et al. [6]. At the same time, at Bell Laboratories, the focus was on the creation of an automatic speech transcription system, which is speaker independent and can handle the acoustic variability arising in speech from different speakers with varying regional accents [7].

This was to fulfill the goal of providing telecommunication services to the people, including voice dialing and command-based automation of phone calls. Another important technique in Bell's approach to ASR is the concept of keyword spotting which attempts to detect only prescribed words or phrases of particular significance in an utterance and neglects the other nonessential portions [8]. This is to accommodate speakers who often prefer to speak natural sentences rather than rigid common words. Conventional authentication mechanisms lose favour in security applications as biometric identification systems take on the lead. Voice is considered to be an important biometric and multimodal recognition systems are springing up to improve the robustness of authenticity as discussed in [9].

Speech recognition can be extended to recognize speakers, exploiting the information present in the speech and various methods including exploiting from the excitation source are reviewed and presented The above-mentioned techniques had a profound impact on the advancements in ASR for the past three decades. In this paper, a review of the various machine learning (ML) techniques for ASR is presented. Humans can recognize speech in different speaking styles and speaking rates (Normal, Fast and Slow), but for the computers it is a difficult task to recognize the speech. Speech recognizers implemented in computers perform relatively poorly when speech rate is very fast or very slow.

In order to improve computer performance, several researchers have proposed that measuring speech rate prior to speech recognition will result in higher success rates of automatic speech recognizers and several ways to automatically measure speech rate in terms of phones or syllables per time unit have been put forward. [7]. Many factors such as stutters, slips of the tongue, interruptions, hesitations, lengthening, filled pauses and laughter affect speech rate [8]. Timing and acoustic realization of syllables or phonemes are frequently affected when the rate of speech is fast or slow. Phone-by-phone length stretching and sentence-by-sentence length stretching are used for normalization [9].

SVM for Continuous Speech Recognition: Continuous speech recognition is a more complex task when compared with isolated recognition, as it poses two major problems: temporal position of words or the number of words in the speech utterance is unknown and the size of the speech databases tends to be larger for continuous recognition tasks and, consequently, the size turns out to be larger than the maximum number of training examples, an SVM can handle. HMM/ SVM systems for continuous recognition have been proposed analogous to HMM/MLP systems, where the phonetic level assignments generated by the HMMs are used by the SVM to classify the phonemes [24,25].

Since each segment may have a different duration, they need to be converted into fixed-length vectors using any one of the methods highlighted in the previous section. The authors suggested dividing the segment into three regions according to a pre-established proportion. Then, the vectors in every region were averaged and concatenated together. However, this method fails to exploit the generalization capabilities of SVMs. Moreover, the efficiency of the system is limited by the errors committed during the segmentation phase. Speech recognition means recognizing the speech and converting it into readable form or text. It is the ability of a machine or program to receive and interpret dictation, or to understand and carry out spoken commands. Speech recognition applications includes voice user interfaces such as voice dialing, call routine, search, simple data entry like entering credit card number etc.

II. LITERATURE SURVEY

M.A.Anusuya et.al. presents a brief survey on Automatic Speech Recognition and discusses the major themes and advances made in the past 60 years of research, so as to provide a technological perspective and an appreciation of the fundamental progress that has been accomplished in this important area of speech communication. The design of Speech Recognition system requires careful attentions to the following issues: Definition of various types of speech classes, speech representation, feature extraction techniques, speech classifiers, data base and performance evaluation. The objective of this review paper is to summarize and compare some of the well known methods used in various stages of speech recognition system and identify research topic and applications which are at the forefront of this exciting and challenging field.

Santosh K.Gaikwad et.al. The Speech is most prominent & primary mode of Communication among of human being. The communication among human computer interaction is called human computer interface. Speech has potential of being important mode of interaction with computer .This paper gives an overview of major technological perspective and appreciation of the fundamental progress of speech recognition and also gives overview technique developed in each stage of speech recognition. This paper helps in choosing the technique along with their relative merits & demerits. A comparative study of different technique is done as per stages. This paper is concludes with the decision on feature direction for developing technique in human computer interface system using Marathi Language.

