SPEAKER IDENTIFICATION USING 2-D DCT, WALSH AND HAAR ON FULL AND BLOCK SPECTROGRAM

Dr. H. B. Kekre
Senior Professor,
MPSTME, SVKM's NMIMS University
Mumbai, 400-056, India.

Dr. Tanuja K. Sarode
Assistant Professor,
Thadomal Shahani Engg. College, Bandra (W),
Mumbai, 400-050, India.

Shachi J. Natu
Lecturer,
Thadomal Shahani Engg. College Bandra (W),
Mumbai, 400-050, India.

Prachi J. Natu
Assistant Professor,
GVAIET, Shelu
Karjat, 410201,
India.

Abstract— **This paper aims to provide different approaches to text dependent speaker identification using DCT, Walsh and Haar transform along with use of spectrograms. Spectrograms obtained from speech samples are used as image database for the study undertaken. This image database is then subjected to various transforms. Using Euclidean distance as measure of similarity, most appropriate speaker match is obtained and is declared as identified speaker. Each transform is applied to spectrograms in two different ways: on full image and on image blocks. In both the ways, effect of different number of coefficients of transformed image is observed. Haar transform on full image reduces multiplications required by DCT and Walsh by 28 times whereas applying Haar transform on image blocks requires 18 times less mathematical computations as compared to DCT and Walsh on image blocks. Transforms when applied to image blocks, yield better or equal identification rates with reduced computational complexity.**

Keywords: Speaker identification; Speaker Recognition; Spectrograms; DCT; WALSH; HAAR;, Image Blocks

I. INTRODUCTION

Multiuser applications and extensive use of internet technology has increased the importance of security. Identifying users and granting access only to those users who are authorized is a key to provide security. Users can be identified using various approaches and their combinations. As the technology is getting advanced, more sophisticated approaches are being used to satisfy the need of security. Some of the most popular techniques are use of login and password, face recognition, fingerprint recognition; iris recognition etc. Use of login and password is becoming less reliable because of the ease with which attackers can steal the password such as sophisticated electronic eavesdropping techniques [1]. Face recognition, fingerprint recognition and iris recognition also carry their own drawbacks. Users should be willing to undergo the tests and should not get upset by these procedures when these techniques are used to identify them. Speaker identification allows non-intrusive monitoring and also achieves high accuracy rates which conform to most security requirements. Speaker recognition is the process of automatically recognizing who is speaking based on some unique characteristics present in speaker's voice [2]. For this recognition purpose, speaker specific characteristics present in speech signal need to be preserved. Job of Speaker recognition can be classified into two main categories, namely speaker identification and speaker verification. Speaker identification deals with distinguishing a speaker from a group of speakers. In contrast, speaker verification aims to determine if a person is the one who he/she claims to be from a speech sample.

Speaker identification problem can be divided into text dependent and text independent Speaker Identification based on relevance to speech contents [2]. When Speaker Identification requires the speaker saying exactly the enrolled or the given password/speech, it is known as Text Dependent Speaker Identification. In contrast, text independent Speaker Identification is a process of verifying the identity without constraint on the speech content. Text independent Speaker Identification is more convenient than text dependent Speaker Identification, because the user can speak freely to the system. However it requires longer training and testing utterances to achieve good performance.

Another classification of speaker identification task is based on set of speakers used for identification purpose. It is categorized as closed set and open set Speaker Identification [2-4]. In closed set problem, from N known speakers, the speaker whose reference template has the maximum degree of similarity with the template of input speech sample of unknown speaker is obtained. This unknown speaker is assumed to be one of the given set of speakers. Thus in closed set problem, system makes a forced decision by choosing the best matching speaker from the speaker database.

In the open set text dependent speaker identification, matching reference template for an unknown speaker's speech sample may not exist. In this paper, closed set text dependent speaker identification is considered. In the proposed method, speaker identification is carried out with spectrograms and transformation techniques such as DCT, WALSH and HAAR [15-18]. Thus an attempt is made to formulate a digital signal processing problem into pattern recognition of images.

The rest of the paper is organized as follows: in section II we present related work carried out in the field of speaker identification. In section III our proposed approach is presented. Section IV elaborates the experiment conducted and results obtained. Analysis of computational complexity is presented in section V. Conclusion has been outlined in section VI.

