International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-2, December 2019

Copy and Move Detection in Audio Recordings using Dynamic Time Warping Algorithm

Kasi prasad Mannepalli, Pelluri vamsi Krishna, Kovi Vamsi Krishna, Kodali Rama Krishna

Abstract: Copy and move forgery technique is a versatile technique used by criminals to change the evidences unlawfully. This is done by removing or adding the segments of an audio recording from another audio using simple software tools . Now-a-days in courts and forensics lab we use digital audio or speech as proof of evidence. With the advancement in digital software and technology it is made possible to modify the original audio data and tamper with it. In this paper we are using a robust method of copy and move detection using MFCC and Dynamic time warping which is robust against several kinds of regularly used post-processing actions and background noise, which highlights its promising potential of the suggested method as an audio tampering detection tool in various practical forensics applications with help of DTW. The proposed method is giving good outcome in differentiating the original sequences and forged sequences. This is done and implemented in mat lab.

Index terms - Audio forensics, Dynamic time warping, MFCC, Speech forgeries, ZCR.

I. INTRODUCTION

Audio recordings are used very abundantly as a digital evidences and proofs in courts[1]. Technology has also been increasing in a rapid way that more powerful software's are developing in such a way that any normal human can tamper or modify the audio recording just by copying and pasting the audio segments in the positions where the user wants it. This is made very easy with the help of a professional software tools. For example "I had been to police station" can be changed very easily to "I had not been to police station" just by inserting not in between "been" and "had". Just by changing the semantics from the same audio file it is possible to modify the whole sentence by using above example. For just to experiment it is fine but, in real life forensics copy move detection of audio recordings plays important role and is very popular. It is not so easy to detect the forged segment in the audio process without proper audio detection tool. Listening to the whole audio recordings over and over until we observe a slight change may take longer hours. Moreover, if the user uses effective filtering techniques and post processing methods it may be very difficult to locate the duplicated segments in the audio file.

This is why we propose a robust copy move detection which is necessary and important in forensics industries that deals with audio and speech recordings. During the recent advancements in technology many software's have been developed like audio splicing[2],[3],[4],[5] forgery detection, image copy move detection, audio recorders detection[6],[7], audio compression detection[8], speech tampering detection[9]. These discoveries have made many researchers enthusiastic towards the things above mentioned.

They have gained a lot of attention of the forensics and researchers towards their discovery in the recent years. Many types and variety of copy move detection have been proposed during our recent study but, only very few of them have published in the field of audio copy move detection. It has been observed from the literature that, speech processing application need to be designed with suitable features. Many researchers are using a combination of various features of speech signal. Some of them are temporal features[10-12].MFCC features[13], pitch Chroma, spectral flux and tonal power ratio[14-16].

To classify the signals for a particular application a classification technique will be used[17-21]. In some of the pattern recognition applications it is also observed that a signal de-noising is required depending upon the data acquisition techniques[22-25]. Transforms are applied to some of the application like MRI and CT images to obtain the enhanced image. Various quantitative analysis is done on the MRI and CT image[25-29].

This research has interested us in moving forward with our project. Even though many software's have come into the world to deal with copy move detection for the audio and speech recordings with the help of powerful post processing techniques the detection rate is very low.

We have to create more powerful algorithm in-order to nullify the error and provide more accurate results. One challenging issues is that to find it's robust features that helps us to minimize the error after post processing techniques. We should be able to anticipate the outcome after using such techniques with a minimum error.

Revised Manuscript Received on December 30, 2019.

* Correspondence Author

Kasiprasad Mannepalli*, Department of electronics and communication, Koneru Lakshmaiah Education Foundation (Deemed to be University), Vaddeswaram, Andhra Pradesh, India.

Pelluri vamsi Krishna, Department of electronics and communication, Koneru Lakshmaiah Education Foundation (Deemed to be University), Vaddeswaram, Andhra Pradesh, India.

Kovi vamsi Krishna, Department of electronics and communication, Koneru Lakshmaiah Education Foundation (Deemed to be University), Vaddeswaram, Guntur (dist), Andhra Pradesh, India.