Shanthi Therese, Chelpa Lingam et.al. Says that speech has evolved as a primary form of communication between humans. The advent of digital technology, gave us highly versatile digital processors with high speed, low cost and high power which enable researchers to transform the analog speech signals in to digital speech signals that can be scientifically studied. Achieving higher recognition accuracy, low word error rate and addressing the issues of sources of variability are the major considerations for developing an efficient Automatic Speech Recognition system. In speech recognition, feature extraction requires much attention because recognition performance depends heavily on this phase. In this paper, an effort has been made to highlight the progress made so far in the feature extraction phase of speech recognition system and an overview of technological perspective of an Automatic Speech Recognition system are discussed.

Sanjib Das et.al. Presents a brief survey on speech is the primary and the most convenient means of communication between people. The communication among human computer interaction is called human computer interface. Speech has potential of being important mode of interaction with computer. This paper gives an overview of major technological perspective and appreciation of the fundamental progress of speech recognition and also gives overview technique developed in each stage of speech recognition. This paper helps in choosing the technique along with their relative merits and demerits. A comparative study of different technique is done as per stages. This paper concludes with the decision on feature direction for developing technique in human computer interface system in different mother tongue and it also discusses the various techniques used in each step of a speech recognition process and attempts to analyze an approach for designing an efficient system for speech recognition. The objective of this review paper is to summarize and compare different speech recognition systems and identify research topics and applications which are at the forefront of this exciting and challenging field.

Nidhi Desai et.al. survey presents speech is the most natural form of human communication and speech processing has been one of the most inspiring expanses of signal processing. Speech recognition is the process of automatically recognizing the spoken words of person based on information in speech signal. Automatic Speech Recognition (ASR) system takes a human speech utterance as an input and requites a string of words as output. This paper introduce a brief survey on Automatic Speech Recognition and discuss the major subjects and improvements made in the past 60 years of research, that provides technological outlook and a respect of the fundamental achievement that has been accomplished in this important area of speech communication. Definition of various types of speech classes, feature extraction techniques, speech classifiers and performance evaluation are issues that require attention in designing of speech recognition system. The objective of this review paper is to summarize some of the well known methods used in several stage of speech recognition system.

Guillaume Gravier, Ashutosh Garg et.al. survey presents Visual speech information from the speaker's mouth region has been successfully shown to improve noise robustness of automatic speech recognizers, thus promising to extend their usability into the human computer interface. In this paper, we review the main components of audio-visual automatic speech recognition and present novel contributions in two main areas: First, the visual front end design, based on a cascade of linear image transforms of an appropriate video region-of-interest, and subsequently, audio-visual speech integration. On the later topic, we discuss new work on feature and decision fusion combination, the modeling of audio-visual speech asynchrony, and incorporating modality reliability estimates to the bimodal recognition process. We also briefly touch upon the issue of audiovisual speaker adaptation. We apply our algorithms to three multi-subject bimodal databases, ranging from small- to large vocabulary recognition tasks, recorded at both visually controlled and challenging environments. Our experiments demonstrate that the visual modality improves automatic speech recognition over all conditions and data considered, however less so for visually challenging environments and large vocabulary tasks.

Li Deng and John C. Platt survey presents that deep learning systems have dramatically improved the accuracy of speech recognition, and various deep architectures and learning methods have been developed with distinct strengths and weaknesses in recent years. How can ensemble learning be applied to these varying deep learning systems to achieve greater recognition accuracy is the focus of this paper. We develop and report linear and log-linear stacking methods for ensemble learning with applications specifically to speech-class posterior probabilities as computed by the convoluional, recurrent, and fully-connected deep neural networks. Convex optimization problems are formulated and solved, with analytical formulas derived for training the ensemble-learning parameters. Experimental results demonstrate a significant increase in phone recognition accuracy after stacking the deep learning subsystems that use different mechanisms for computing high-level, hierarchical features from the raw acoustic signals in speech.