II. RELATED WORK

Speaker identification problem basically consists of two stages: feature extraction stage and pattern classification stage. In literature there are many approaches available for speaker identification process based on various approaches for feature extraction. Feature extraction is the process of extracting subset of features from the entire feature set. The basic idea behind the feature extraction is that the entire feature set is not always necessary for the identification process.

One of the popular approaches for feature extraction is the Mel Frequency Cepstrum Coefficients (MFCC). The MFCC parameter as proposed by Davis and Mermelstein [5] describes the energy distribution of speech signal in a frequency field. Wang Yutai et. al. [6] has proposed a speaker recognition system based on dynamic MFCC parameters. This technique combines the speaker information obtained by MFCC with the pitch to dynamically construct a set of the Mel-filters. These Mel-filters are further used to extract the dynamic MFCC parameters which represent characteristics of speaker's identity. A histogram based technique was proposed by Sleit, Serhan and Nemir [7] which uses a reduced set of features generated using MFCC method. For these features, histograms are created using predefined interval length. These histograms are generated first for all data in feature set for every speaker. In second approach, histograms are generated for each feature column in feature set of each speaker. Another widely used method for feature extraction is use of linear Prediction Coefficients (LPC). LPCs capture the information about short time spectral envelope of speech. LPCs represent important speech characteristics such as formant speech frequency and bandwidth [8].

Vector Quantization (VQ) is yet another approach of feature extraction [19-22]. In Vector Quantization based speaker recognition systems; each speaker is characterized with several prototypes known as code vectors [9]. Speaker recognition based on non-parametric vector quantization was proposed by Pati and Prasanna [10]. Speech is produced due to excitation of vocal tract. In this approach, excitation information can be captured using LP analysis of speech signal and is called as LP residual. This LP residual is further subjected to non-parametric Vector Quantization to generate codebooks of sufficiently large size. Combining nonparametric Vector Quantization on excitation information with vocal tract information obtained by MFCC was also introduced by them.

III. PROPOSED APPROACH

In the proposed approach, first we converted the speech samples collected from various speakers into spectrograms [11]. Spectrograms were created using Short Time Fourier Transfer method as discussed below:

In the approach using STFT, digitally sampled data are divided into chunks of specific size say 128, 256 etc. which usually overlap. Fourier transform is then obtained to calculate the magnitude of the frequency spectrum for each chunk. Each chunk then corresponds to a vertical line in the image, which is a measurement of magnitude versus frequency for a specific moment in time.

Thus we converted the speech database into image database. Different transformation techniques such as Discrete Cosine Transform [12], Walsh transform and Haar transform are then applied to these images in two different ways to obtain their feature vectors.

First, every transform is applied on full image to obtain feature vector of an image. From this feature vector obtained, partial feature vector was used to identify speaker. Second, transform is applied to image blocks obtained by dividing an image into four equal and non-overlapping blocks to get the feature vector of an image. From this feature vector again identification rate is obtained for various portions selected from the feature vector i.e. for partial feature vector. Out of total database, 80% of images were used as trainee images and 20% images were used as test images. Euclidean distance between test image and trainee image is used as a measure of similarity. Euclidean distance between the points X(X1, X2, etc.) and point Y (Y1, Y2, etc.) is calculated using the formula shown in equation. (1).

$$D = \sqrt{\sum_{i=1}^{n} (X_i - Y_i)^2} \qquad (1)$$

Smallest Euclidean distance between test image and trainee image means the most probable match of speaker. Algorithms for transformation technique on full image and transformation techniques on image blocks are given below.

A. *Transformation techniques on full image:*

While applying transformation technique on full image [27, 28], first image was resized to size 256*256. Then DCT/ Walsh/ Haar were applied to image. This generates the feature vector for an image. This process of generating feature vector was carried out for every trainee image in the database and the feature vectors were saved for decision making process. Similarly, feature vectors for each test images were generated and saved. Further, for every test image, Euclidean distance between test image and trainee image corresponding to the same sentence was calculated. The trainee image that gives smallest Euclidean distance corresponds to the 'identified' speaker. Size chosen for partial feature vectors is 192*192, 128*128, 64*64, 32*32, 20*20 and 16*16. This selection of feature vector is illustrated in following Fig. 1 and is based on the number of rows and columns that we selected from the feature vector of an image.