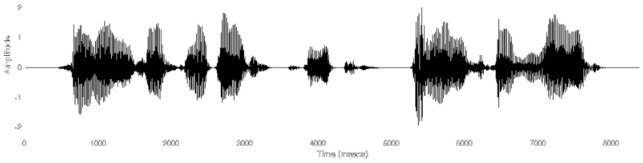
Kodali rama Krishna, Department of electronics and communication, Koneru Lakshmaiah Education Foundation (Deemed to be University), Vaddeswaram, Guntur (dist), Andhra Pradesh, India.

II. COPY AND MOVE FORGERY IN SPEECH RECORDINGS.

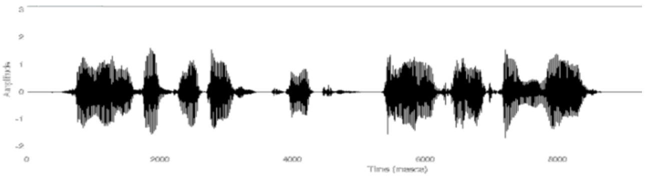
In this segment we will introduce the topic copy move forgery and some proposed methodology to deal and analyse the robustness against many post processing techniques using machine learning algorithms like Dynamic time warping.

A. Speech features of Copy move forgery

Copy move detection is a commonly used technique to forge digital audio recordings. Any person can easily manipulate the audio recording just by copying the segment of an audio at one position and pasting the same segment in other position. The segment of one audio can be pasted in the various other audio files through help of many software's available in the world. Furthermore, with help of powerful post-processing techniques like adding noise in the segments, sampling, quantization, filtering, compressing etc.. it is very difficult to find or trace the audio segments.



a) Copy move original image

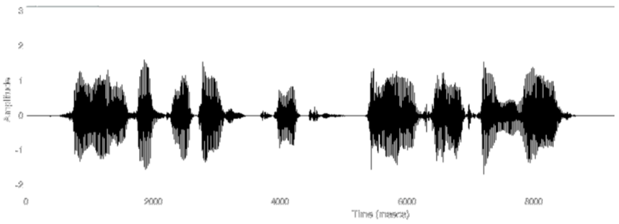


b) Copy move forged signal

Fig.1: It represents the copy move method of forged and original signal. a) Represents the original image. B) it represents the copy move forged audio signal.

B. Zero Crossing Rate

We mostly use zero crossing frequency in audio recognition and recovery of music information, which is a key feature for classifying percussive sounds. It is the frequency at which the signal switches from positive to zero to negative and negative to zero to positive. We can use this as a pitch detection algorithm in voice activity detection, such as whether the sound in an audio is generated by human or machine.



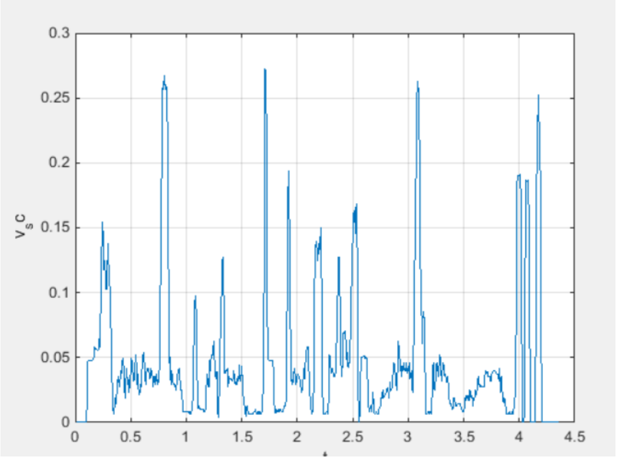
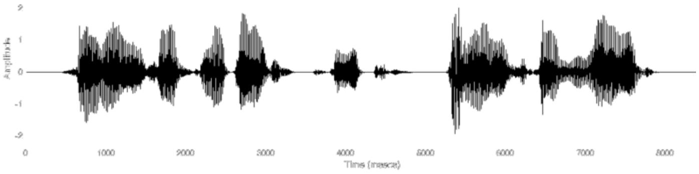