Li Deng, Jinyu et.al. Survey describe that deep learning is becoming a mainstream technology for speech recognition at industrial scale. In this paper, we provide an overview of the work by Microsoft speech researchers since 2009 in this area, focusing on more recent advances which shed light to the basic capabilities and limitations of the current deep learning technology. We organize this overview along the feature-domain and model-domain dimensions according to the conventional approach to analyzing speech systems. Selected experimental results, including speech recognition and related applications such as spoken dialogue and language modeling, are presented to demonstrate and analyze the strengths and weaknesses of the techniques described in the paper. Potential improvement of these techniques and future research directions are discussed.

Pato, J.N. et.al: The face provides complex information, including age, gender, race, identity, personality, intent, and emotion. In addition, the ability to speak has a significant impact on facial expressions as well. All of these aspects govern human relationships at all levels. While we can take this complex mix of ideas to avoid costly causes, the existence of all these sources of life makes the process more difficult for the machine-minded. For example, as emotions move away from neutral emotions, the functioning of normal affective states decreases. Likewise, effective recognition systems should be firmly rooted in the changes in the personality structure of individuals. In the context of facial recognition, previous research has suggested ways to compensate for headaches and subject differences. However, it is not known how to reduce the variability of verbal messages during verbal interaction. In honest conversation, the process of pronunciation affects the facial expressions. Recovery of emotions from facial expressions requires the separation of language and emotional information. Functioning and compensating for the content of the basic vocabulary improved the comprehension of the emotion.

However, this requires transcriptional and phoneme information, which is not available in many types of applications. This study uses an asymmetric bilinear factor model to soften language and emotion (if not provided) The emotional evaluation of the IEMOCAP database shows that this approach can separate these items from face to face, resulting in significant performance improvements. The improvement is similar to the known phonetic translation of truth. Similarly, experiments conducted on the SEMAINE site using image-based methods have proven the technology's success in real-life situations. The IEMOCAP corpus is a collection of media information examining the human interaction of written comments with translators (5 men and 5 women).

Over the course of five sessions, a film and actress performed a series of screenwriting and editing. Choosing the circumstances and circumstances for which it is intended to make an impression (for example, losing an airport and enrolling in a school). Dialogues are divided into circuits, which are legally translated. Phoneme and boundary words are obtained by forced alignment algorithm. Transition is emphasized by three reviewers using the following tags, which use the following hash tags: anger, happiness, sadness, shock, discomfort, frustration, fear, jealousy, neutrality, etc.

Muthukrishnan R et.al Hands and shoulders are widely used biometric. In this work, these tools are used to create the feature height. Haar ports are known for their difficulty in reducing them. In both cases, Haar wavelet technology is used for feature extraction. This work introduces a new approach to layer layer physics and testing. This layered look makes for a better display than the usual features. The proposed algorithm optimizes each model's performance and efficiency. This is evident from the simulation results. Due to its simplicity, the ongoing research conducted in this field (multimodal) has focused on biometric publishing in terms of height or category matching. However, large scale biometric applications still require improved performance. It takes more emphasis and time than just checking. Therefore, a more specific biometric system should be implemented to not only achieve the desired improvement, but also reduce the execution time. Common classes used in different biological features are vector machines (SVMs) with different kernels (especially Gaussian and polynomial), classes based on Gaussian mixed models, neural networks and multilayer perceptron. Most of them offer great performance improvements, but the results are entirely dependent on the available data sets. High-speed learner models based on SVM are used in the work presented to assess the performance of the proposed system. The idea behind the SVM was to use a kernel trick to map vector entries to high-speed data streams, and then create a grid mapping within that space to to separate the series of files with the endpoint. In general, the number of modern mining technologies implemented by modern systems is often the same as the number of biometric technologies considered.

The use of a single mining technique to exploit the features of the two participants will further strengthen the design system. This research effort is aimed at reducing false rejection, false recognition and training and testing sessions to obtain accurate information. The higher the quality of biological samples collected in various packaging materials, the higher the reliability of the signal. For the same biometric sample, the better the method is accepted, the higher the confidence level. Therefore, this paper proposes a biometric algorithm that utilizes the quality of a biometric sample and the confidence of a recognized expert (called PSVM). First, the sample code for punctuation and memory consolidation is derived from sample quality and expert confidence, and then the results of the exam results. XM2VTS data were used to compare HTER, Bayesian, FLD, MLP, Mean method and SVM of PSV. Experimental results show that PSVM fusion algorithm has lower HTER.

