A Novel Technique for Speech Identification and Recognition using Row and Column Features (RCF)

G.Balakrishnan*
Ph.D Research Scholar & Associate Professor, Department of Computer Science,
Navarasam College of Arts & Science for Women, Erode
Email: balakrish1972@gmail.com
Dr.S.Pannirselvam
Research Supervisor & Head
Department of Computer Science, Erode Arts & Science College (Autonomous), Erode, Tamil Nadu, India.
Email: pannirselvam08@gmail.com

Abstract— Today, image processing penetrates into various fields, but till it is struggling in identification and recognition issues. Speech recognition is developed into a very active research area specializing on how to extract and recognize within images. The text based Speech identification and recognition is widely used biometric application for security and identification concern. The various methods have been proposed for speech identification and recognition each method has advantages and drawbacks. The complexity in identification and recognition, other issues affects performance of existing system makes insufficient. In this paper presents speech identification and recognition on full image and on Row suggest of an image. In each of the methods, effect of different quantity of coefficients of transformed picture is determined. The Row and Column Feature (RCF) vector are calculated separately and stored. The feature is generated and matching is done by Euclidean distance classification is used to measure a distance between diagnosed speech. The experimental result shows that RCF provides better recognition rate when compared with the existing methods.
Keywords— DCT, WALSH, HAAR, RCF.

1. INTRODUCTION

Security protection has become an exceedingly vital problem due to widespread use of Net technology as well as because of multi-user applications. Identifying customers and granting get admission to only to those users who are authorized is a key to provide security. Users can be recognized the use of numerous strategies and their combinations. Because the generation is getting advanced, extra state-of-the-art approaches are being used to satisfy the want of safety. Speech identity problem may be further labeled as textual content based and text independent speech identity based totally on relevance to speech contents. Text dependent speech identity calls for the speech pronouncing precisely the enrolled or the given password/speech.

Textual content impartial speech identity is a system of verifying the identity without constraint on the speech content material. Speech identification assignment also can be categorized into closed set and open set speech identity. In closed set hassle, from N acknowledged audio system, the Speech whose reference template has the maximum diploma of similarity with the template of input speech sample of unknown Speech is received. This unknown speech is assumed to be one of the given set of speech. As a result in closed set problem, system makes a compelled selection by selecting the best matching speech from the speech database. In the open set, text structured speech identity matching reference template for an unknown audio system speech pattern may not exist.

2. LITERATURE SURVEY

Speech identity trouble essentially includes characteristic extraction level and pattern class stage. In literature there are many strategies to be had for Speech identity process based totally on various processes for feature extraction.

Davis [1] proposed one of the famous procedures for feature extraction is the Mel Frequency Cepstrum Coefficients (MFCC). The MFCC parameter as by means of describes the power distribution of speech sign in a frequency area.

Wang Yutai et.al. [2] proposed a Speech popularity device based on dynamic MFCC parameters. This approach combines the Speech data received by MFCC with the pitch to dynamically construct a fixed of the Mel-filters. Those Mel-filters are in addition used to extract the dynamic MFCC parameters which constitute characteristics of speech identity.

Sleit et al. [3] proposed a histogram primarily based technique turned into by way uses a reduced set of functions generated using MFCC method. For those features, histograms are created the use of predefined c programming language length. Histograms are generated first for all records in function set for each Speech and then for each characteristic column in feature set of every Speech.

Every other extensively used technique for feature extraction is located of linear Prediction Coefficients (LPC). LPCs capture the facts about brief time spectral envelope of speech. LPCs constitute critical speech traits inclusive of formant speech frequency and bandwidth [4].

Vector Quantization (VQ) is yet another technique of function extraction based totally Speech popularity structures every Speech is characterized with numerous prototypes called code vectors [5].

Pati et al. [6] developed Speech recognition based totally on non-parametric vector quantization. Speech is produced due to excitation of vocal tract. In this technique, excitation records may be captured using LP analysis of speech signal and is known as LP residual. This LP residual is in addition subjected to nonparametric Vector Quantization to generate codebooks of sufficiently massive length.