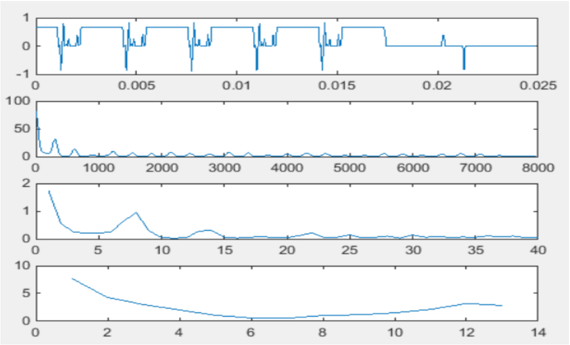
Fig.2: zero crossing rate of audio signal. we use this method for detection of audio signal which gives the change in frequency values from positive to negative or negative to positive.

C. MFCC

In audio hearing, we for the most part utilize this MFC (cephalic frequency coefficients) as a transient power range of a sound. MFCCs are only cephalic coefficients of mel frequency that are coming about because of the cephalic portrayal of a sound clasp. The distinction between the cepstrum and the cepstrum of the mel recurrence is that in the MFC, the recurrence groups are equidistant on the mel scale, which is nearer to the reaction of the human sound-related framework contrasted with the directly dispersed recurrence groups utilized in the ordinary hedgehog. This distortion in recurrence it can permit a superior portrayal of the sound, for instance in sound pressure we utilize this strategy. These coefficients make the entire audio signal into shorter frames. We then calculate the power spectrum of each frame and all the frames which are of unvoiced data are removed by mel filter bank.

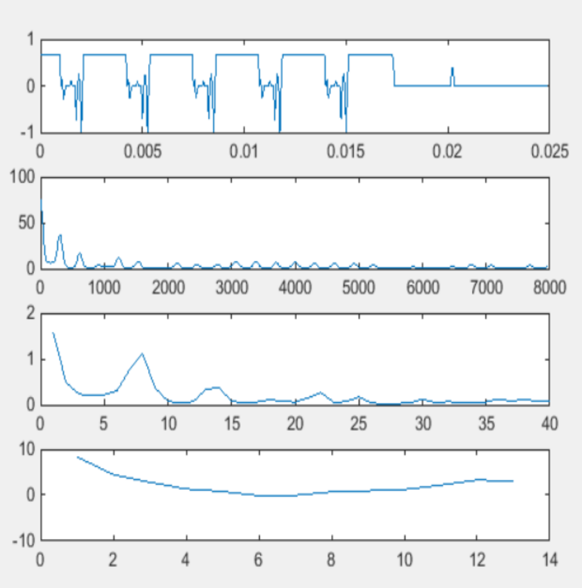


a) Audio signal representation in terms of amplitude and time



(b) MFCC of the original signal.

(c) Copy-move forged signal and its representation in terms of amplitude and time.



(a) MFCC of forged signal

Fig.3: It represents the feature extraction of the signal with help of MFCC. (a) It represents the MFCC of the original signal. (b) It represents the MFCC of forged signal with noise removal.

III. COLLECTION OF DATA

We have collected data samples of the audio recordings using phone recorder with a standard bit rate of 100kb/sec in a noise free environment. We have used different sentences in the recording like both positive and negative with human voice and also forged the signals to observe the difference in it. Around 200 samples have been collected from 5 persons each speaking 10 sentences in both positive and negative. We have pre-processed the data samples and extracted the features like mel frequency cepstral coefficients and Zero crossing rate for further processing the data.

IV. METHODOLOGY

In this paper we are proposing a robust method to detect speech forgery in the audio recordings. First we need to extract features from the audio recording files. To extract features we use zero crossing rate and MFCC(mel frequency cepstrum coefficients). After extracting the features we use DTW algorithm to compute the features like ceptrum frequencies between original and forged which we want and making those features computable to dtw algorithm. Finally we detect the forged speech recording with help of DTW distance between original and forged signals.