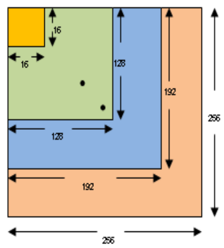

Fig. 1: Selection of partial feature vector

Same steps are repeated except the transformation technique used, to get feature vectors for image database using Walsh and Haar transform.

B. Transformation technique on image blocks

In the second approach, all three transformation techniques are applied to image blocks [27]. These image blocks are obtained by dividing an image into four equal and non-overlapping blocks. Fig. 2 shows how the image is divided and image blocks are obtained. Thus when N*N image is divided into four equal and non-overlapping blocks, we get blocks of size N/2 * N/2.

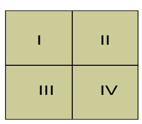

Fig. 2: Image divided into four equal non-overlapping blocks

While applying transformation technique on image blocks, first image was resized to size 256*256. Then it was divided into four equal and non-overlapping blocks as described above. On each block, DCT/ Walsh/ Haar transform was applied. These transformed blocks when appended one after the other, give feature vector of an image which is of size 128*512. Feature vectors for trainee as well as test images were generated and saved for decision making. While selecting partial feature vectors in this case, it was selected from each transformed block and then appended one after the other to form partial feature vector of an image. In this method, partial feature vectors of size 96*384, 64*256, 32*128, 16*64 and 8*32 were selected.

IV. EXPERIMENTS AND RESULTS

Implementation for the proposed approach was done on Intel Core 2 Duo Processor, 2.0 GHz, and 3 GB of RAM. Operating System used is Windows XP and softwares used are MATLAB 7.0 and Sound forge 8.0. To study the proposed approach we recorded six distinct sentences from 30 speakers: 11 males and 19 females. These sentences are taken from VidTIMIT database [13] and ELSDSR database [14]. For every speaker 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 spectrograms were created with window size 256 and overlap of 128. Before creation of spectrograms, DC offset present in speech samples was removed so that signals are vertically centered at 0. After removal of DC offset, speech samples were normalized with respect to amplitude to -3 dB and also with respect to time. Spectrograms generated from these speech samples form the image database for our experiment. In all we had 1800 spectrograms in our database.

Eight spectrograms per speaker have been used as trainee images and two spectrograms per speaker as test images. Thus in all we had 1440 spectrograms for training purpose and 360 spectrograms for testing purpose. In the first approach as stated earlier, transformation techniques i.e. DCT, Walsh and Haar were applied on full image to obtain feature vector of image. Later by selecting partial feature vectors, identification rate was obtained. Similarly results for second approach were obtained using full and partial feature vectors. In both approaches, Euclidean distance between feature vectors of test images and trainee images was used as a measure of similarity between images.

Since our work is restricted to text dependent approach, Euclidean distance for a test image of speaker say ‘x’ for a particular sentence say ‘s1’ is obtained by comparing the feature vector of that test image with the feature vectors of all the trainee images corresponding to sentence ‘s1’. Results are calculated for set of test images corresponding to each sentence. Fig.3 shows that the spectrogram for the same sentence, uttered by different speakers is different.

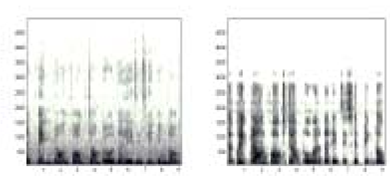

Fig.3: Spectrograms of sentence s1 for speaker 1 and speaker 5

A. Results for DCT on Spectrograms

1) Results for DCT on full image

Table I shows the identification rate for six sentences s1 to s6 when DCT is applied on full image and different numbers of DCT coefficients are taken to find the matching spectrogram. Table II shows the overall identification rate considering all sentences, for varying size of portion selected from feature vector It also shows the number of DCT coefficients used for identifying speaker for corresponding selected portion of feature vector.