3. EXISTING METHODOLOGY

3.1 Discrete Cosine Transform (DCT)

DCT image is split into blocks. However it gets rid of correlation throughout the bounds and subsequently consequences in blockading artifacts. This disadvantage may be averted by means of the use of wavelet transforms. Its extremely good strength compaction assets have made wavelets extra famous in current years. Extra energy compaction gives higher compression ratio**.**

3.1 Walsh Transform

It is non-sinusoidal orthogonal transform that decomposes a signal into a set of orthogonal square waveforms called Walsh capabilities. The transformation has no multipliers and is real due to the fact the amplitude of Walsh features has best two values +1 or - 1. Walsh functions are square or rectangular waveforms with values of -1 or +1. An essential function of Walsh capabilities is sequenced that's decided from the wide variety of 0-crossings per unit time interval. Every Walsh function has a completely unique sequence price.

3.3 Haar Transform

This sequence becomes proposed in 1909 by means of Alfred Haar. Haar used these features to provide an example of a countable orthonormal machine for the distance of square-integrable functions at the real line. The Haar remodel is derived from the Haar matrix.

3.3 PROPOSED METHODOLOGY

The first step within the Speech identification system to transform the speech signal into a wave. A wave format is time-various spectral illustration that indicates how the spectral density of a sign varies with time.

Text speech is commonly created in one of two methods: approximated as a clear out financial institution that consequences from a chain or calculated from the time signal using the short-time Fourier rework. Increasing a wave the use of sampled records, within the time domain is damaged up into portion, which normally overlap and Fourier transformed to calculate the importance of the frequency spectrum for each portion. The speech sign is first divided into frames is arranged column smart to form a matrix. Divide the every frame samples with an overlap of 25% between consecutive frames. These frames are then arranged column quick to form a matrix. The feature is then plotted as the squared magnitude of this column matrix.

Each transform is carried out on full image and from the feature vectors obtained, one of a different numbers of coefficients have been used to pick out Speech. Second, transform is carried out to row mean of an image to get the function vector of an image. From this feature vector again identification rate is acquired for various portions selected from the feature vector. Speech has been used as trainee images and testing speech images.

In this approach, transformation method has been applied on complete image to attain feature vector of image. Further it decided on partial feature vectors, identification rate changed into received. This option of function vector is primarily based on the wide variety of rows and columns that we selected from the characteristic vector of image. For these exclusive sizes, identity price changed into acquired. The row mean of these image are calculated and then transformation strategies have been carried out to them to form feature vectors of images and also images have been divided into N equal and non-overlapping blocks. Row mean of those blocks became calculated to get characteristic vectors of images.

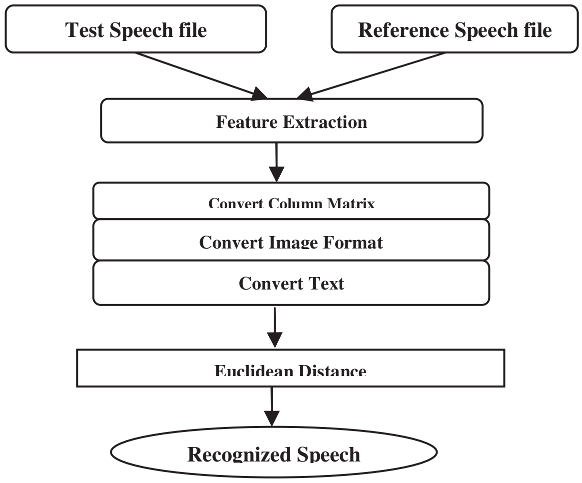

Fig.1 Process Flow

3.3.2 FEATURE VECTOR EXTRACTION

The feature vectors of all of the reference speech samples are stored in the database inside the segment. The matching segment, the check sample this is to be diagnosed is taken and similarly processed as within the training phase to form the characteristic vector. The saved characteristic vector which gives the minimal Euclidean distance with the input pattern function vector is said as the Speech identified. The process for feature vector extraction is Column transform is implemented. The speech sign after which suggest of the absolute values of the rows of the remodel matrix is then calculated. Those row approach form column vector paperwork the characteristic vector for the speech sample are calculated for extraordinary values of n and saved within the database.