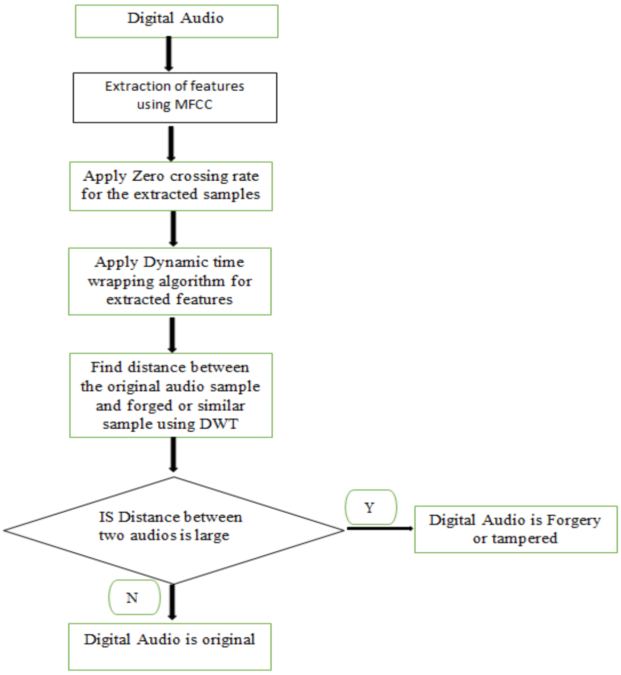

Digital Audio
Extraction of features using MFCC
Apply Zero crossing rate for the extracted samples
Apply Dynamic time wrapping algorithm for extracted features
Find distance between the original audio sample and forged or similar sample using DWT
IS Distance between two audios is large
Y
Digital Audio is Forgery or tampered
N
Digital Audio is original

Fig.4: Flow chart for overall analysis of copy and move detection using dynamic time warping algorithm.

A. Extraction of features from the audio recordings.

In this we are extracting the features of the audio by using MFC coefficients and zero crossing rate. The features are compared for similarity between them and we use these features which are best from both techniques and helps us to obtain more unique features. The change in zero crossing rate helps us to identify when the audio signal is changing from one value to another value . The values which are zero can be neglected as there is no change in the audio which means it is unvoiced data. This unvoiced data can be segmented separately from the voiced data.

$$ZCR=\frac{1}{T-1}\sum_{t=1}^{T-1} 1_{R<0}(S_t S_{t-1}) \quad (1)$$

where S is the signal length, T ,1(R<0) is indicator function.

We are using MFCC which is mel frequency coefficients with help of these coefficient values we are calculating the variance and standard deviation and observing the difference between original and forged audio. We have taken the coefficient values of multiple samples and calculated the variance and standard deviation.

B. Similarity computation and dynamic time warping

The Features which we obtained are classified in a sequential order. Each feature obtained is kept in multiple data files to find its unique features. Dynamic time warping is a kind of algorithm used for measuring the likeness between two data sequences which might vary in speeds. For example in speech recognition two speeches may not end at same time but with help of DWT we can identify which is speech is faster.

We have applied DWT to each and every feature obtained from MFCC and have found the distance between two subsequent samples like distance between original and forged audio signals, difference between two original samples and distance between two forged data samples. This is to know the similarity's and difference between samples. DWT is a method that calculates the optimal match between two sequences. Each and every index is matched with one or two more indices from other sequences and vice versa.

V. RESULTS AND DISCUSSIONS

The dynamic time warping algorithm is used in measuring similarity between data samples, with help of this algorithm we have calculated the distance between original to original audio segments and original to forged audio segments and forged to forged audio segments as shown in table.1.

A. Selection of threshold value

In this sub section we will see how we have chosen the threshold value. In a speech to detect whether the segment is forged or not proper threshold value is required .if the selection of threshold is too small, then some of the duplicated segments may be missed. However, if the threshold value is very high the accuracy is increased but results in false detection of forged segments. selection of threshold plays an important role in speech forgery detection.