TABLE I. IDENTIFICATION RATE FOR SENTENCES S1 TO S6 FOR VARYING PORTION OF FEATURE VECTOR WHEN DCT IS APPLIED TO FULL IMAGE

Portion of feature vector selected	Sentence					
	S1	*S2*	*S3*	*S4*	*S5*	*S6*
256*256	63.33	66.67	75	66.67	76.67	76.67
192*192	73.33	70	76.67	75	78.33	78.33
128*128	78.33	73.33	80	78.33	81.67	81.67
64*64	80	80	78.33	86.67	83.33	88.33
32*32	90	86.67	86.67	86.67	86.67	90
20*20	86.67	86.67	86.67	88.33	90	90
16*16	85	85	86.67	86.67	91.67	90

TABLE II. OVERALL IDENTIFICATION RATE FOR VARYING NUMBER OF DCT COEFFICIENTS WHEN DCT IS APPLIED TO FULL IMAGE

Portion of feature vector selected	Number of DCT coefficients	Identification rate (%)
256*256	65536	70.83
192*192	36864	75.27
128*128	16384	78.88
64*64	4096	82.77
32*32	1024	87.77
20*20	400	**88.05**
16*16	256	87.5

2) Results of DCT on image blocks

Table III shows sentence wise results obtained when DCT of image blocks is taken by dividing an image into four non-overlapping blocks. The overall identification rates when DCT of image blocks is taken by dividing an image into four non-overlapping blocks in Table IV.

TABLE III. IDENTIFICATION RATE FOR SENTENCES S1 TO S6 FOR VARYING PORTION OF FEATURE VECTOR USING DCT ON IMAGE BLOCKS

Portion of feature vector selected	Sentence					
	S1	*S2*	*S3*	*S4*	*S5*	*S6*
128*512	63.33	66.67	75	66.67	76.67	76.67
96*384	71.67	70	76.67	75	78.33	78.33
64*256	78.33	73.33	80	76.67	81.67	81.67
32*128	78.33	80	78.33	86.67	83.33	86.67
16*64	90	88.33	86.67	90	86.67	88.33
8*32	88.33	88.33	85	86.67	90	86.67

TABLE IV. OVERALL IDENTIFICATION RATE FOR DCT ON IMAGE BLOCKS

Portion of feature vector selected	Number of HAAR coefficients	Identification rate (%)
128*512	65536	70.83
96*384	36864	76.39
64*256	16384	80
32*128	4096	84.44
16*64	1024	**85.55**
8*32	256	85.27

B. Results of Walsh transform on Spectrograms

1) Results for Walsh on full image

Results of Walsh transform on Spectrograms are tabulated below. Table V shows the identification rate for sentences s1 to s6 when different numbers of Walsh transform's coefficients are taken to find the matching spectrogram i.e. to identify speaker using first approach. Table VI shows the overall identification rate considering all sentences, for various percentages of Walsh coefficients i.e. for partial feature vectors.

TABLE V. IDENTIFICATION RATE FOR SENTENCES S1 TO S6 FOR VARYING PORTION OF FEATURE VECTOR WHEN WALSH TRANSFORM IS APPLIED TO FULL IMAGE

Portion of feature vector selected	Sentence					
	S1	*S2*	*S3*	*S4*	*S5*	*S6*
256*256	63.33	66.67	75	66.67	76.67	76.67
192*192	75	71.67	76.67	73.33	78.33	81.67
128*128	80	75	78.33	83.33	81.67	81.67
64*64	86.67	83.33	81.67	85	83.33	85
32*32	86.67	81.67	81.67	88.33	83.33	91.67
20*20	91.67	78.33	83.33	85	86.67	83.33
16*16	86.67	85	83.33	85	83.33	86.67

TABLE VI. OVERALL IDENTIFICATION RATE FOR VARYING NUMBER OF COEFFICIENTS WHEN WALSH IS APPLIED TO FULL IMAGE

Portion of feature vector selected	Number of Walsh coefficients	Identification rate (%)
256*256	65536	70.83
192*192	36864	76.11
128*128	16384	80
64*64	4096	84.16
32*32	1024	**85.55**
20*20	400	84.72
16*16	256	85

2) Results of Walsh on image blocks

Table VII shows the sentence wise identification rate and Table VIII shows overall identification rate when Walsh transform is applied to image blocks by dividing an imaged into four non-overlapping, equal sized blocks.