3.3.2.1 WAVE File Format

Waveform Audio File Format (WAVE) is an application of RIFF or Resource Interchange File Format which stores audio bit streams in “amy”. WAVE encodes the sound in Linear Pulse Code Modulation format. Sound is basically a pressure wave or mechanical energy having pressure variance in an elastic medium. The variance propagates as compression and rarefaction wherein compression occurs when pressure is higher than the ambient pressure and rarefaction occurs when the pressure of the propagating wave is less than the ambient pressure. Exactly in the same manner a WAVE file just represents the sampled sound waves. In this work using an “amy.wav” wave file to show the proposed algorithm of encrypting the sound file in various image formats. As already mentioned a wave file consists of positive and negative values over its entire range of samples. Here for simplicity will using only the samples having positive values.

3.3.2.2 Image Formats

Digital image formats are means of storing digital images in either uncompressed, compressed and vector formats. On rasterization an image is converted into a grid of pixels. In lossless compression the entire digital data is preserved during compression thus preserving image quality. In lossy compressions, the digital data preservation takes place by compromising image quality. Here discussed only for JPEG formats and these are the very formats in which the wave files.

3.3.2.3 Data of wave file in column matrix

The wave file with graphical representation is provided with the sampling length of this tone and as discussed above using only those samples which have positive values. MATLAB code which fetches the wave file using 'wavread' function. Amplitude values are obtained in the range of 0 and +1. It is to be noted that the variable D is basically a column vector.

3.3.2.4 Convert column matrix into M x N matrix

A grayscale image of M by N pixels is represented in MATLAB as an M X N matrix having "double" data type wherein each element of the matrix denotes a pixel within an intensity of 0 and 1. It is to be noted that the variable D is a column matrix with "double" type and intensity within 0 and 1. So to convert variable D in an image format have to transform D into a 1000 X 2000 matrix.

3.3.2.5 Convert matrix into Image File

To convert matrix A into JPEG formats using MATLAB function called "imwrite". It stores matrix A in the file path mentioned and also save column matrix D in a new wave file using "wavwrite" function. It clearly describes that JPEG stores the wave file.

3.3.2.6 Convert Image Matrix into Column Matrix

The above function is used to save X column vector in the given 'filename' with a desired frequency 'FS'. The column vector X is obtained by converting image matrix of double precision into column matrix.

3.4 PROPOSED ALGORITHM

The entire retrieval procedure with the orientation features is presented as simple algorithms hereunder using MATLAB. In order to identify and recognition of feature based images from the databases are followed.

3.4.1 Algorithm – I

// Transformation on full image //
Begin
Step 1: *Read an image from the image database (IDB) of size M×N (256 X 256).*
Step 2: *Calculate Row Mean of an image.*
Step 3: *Perform* ***procedure training_feature ( )***
Step 3: *Perform* ***procedure*** **testing**_***feature ( )***
Step 4: *Repeat* ***Step1*** *through* ***Step 3*** *for all the images in IDB.*
Step 5: *Establish feature database set.*
End

Perform procedure training_feature ( )
{
Step 1: *Apply the transformation on resized image to obtain its feature vector.*
Step 2: *Save these feature vectors for further comparison.*
Step 3: *Read the query image.*
Step 4: *Repeat step 1 to step 3 for each training image in the database to extract their feature vector.*
Step 5: *Perform* ***procedure Eucli_Dist ()***
{
Compute the distance measures for number of images from IDB with the target image using the equation 5.1.
}
Step 6: *Declare the Speech corresponding to this trainee image as identified Speech.*
Step 7: *Repeat the Step 5 and Step 6 are repeated for selected portion of feature vector.*
Step 8: *Return*
}

Perform procedure testing_feature ( )
{
Step 1: *Apply the transformation on row mean to obtain its feature vector.*
Step 2: *Save feature vectors for further comparison.*
Step 3: *Read the query image.*
Step 4: *Repeat step 1 to step 3 for each test image in the database to extract their feature vector.*
Step 5: *Perform* ***procedure Eucli_Dist ()***
{
Compute the distance measures for number of images from IDB with the target image using the equation 5.1.
}
Step 6: *Declare the Speech corresponding to this trainee image as identified Speech.*
Step 7: *Repeat the Step 5 and Step 6 are repeated for selected portion of feature vector.*
Step 8: *Return*
}

4. EXPERIMENTATION & RESULTS

The experimentation is carried out by MATLAB. It stands for MATrix LABoratory. MATLAB® is a high-performance language for technical computing. It integrates computation, visualization and programming in an easy-to-use environment where problems and solutions are expressed in familiar mathematical notation.