TABLE 1

RESULTS OF DTW DISTANCE [%].

DTW distance	0 ~ 0.04	0 ~ 0.5	0 ~ 1
Forged audio segments	93.62	99.32	99.96
Non forged audio segments	0	0.82	3.02

in computing the similarities between non forged mfcc values and forged mfcc values .

From table.1 we can observe that the forged audio segment pairs which are having the difference in distance between 0 to 0.04 are of 0 percent and the forged audio segment pairs which are having a difference in distance from 0 to 0.5 are of 99.32 percent, and remaining pairs which are having a difference in distance from 0 to 1 are of 99.96 percent. The difference in distance is calculated from the Table.2. If we choose a threshold value of 1 then maximum number of audio segments get detected as forged and this results in false prediction. As we can see that for a threshold value of 0.5 the number of forged segments are 99.32% and number of non-forged segments are of 0.81 % pairs. Hence we are taking the threshold value as 0.5 for the proposed method. If the DTW Difference for two audio segments is less than the Threshold value i.e. 0.5 the audio segments are non forged or original audio segments. If the difference exceeds the threshold value then it is considered as forged signal or duplicated segments.

B. Computing using Dynamic time Warping

The obtained features from MFCC of the sample data are arranged in a sequential manner and are given to the DTW algorithm. From Table.2. It is observed that the different Mfcc samples were taken into consideration and distance between each sample is calculated. From table 2 it is observed that the distance from two original and forged segments is much greater than the distance between two forged segments or distance between two non-forged signals. The distance between two non-forged segments is very low. If we compare, the difference between original to forged. we are taking the maximum difference into consideration i.e. maximum difference between the two distances. The maximum difference values are stated in the table.2. Based on the maximum difference in values we choose the threshold limit. The DTW algorithm helps in computing the similarities between non forged mfcc values and forged mfcc values.

TABLE 2

DTW FINAL DISTANCE VALUES AND MAXIMUM DIFFERENCE IN FORGED SPEECH

DTW distance values	MFCC 1	MFCC 2	MFCC 3	MFCC4	MFCC 5	MFCC 6	MFCC 7	MFCC 8	MFCC 9	MFCC 10	MFCC 11	MFCC 12	MFCC 13
Distance between two Original segments	0.037488716	0.016916916	0.062906994	0.099728541	0.057980463	0.019956947	0.074717154	0.030848861	0.027433474	0.075570015	0.062906994	0.098728541	0.026906884

Distance between original and forged segments	**0.22703303**	**0.24028382**	**0.35181043**	**0.14082748**	**0.2913187**	**0.17052413**	**0.15305472**	**22.6241142**	**0.47406236**	**0.30168126**	**0.45210112**	**0.15682748**	**0.31581034**
Distance between two forged segments	0.166263792	0.070537733	0.018717965	0.047922644	0.104865704	0.131944331	0	1.148871531	0.140827477	0.070537733	0.018717965	0.054922644	0.017817965

VI . CONCLUSION

In this paper, we have investigated the detection of copy move forgery in an audio recording. The experiment analysis has made us to propose a robust copy move detection method with help of the features which we extracted through using MFCC and zero crossing rate. The dynamic time warping method helped in computing the distance and similarities between two audio sequences. Our proposed method can effectively detect and consider whether the audio recording is forged or not based on the experimental analysis. However, the proposed method still has some limitations, in the future work a combination of spectral, prosodic features may be utilized along with dimensionality reduction techniques, To improve the accuracy even when the forgery is done with powerful post processing techniques are used in the audio segments that can prevent the change in pitch levels and normalize the copy moved data then our proposed method may not be robust against such techniques. In our future work, we will improve our methodology, so that it can detect a forged audio under such circumstances.