TABLE VII. IDENTIFICATION RATE FOR SENTENCES S1 TO S6 FOR VARYING PORTION OF FEATURE VECTOR USING WALSH ON IMAGE BLOCKS

Portion of feature vector selected	Number Of DCT coefficients	Identification rate (%)
128*512	65536	70.83
96*384	36864	75
64*256	16384	78.61
32*128	4096	82.22
16*64	1024	**88.33**
8*32	256	86.67

TABLE VIII. OVERALL IDENTIFICATION RATE FOR WALSH TRANSFORM ON ROW MEAN OF AN IMAGE WHEN IMAGE IS DIVIDED INTO DIFFERENT NUMBER OF NON-OVERLAPPING BLOCKS OF EQUAL SIZE

Portion of feature vector selected	Sentence					
	S1	*S2*	*S3*	*S4*	*S5*	*S6*
128*512	63.33	66.67	75	66.67	76.67	76.67
96*384	75	71.67	76.67	73.33	78.33	81.67
64*256	80	75	78.33	83.33	81.67	81.67
32*128	86.67	83.33	81.67	85	83.33	85
16*64	86.67	81.67	81.67	88.33	83.33	91.67
8*32	86.67	85	83.33	85	83.33	86.67

C. *Results of Haar transform on Spectrograms*

1) *Results for Haar on full image*

Table IX shows sentencewise identification rate when 2-D Haar transform is applied to full image and different numbers of coefficients are taken. Overall identification rate for both is shown in Table X.

TABLE IX. IDENTIFICATION RATE FOR SENTENCES S1 TO S6 FOR PARTIAL VECTORS OF DIFFERENT SIZES WHEN HAAR TRANSFORM IS APPLIED TO FULL IMAGE

Portion of feature vector selected	Sentence					
	S1	*S2*	*S3*	*S4*	*S5*	*S6*
256*256	63.33	66.67	75	66.67	76.67	76.67
192*192	80	73.33	78.33	76.67	78.33	78.33
128*128	80	75	78.33	83.33	81.67	81.67
64*64	86.67	83.33	81.67	85	83.33	85
32*32	86.67	81.67	81.67	88.33	83.33	91.67
20*20	86.67	88.33	86.67	85	85	86.67
16*16	86.67	85	83.33	85	83.33	86.67

TABLE X. OVERALL IDENTIFICATION RATE FOR HAAR TRANSFORM ON FULL IMAGE

Portion of feature vector selected	Number of HAAR coefficients	Identification rate (%)
256*256	65536	70.83
192*192	36864	77.5
128*128	16384	80
64*64	4096	84.16
32*32	1024	85.55
20*20	400	**86.39**
16*16	256	85

2) *Results for Haar on image blocks*

Table XI shows identification rate for each sentence when 2-D Haar transform was applied on image blocks obtained by dividing an image into four non-overlapping and equal sized blocks. Overall identification rate for Haar transform on image blocks is shown in Table XII.

TABLE XI. IDENTIFICATION RATE FOR SENTENCES S1 TO S6 FOR VARYING PORTION OF FEATURE VECTOR USING HAAR ON IMAGE BLOCKS

Portion of feature vector selected	Sentence					
	S1	**S2**	**S3**	**S4**	**S5**	**S6**
128*512	63.33	66.67	75	66.67	76.67	76.67
96*384	73.33	70	78.33	76.67	78.33	81.67
64*256	80	75	78.33	83.33	81.67	81.67
32*128	86.67	83.33	81.67	85	83.33	85
16*64	86.67	81.67	81.67	88.33	83.33	91.67
8*32	86.67	85	83.33	85	83.33	86.67

TABLE XII. OVERALL IDENTIFICATION RATE FOR HAAR TRANSFORM ON IMAGE BLOCKS

Portion of feature vector selected	Number of Walsh coefficients	Identification rate (%)
128*512	65536	70.83
96*384	36864	76.11
64*256	16384	80
32*128	4096	84.16
16*64	1024	**85.55**
8*32	256	85

V. COMPLEXITY ANALYSIS

For 2-D DCT on N*N image, $2N^3$ multiplications are required and $2N^2(N-1)$ additions are required. For 2-D Walsh on N*N image, $2N^2(N-1)$ additions are required. For 2-D Haar transform on N*N image where N=2m , number of multiplications required are (m+1)N and number of additions required are $2mn^2$. For 1-D DCT on 1*N image, N^2 multiplications are needed and N(N-1) additions are needed. For 1-D WALSH on 1*N image, N(N-1) additions are needed. For 1-D Haar transform on N*1 image, number of multiplications required are (m+1)N and number of additions are mn. These details are summarized in table XIII.