To study the proposed approach recorded every Speech 10 occurrences of each sentence were recorded. Recording was done at varying times. This forms the closed set for our experiment. From these speech samples were created with window size 256 and overlap of 128.

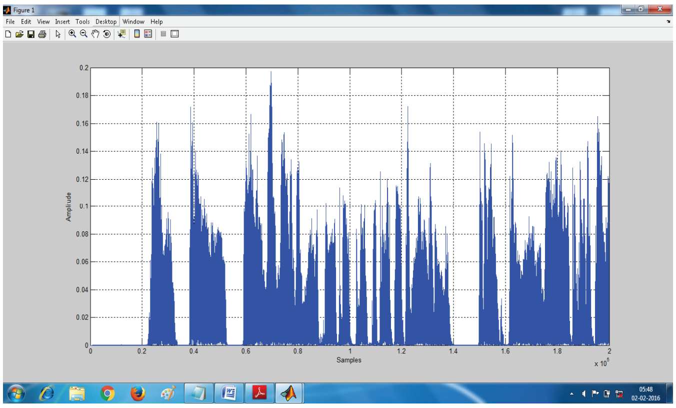

Fig.2 Original Wave file

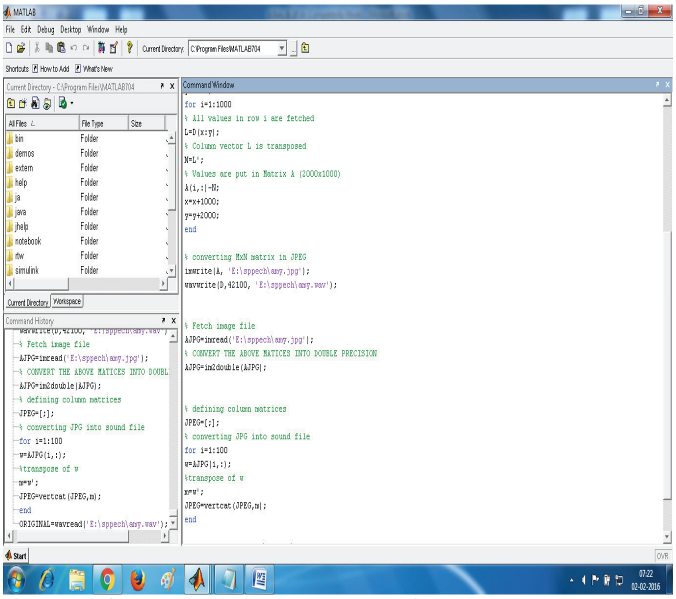

Fig.3 Code for Wave File into Column Matrix and JPEG format

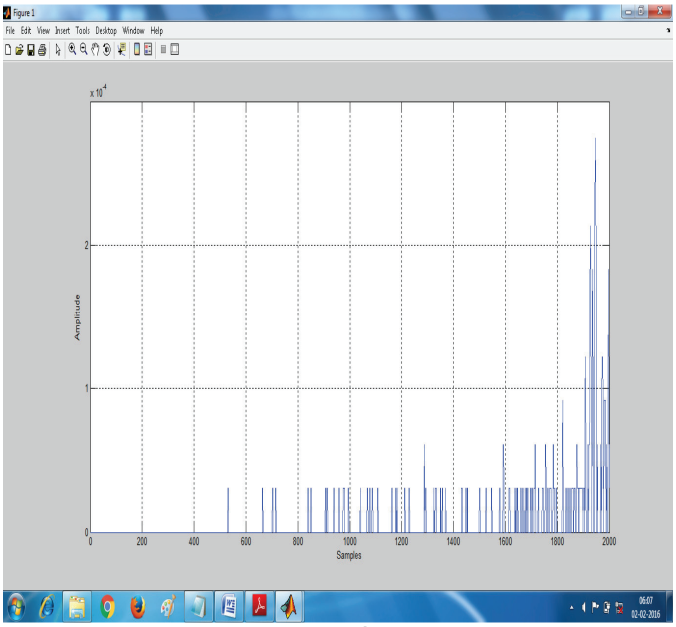

Fig.3 Wave File into Column Matrix

5. SIMILARITY AND PERFORMANCE MEASURES

To find the similarity measures between the images, various metrics are used to measure the distance between features of the images. Some of the well known distance metrics used in for image retrieval is presented below. The Euclidean Distance is calculated as below

$$d_E(x_1, x_2) = \sqrt{\sum_{i=1}^{i=n} \left(x_1(i) - x_2(i)\right)^2} \quad \ldots (5.1)$$

Where $x_1(i)$ is the feature vector of input image i and $x_2(i)$ is the feature vector of the target image i in the image database.