REFERENCES

1. Qi Yan, Student Member,Rui Yang,Jiwu Huang."Robust Copy-move Detection of Speech Recording Using Similarities of Pitch andFormant",DOI-10.1109/TIFS.2019.2895965, IEEE.
2. Alan J Cooper. Detecting butt-spliced edits in forensic digital audio recordings. In Proc. of 39th Int. Conf. Audio Forensics: Practices and Challenges. Audio Engineering Society, 2010.
3. Jiaorong Chen, Shijun Xiang, Hongbin Huang, and Weiping Liu. Detecting and locating digital audio forgeries based on singularity analysis with wavelet packet. Multimedia Tools and Applications, 75(4):2303–2325, 2016.
4. Xunyu Pan, Xing Zhang, and Siwei Lyu. Detecting splicing in digital audios using local noise level estimation. In Proc. of 2012 IEEE Int.Conf. Acoustics, Speech and Signal Processing (ICASSP), pages 1841–1844. IEEE, 2012.
5. Hong Zhao, Yifan Chen, Rui Wang, and Hafiz Malik. Audio splicing detection and localization using environmental signature. Multimedia Tools and Applications, 76(12):13897–13927, 2017.
6. Christian Kraetzer, Andrea Oermann, Jana Dittmann, and Andreas Lang. Digital audio forensics: A first practical evaluation on microphone and environment classification. In Proc. of the 9th Workshop on Multimedia & Security, pages 63–74. ACM, 2007.
7. Huy Quan Vu, Shaowu Liu, Xinghua Yang, Zhi Li, and Yongli Ren. Identifying microphone from noisy recordings by using representative instance one class-classification approach. Journal of networks, 7(6):908–917, 2012.
8. Mengyu Qiao, Andrew H Sung, and Qingzhong Liu. Improved detection of MP3 double compression using content-independent features. In Proc. of 2013 IEEE Int. Conf. Signal Processing, Communication and Computing (ICSPCC), pages 1–4. IEEE, 2013.
9. William M Campbell, Kevin J Brady, Joseph P Campbell, R Granville, and Douglas A Reynolds. Understanding scores in forensic speaker recognition. In Proc. of IEEE Odyssey, ISCA Speaker Recognition Workshop, pages 1–8. IEEE, 2006.
10. Mannepalli, K., Sastry, P.N., Suman, M., "Emotion recognition in speech signals using optimization based multi-SVNN classifier", Journal of King
11. Mannepalli, K., Sastry, P.N., Suman, M.,"Accent recognition system using deep belief networks for telugu speech signals", Advances in Intelligent Systems and Computing, 515, 2017, pp. 99-105. DOI: 10.1007/978-981-10-3153-3_10.
12. Mannepalli, K., Sastry, P.N., Rajesh, V., "Accent detection of Telugu speech using prosodic and formant features", International Conference on Signal Processing and Communication Engineering Systems - Proceedings of SPACES 2015, (2015),in Association with IEEE, art. no. 7058274, pp. 318-322. DOI: 10.1109/SPACES.2015.7058274.
13. Mannepalli, K., Sastry, P.N., Suman. M, "MFCC-GMM based accent recognition system for Telugu speech signals", International Journal of Speech Technology, 19 (1), (2016), pp. 87-93.
14. Mannepalli, K., Sastry, P.N., Suman, M.,"A novel Adaptive Fractional Deep Belief Networks for speaker emotion recognition", Alexandria Engineering Journal, 56 (4), 2017, pp. 485-497. DOI: 10.1016/j.aej.2016.09.002.
15. Mannepalli, K., Sastry, P.N., Suman, M., "FDBN: Design and development of Fractional Deep Belief Networks for speaker emotion recognition",International Journal of Speech Technology, 19 (4), 2016, pp. 779-790.
16. Mannepalli, K., Sastry, P.N., Suman, M., "Emotion recognition in speech signals using optimization based multi-SVNN classifier", Journal of King Saud University - Computer and Information Sciences, 2018, DOI: 10.1016/j.jksuci.2018.11.012.
17. Srinivasa Reddy S., Suman M. .," Microaneurysm extraction with contrast enhancement using deep neural network ", 2018, Lecture Notes in Electrical Engineering ,Vol: 434,pp: 229- 238 ,DOI: 10.1007/978-981-10-4280-5_24.
18. Reddy S.S., Suman M., Prakash K.N. .," Micro aneurysms detection using artificial neural networks ", 2018, Lecture Notes in Electrical Engineering ,Vol: 471,pp: 273 -282 ,DOI: 10.1007/978-981-10-7329-8_28.
19. Bojja P., Sanam N., Design and development of artificial intelligence system for weather forecasting using soft computing techniques, ARPN Journal of Engineering and Applied Sciences, Vol:12, issue:3, 2017, pp: 685-689, ISSN: 18196608.
20. Vallabhaneni R.B., Rajesh V., Brain tumor detection using mean shift clustering and glcm features with edge adaptive total variation denoising technique, ARPN Journal of Engineering and Applied Sciences, Vol:12, issue:3, 2017,pp: 666-671, ISSN: 18196608.
21. Gattim N.K., Rajesh V., Partheepan R., Karunakaran S., Reddy K.N.,Multimodal image fusion using curvelet and genetic algorithm, Journal of Scientific and Industrial Research,Vol:76, issue:11, 2017 ,pp: 694-696, ISSN: 224456.