TABLE XIII. COMPUTATIONAL DETAILS FOR 2-D DCT ON FULL N*N IMAGE, 2-D WALSH ON FULL N*N IMAGE, 2-D HAAR ON FULL N*N IMAGE, 2-D DCT ON FOUR IMAGE BLOCKS OF SIZE N/2*N/2 EACH, 2-D WALSH ON FOUR IMAGE BLOCKS OF SIZE N/2*N/2 EACH AND 2-D HAAR ON FOUR IMAGE BLOCKS OF SIZE N/2*N/2 EACH RESPECTIVELY.

Algorithm	Parameter	
	Number of multiplications	*Number of additions*
DCT on full image	$2N^3$	$2N^2(N-1)$
Walsh on full image	0	$2N^2(N-1)$
Haar on full image	$2(m+1)N^2$	$2mN^2$
DCT on image blocks	N^3	$N^2(N-2)$
Walsh on image blocks	0	$N^2(N-2)$
Haar on Image Blocks	$2(m+1)N^2$	$2mN^2$

Table XIV shows comparison of number of multiplications required, number of additions required and identification rate when DCT, Walsh and Haar are applied on full image of size 256*256 and on Row Mean of an image.

TABLE XIV. NUMBER OF MULTIPLICATIONS AND NUMBER OF ADDITIONS FOR DCT ON FULL IMAGE, 2-D WALSH ON FULL IMAGE, 2-D HAAR ON FULL IMAGE (IMAGE SIZE 256*256), 2-D DCT ON IMAGE BLOCKS, 2-D WALSH ON IMAGE BLOCKS AND 2-D HAAR ON IMAGE BLOCKS (BLOCK SIZE 128*128).

Algorithm	Parameter		
	Number of multiplications	*Number of additions*	*Identification rate (%)*
DCT on full image	$3.3*10^7$	$3.3*10^7$	88.05
Walsh on full image	0	$3.3*10^7$	85.55
Haar on full image	$1.1*10^6$	10^6	86.39
DCT on image blocks	$1.6*10^7$	$1.6*10^7$	88.33
Walsh on image blocks	0	$1.6*10^7$	85.55
Haar on Image Blocks	10^6	$9.1*10^5$	85.55

VI. CONCLUSION

In this paper we considered closed set text dependent speaker identification rate using three different transformation techniques: DCT, Walsh and Haar. For each transformation technique identification rate was obtained using two approaches. First, by applying transformation technique on full image and second, by applying transformation technique on image blocks by dividing image into four equal sized and non-overlapping blocks. It has been observed that in the first approach, as the number of coefficients chosen is smaller up to a certain limit; better identification rate is achieved in all three transformation techniques. DCT on full image gives its best identification rate of 88.05% when 20*20 portion of feature vector is selected i.e. for 400 DCT coefficients. Walsh transform gives its best identification rate which is 85.55% for 32*32 portion of feature vector (1024 coefficients) and Haar transform gives its best performance of 86.39% for feature vector portion of size 20*20 (400 coefficients).

Further when the identification rates for all three transformation techniques on image blocks are compared, it has been observed that all three transformation techniques except HAAR transform show slightly increased identification rate that too with reduced number of mathematical computations. Though number of coefficients used in transformation technique on image blocks method is higher in case of DCT and Haar, as compared to number of coefficients selected from feature vector of full image, number of multiplications and additions reduce drastically as compared to transformation on full image.

In Walsh transform on full image, numbers of mathematical computations required are greatly reduced as compared to DCT since no multiplications are required in Walsh. These computations are further reduced by use of Haar transform but at the slight expense of identification rate. DCT on full image requires 28 times more multiplications and 32 times more additions as compared to Haar transform on full image. Walsh transform on full image 32 times more additions as compared to Haar transform on full image. Though the number of multiplications required in Walsh transform on full image is zero, total CPU time required by Haar transform is less than that of Walsh transform. Smallest numbers of mathematical computations are required for Haar transform on Row Mean of an image.