Accuracy

The accuracy of the identification system is calculated by

$$A\,(\%) = \frac{\text{No. of matches}}{\text{No. of Samples Tested}} \times 100 \times 100 \text{ -- } (5.3)$$

6. PERFORMANCE EVALUATION

The proposed feature extraction is experimented with the images collected from the standard VidTIMIT database and generated feature set images considered for this experiment are of the size. From the below Table 1.1 shows that recognition percentage of query images with Proposed Model gives the higher retrieval accuracy of 86.34%. The performance was evaluated using the Euclidean Distance classification by analysis of the values in the table the Proposed model is better for Speech identification.

Table 1.1 Recognition Accuracy of Full images

Portions feature selected	Number of Coefficient	DCT	WALSH	HAAR	Proposed RCF
256*256	65536	70.83	70.83	70.83	71.34
192*192	36864	75.27	76.11	77.5	79.56
128*128	16384	78.88	80	80	80.91
64*64	4096	82.77	84.16	84.16	84.82
32*32	1024	87.77	85.55	85.55	86
20*20	400	88.05	84.72	86.39	89.51
16*16	256	87.5	85	85	86.34

Fig.6 Comparison Graph with Existing Model

From the above fig.6 shows the pictorial representation of the performance evaluated. By analyzing the obtained results the Proposed RCF produced the best results.

7. CONCLUSION

In this paper, the speech recognition and distance based retrieval with feature extraction images based on DCT, WALSH and HAAR models has been presented. The experimental result proves the effectiveness of the proposed RCF methods provides good identification rate and Euclidean distance gives better for recognition of speech when compared to existing methods. The proposed RCF produces better results with 86.34% accuracy compared with existing methods.

8. REFERENCES

[1] S. Davis and P. Mermelstein, “Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences,” IEEE Transaction Acoustics Speech and Signal Processing, vol. 4, pp. 375-366, 1980.

[2] Wang Yutai, Li Bo, Jiang Xiaoqing, Liu Feng, Wang Lihao, “Speaker Recognition Based on Dynamic MFCC Parameters”, International Conference on Image Analysis and Signal Processing, pp. 406-409, 2009.

[3] Azzam Sleit, Sami Serhan, and Loai Nemir, “A histogram based speaker identification technique”, International Conference on ICADIWT, pp. 384-388, 2008.

[4] B. S. Atal, “Automatic Recognition of speakers from their voices”, Proc. IEEE, vol. 64, pp. 460-475, 1976.

[5] Jialong He, Li Liu, and G¨unther Palm, “A discriminative training algorithm for VQ-based speaker Identification”, IEEE Transactions on speech and audio processing, vol.7,No.3, pp. 353-356,1999.

[6] Debadatta Pati, S. R. Mahadeva Prasanna, “Non-Parametric Vector Quantization of Excitation Source Information for Speaker Recognition”, IEEE Region 10 Conference, pp. 1-4, 2008.

[7] Tridibesh Dutta and Gopal K. Basak, “Text dependent speaker identification using similar patterns in spectrograms”, PRIP'2007 Proceedings, Volume 1, pp.87-92, Minsk, 2007.

[8] Andrew B. Watson, “Image compression using the Discrete Cosine Transform”, Mathematica journal, 4(1), pp. 81-88,1994.

[9] Evgeniy Gabrilovich, Alberto D. Berstin: “Speaker recognition: using a vector quantization approach for robust text-independent speaker identification”, Technical report DSPG-95-9-001, 1995.

[10] Tridibesh Dutta, “Text dependent speaker identification based on spectrograms”, Proceedings of Image and vision computing, pp. 238-243, New Zealand 2007.

[11] J.P.Campbell, “Speaker recognition: a tutorial”, Proc. IEEE, vol. 85, no. 9, pp. 1437-1462, 1997.

[12] D. O.Shaughnessy, “Speech communications- Man and Machine”, New York, IEEE Press, 2nd Ed., pp. 199, pp. 437-458, 2000.