III. PROPOSED APPROACH

In this process, we propose a face tracking algorithm with temporal-spatial information and trajectory of confidence. The whole process is divided into Video and Speech association. Trajectories with high confidence are associated with the detection result of the current frame during local association, whereas trajectories with low confidence are associated with the detection results of the current frame are not matched during global association. We determine the association results using a combined model of digital image processing and digital signal processing. The major steps carried out are, Detection and recognition. Face detection is a computer technology being used in a variety of applications that identifies human faces in digital images.

Face detection also refers to the psychological process by which humans locate and attend to faces in a visual scene.In face based systems this can be in the form of a change in the illumination direction and/or face pose variations. Multi-modal systems use more than one biometric at the same time. In this face detection process is implemented the viola jones is used to detect the face region and the detected region will extracted. In the feature extraction process, we can implement the LBP for pattern extraction in image, and MFCC, Energy features extraction in the speech signal. We implement the Multi SVM Classifier is used to recognize the person and then display the result.

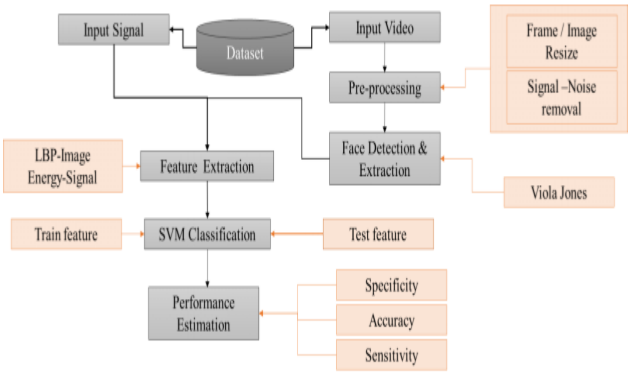

Fig .1. proposed flow chart.

Face Detection & Extraction: Face detection can be regarded as a specific case of object-class detection. In object-class detection, the task is to find the locations and sizes of all objects in an image that belong to a given class. Examples include upper torsos, pedestrians, and cars. Face-detection algorithms focus on the detection of frontal human faces. It is analogous to image detection in which the image of a person is matched bit by bit. Image matches with the image stores in database. Any facial feature changes in the database will invalidate the matching process.A reliable face-detection approach based on the genetic algorithm and the eigen-face technique: Firstly, the possible human eye regions are detected by testing all the valley regions in the gray-level image.

Then the genetic algorithm is used to generate all the possible face regions which include the eyebrows, the iris, the nostril and the mouth corners. Each possible face candidate is normalized to reduce both the lightning effect, which is caused by uneven illumination; and the shirring effect, which is due to head movement. The fitness value of each candidate is measured based on its projection on the eigen-faces. After a number of iterations, all the face candidates with a high fitness value are selected for further verification. The Viola-Jones algorithm is a widely used mechanism for object detection. The main property of this algorithm is that training is slow, but detection is fast. This algorithm uses Haar basis feature filters, so it does not use multiplications. The efficiency of the Viola-Jones algorithm can be significantly increased by first generating the integral image.

Feature Extraction : In machine learning, pattern recognition and in image processing, feature extraction starts from an initial set of measured data and builds derived values (features) intended to be informative and non-redundant, facilitating the subsequent learning and generalization steps, and in some cases leading to better human interpretations. Feature extraction is a dimensionality reduction process, where an initial set of raw variables is reduced to more manageable groups (features) for processing, while still accurately and completely describing the original data set. When the input data to an algorithm is too large to be processed and it is suspected to be redundant (e.g. the same measurement in both feet and meters, or the repetitiveness of images presented as pixels), then it can be transformed into a reduced set of features (also named a feature vector). Determining a subset of the initial features is called feature selection.