Similarly Haar transform on image blocks gives trade of between number of mathematical computations and identification rate when compared to DCT on image blocks and Walsh on image blocks. DCT on image blocks requires 16 times of multiplications and 18 times more additions as compared to Haar transform on image blocks. Again here though the number of multiplications required in Walsh transform on full image is zero, because of large number of additions, it requires more CPU time. After overall comparison, we conclude that transformation techniques on image blocks provide better performance with respect to computations required. In that, Haar transform on image blocks gives slightly less identification rate than other two transforms but with reduced number of computation and hence it is faster.

REFERENCES

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

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

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

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

[5] 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.

[6] 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

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

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

[9] 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, May 1999.

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

[11] 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.

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

[13] http://www.itee.uq.edu.au/~conrad/vidtimit/

[14] http://www2.imm.dtu.dk/~lf/elsdsr/

[15] H.B.Kekre, Sudeep D. Thepade, “Improving the Performance of Image Retrieval using Partial Coefficients of Transformed Image”, International Journal of Information Retrieval (IJIR), Serials Publications, Volume 2, Issue 1, pp. 72-79 (ISSN: 0974-6285), 2009.

[16] H.B.Kekre, Tanuja Sarode, Sudeep D. Thepade, “DCT Applied to Row Mean and Column Vectors in Fingerprint Identification”, In Proceedings of International Conference on Computer Networks and Security (ICCNS), 27-28 Sept. 2008, VIT, Pune.

[17] H.B.Kekre, Sudeep D. Thepade, Archana Athawale, Anant Shah, Prathmesh Verlekar, Suraj Shirke,“Energy Compaction and Image Splitting for Image Retrieval using Kekre Transform over Row and Column Feature Vectors”, International Journal of Computer Science and Network Security (IJCSNS),Volume:10, Number 1, January 2010, (ISSN: 1738-7906) Available at www.IJCSNS.org.

[18] H.B.Kekre, Sudeep D. Thepade, Archana Athawale, Anant Shah, Prathmesh Verlekar, Suraj Shirke, “Performance Evaluation of Image Retrieval using Energy Compaction and Image Tiling over DCT Row Mean and DCT Column Mean”, Springer-International Conference on Contours of Computing Technology (Thinkquest-2010), Babasaheb Gawde Institute of Technology, Mumbai, 13-14 March 2010, The paper will be uploaded on online Springerlink.

[19] H.B.Kekre, Tanuja K. Sarode, Sudeep D. Thepade, Vaishali Suryavanshi, “Improved Texture Feature Based Image Retrieval using Kekre’s Fast Codebook Generation Algorithm”, Springer-International Conference on Contours of Computing Technology (Thinkquest-2010), Babasaheb Gawde Institute of Technology, Mumbai, 13-14 March 2010, The paper will be uploaded on online Springerlink.

[20] H. B. Kekre, Tanuja K. Sarode, Sudeep D. Thepade, “Image Retrieval by Kekre’s Transform Applied on Each Row of Walsh Transformed VQ Codebook”, (Invited), ACM-International Conference and Workshop on Emerging Trends in Technology (ICWET 2010),Thakur College of Engg. And Tech., Mumbai, 26-27 Feb 2010, The paper is invited at ICWET 2010. Also will be uploaded on online ACM Portal.

[21] H. B. Kekre, Tanuja Sarode, Sudeep D. Thepade, “Color-Texture Feature based Image Retrieval using DCT applied on Kekre’s Median Codebook”, International Journal on Imaging (IJI), Volume 2, Number A09, Autumn 2009,pp. 55-65. Available online at www.ceser.res.in/iji.html (ISSN: 0974-0627).

[22] H. B. Kekre, Ms. Tanuja K. Sarode, Sudeep D. Thepade, "Image Retrieval using Color-Texture Features from DCT on VQ Codevectors obtained by Kekre’s Fast Codebook Generation", ICGST-International Journal on Graphics, Vision and Image Processing (GVIP), Volume 9, Issue 5, pp.: 1-8, September 2009. Available online at http: //www.icgst.com /gvip /Volume9 /Issue5 /P1150921752.html.

[23] H. B. Kekre, Sudeep Thepade, Akshay Maloo, “Image Retrieval using Fractional Coefficients of Transformed Image using DCT and Walsh Transform”, International Journal of Engineering Science and Technology, Vol.. 2, No. 4, 2010, 362-371

[24] H. B. Kekre, Sudeep Thepade, Akshay Maloo,”Performance Comparison of Image Retrieval Using Fractional Coefficients of Transformed Image Using DCT, Walsh, Haar and Kekre’s Transform”, CSC-International Journal of Image processing (IJIP), Vol.. 4, No.2, pp.:142-155, May 2010.