IV. CONCLUSION

Speech is basic mode of communication between human beings, so a feasible interface is required to connect human with machines. Although this field has gained a wide approval to automate the services and applications but there are several parameters which affect the accuracy and efficiency of speech recognition system. The most of speech variability involves speech rate, environmental conditions, channel and context of utterance. Robustness of speech system depends on some stable parameters/ features of speech signal. To enhance the power of speech recognition system, it is required to design speech recognizers in local languages. Multilingual is new evolving field in area of speech recognition. There is a lot of development and research in the field of foreign languages but to enhance its power and utility for native people, it's essential to use this technology in native languages.A comprehensive review of common machine learning techniques, support vector machines will be employed in ASR. A thorough review on the recent developments in deep learning which has provided significant improvements in ASR performance,

REFERENCES

[1]. Li Deng, Jinyu Li, Jui-Ting Huang, Kaisheng Yao, Dong Yu, Frank SeideMichael L. Seltzer, Geoff Zweig, Xiaodong He, Jason Williams, Yifan Gong, and Alex Acero Microsoft Corporation, One Microsoft Way, Redmond, WA 98052, USA 2009

[2]. M.A.Anusuya and S.K.Katti ,Department of Computer Science and Engineering,Sri Jayachamarajendra College of Engineering, Mysore, India, (IJCSIS) International Journal of Computer Science and Inform4 ation Security,2009.

[3]. Shanthi Therese ,Chelpa Lingam, International Journal of Scientific Engineering and Technology ,

June 2013.,Review of Feature Extraction Techniques in Automatic Speech Recognition.

[4]. Speech Recognition Technique: A Review Sanjib Das Department of Computer Science, Sukanta Mahavidyalaya, (University of North Bengal), India, International Journal of Engineering Research and Applications (IJERA) MayJun 2012.

[5]. Nidhi Desai1, Prof.Kinnal Dhameliya2, Prof.Vijayendra Desai3, International Journal of Emerging Technology and Advanced Engineering , December 2013, Feature Extraction and Classification Techniques for Speech Recognition: A Review.

[6]. Li Deng and John C. Platt, Microsoft Research, One Microsoft Way, Redmond, WA, USA, November 2010, Ensemble Deep Learning for Speech Recognition.

[7]. Santosh K.Gaikwad, Dr.Babasaheb Ambedkar Marathwada, Bharti W.Gawali, 2011, A Review on Speech Recognition Technique.

[8]. Samy Bengio and Georg Heigold, Google Inc, Mountain View, CA, USA, feb. 2007, Word Embeddings for Speech Recognition. Rubi, International Journal of Computer Science and Mobile Computing, Vol.4 Issue.5, May- 2015, pg. 1017-1024 © 2015, IJCSMC All Rights Reserved 1024

[9]. Audio-Visual Speech Gerasimos Potamianos, Member, IEEE, Chalapathy Neti, Member, IEEE, Guillaume Gravier,, Ashutosh Garg, Student Member, IEEE, and Andrew W. Senior, Member, IEEE 2006, Recent Advances in the Automatic Recognition.

[10]. 10.Dandan Mo,December 4, 2012, A survey on deep learning: one small step toward AI. 11. Aalto University publication series, Foundations and Advances in Deep Learning, Kyunghyun Cho, 2014.

[11]. Abboud, A. J., Sellahewa, H. and Jassim, S. A. “Quality approach for adaptive face recognition”, in Proc. Mobile Multimedia/Image Processing Security, and Applications, SPIE Vol. 7351, 73510 N, 2009.

[12]. Aloysius G., “Efficient High Dimension Data Clustering using ConstraintPartitioning KMeans Algorithm,” the International Arab Journal of Information Technology, Vol. 10, No. 5, pp. 467-476, 2013.

[13]. Alsaade.F and Zahrani.M, “Enhancement of Multimodal Biometric Verification Using a Combination of Fusion Methods”,5th International Conference: Sciences of Electronic, Technologies of Information and Telecommunications March 22-26, 2009.

[14]. Amoli.G, Thapliyal.N, Sethi.N: Iris Preprocessing. International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 2, No. 6, pp. 301-304, 2012.

[15]. Ang.R. Safavi-Naini.R, McAven.L:. Cancelable Key-based Fingerprint Templates. In C. Boyd and J. Gonzalez Nieto (Eds.), Australasian Conference on Information Security and Privacy, pp. 242-252, 2005.