[25] H. B. Kekre, Tanuja Sarode “Two Level Vector Quantization Method for Codebook Generation using Kekre’s Proportionate Error Algorithm” , CSC-International Journal of Image Processing, Vol.4, Issue 1, pp.1-10, January-February 2010

[26] H. B. Kekre, Sudeep Thepade, Akshay Maloo, “Eigenvectors of Covariance Matrix using Row Mean and Column Mean Sequences for Face Recognition”, CSC-International Journal of Biometrics and Bioinformatics (IJBB), Volume (4): Issue (2), pp. 42-50, May 2010.

[27] H. B. Kekre, Tanuja Sarode, Shachi Natu, Prachi Natu, “Performance Comparison Of 2-D DCT On Full/Block Spectrogram And 1-D DCT On Row Mean Of Spectrogram For Speaker Identification”, (Selected), CSC-International Journal of Biometrics and Bioinformatics (IJBB), Volume (4): Issue (3), in press.

[28] H. B. Kekre, Tanuja Sarode, Shachi Natu, Prachi Natu, “Performance Comparison of Speaker Identification Using DCT, Walsh, Haar On Full And Row Mean Of Spectrogram” , (Selected), International Journal of Computer Applications, August 2010 Edition, in press.

AUTHORS PROFILE

Dr. H. B. Kekre has received B.E. (Hons.) in Telecomm. Engg. from Jabalpur University in 1958, M.Tech (Industrial Electronics) from IIT Bombay in 1960, M.S.Engg. (Electrical Engg.) from University of Ottawa in 1965 and Ph.D. (System Identification) from IIT Bombay in 1970. He has worked Over 35 years as Faculty of Electrical Engineering and then HOD Computer Science and Engg. at IIT Bombay. For last 13 years worked as a Professor in Department of Computer Engg. at Thadomal Shahani Engineering College, Mumbai. He is currently Senior Professor working with Mukesh Patel School of Technology Management and Engineering, SVKM’s NMIMS University, Vile Parle (w), Mumbai, INDIA. He ha guided 17 Ph.D.s, 150 M.E./M.Tech Projects and several B.E./B.Tech Projects. His areas of interest are Digital Signal processing, Image Processing and Computer Networks. He has more than 250 papers in National / International Conferences / Journals to his credit. Recently six students working under his guidance have received best paper awards. Currently he is guiding ten Ph.D. students.

Dr. Tanuja K. Sarode has received M.E. (Computer Engineering) degree from Mumbai University in 2004 and Ph.D. from Mukesh Patel School of Technology, Management and Engg. SVKM’s NMIMS University, Vile-Parle (W), Mumbai, INDIA, in 2010. She has more than 10 years of experience in teaching. Currently working as Assistant Professor in Dept. of Computer Engineering at Thadomal Shahani Engineering College, Mumbai. She is member of International Association of Engineers (IAENG) and International Association of Computer Science and Information Technology (IACSIT). Her areas of interest are Image Processing, Signal Processing and Computer Graphics. She has 70 papers in National /International Conferences/journal to her credit.

Shachi Natu has received B.E. (Computer) degree from Mumbai University with first class in 2004. Currently Purusing M.E. in Computer Engineering from University of Mumbai. She has 05 years of experience in teaching. Currently working as Lecturer in department of Information Technology at Thadomal Shahani Engineering College, Bandra (w), Mumbai. Her areas of interest are Image Processing, Data Structure, Database Management Systems and operating systems. She has 2 papers in National / International Conferences /journal to her credit.

Prachi Natu has received B.E. (Electronics and Telecommunication) degree from Mumbai University with first class in 2004. Currently Purusing M.E. in Computer Engineering from University of Mumbai. She has 04 years of experience in teaching. Currently working as Lecturer in Computer Engineering department at G. V. Acharya Institute of Engineering and Technology, Shelu. Mumbai. Her areas of interest are Image Processing, Database Management Systems and operating systems. She has 2 papers in National / International Conferences /journal to her credit.