22. Bhavana D., Rajesh V., A new pixel level image fusion method based on genetic algorithm , Indian Journal of Science and Technology, Vol: 9, Issue: 45, 2016, pp: 1 - 8, ISSN 9746846.
23. Bhavana D., Rajesh V., Kumar K.K., Implementation of plateau histogram equalization technique on thermal images , Indian Journal of Science and Technology, Vol: 9, Issue: 32, 2016, pp: 1 - 4, ISSN 9746846.
24. Bhavana D., Rajesh V., Koteswara Rao C.H., Multispectral image fusion using integrated wavelets and principal component analysis , International Journal of Control Theory and Applications, Vol: 9, Issue: 34, 2016, pp: 737 - 743, ISSN 9745572.
25. Hari Priya D., Sastry A.S.C.S., Rao K.S., Low power cmos circuit design for R wave detection and shaping in ECG ,ARPN Journal of Engineering and Applied Sciences, Vol: 11, Issue: 24, 2016, pp: 14491 - 14496, ISSN 18196608.
26. Naveen Kishore Gattim, Dr. V. Rajesh, “Rotation and Scale Invariant Feature Extraction for MRI Brain Images”, Journal of Theoretical and Applied Information Technology, Vol.70 No.1, Dec 2014, Page 62-67, ISSN: 1817-3195.
27. Naveen Kishore Gattim, Dr. V. Rajesh, “Multimodal Medical Image Fusion under Redundant Transforms”, International Review On Computers and Software, Vol. 10, No. 3, March 2015, pp. 241-248. ISSN: 1828-6003.
28. Bennilo Fernandes, J et. al, "Fuzzy utilization in speech recognition and its different application", International Journal of Engineering and Advanced Technology (2019), 8 (5 Special Issue 3), pp. 261-266. DOI: 10.35940/ijeat.E1058.0785S319.
29. Bennilo Fernandes, J. et. al,"Reversible image watermarking technique using LCWT and DGT", International Journal of Engineering and Technology(UAE) (2019), 7 (1), pp. 42-47.

AUTHORS PROFILE

Kasiprasad Mannepalli, is a doctorate in Electronics & communication Engineering and currently working as an Associate Professor in the Department of ECE in Koneru Lakshmaiah Education Foundation (Deemed to be University), Vaddeswaram, Vijayawada, Guntur (dist), Andhra Pradesh, India. His research interests are speech signal processing, Artificial neural networks and pattern recognition. He worked in the area of speech signal processing for Telugu language speeches particularly for accent recognition and emotion recognition from Telugu speech. He has published his research articles in various reputed international journals and also presented papers in various national and international conferences. He is a review member for various international journals.

Pelluri Vamsi Krishna, is a Under Graduate student, currently Studying B-tech final year in the stream Electronics and communication engineering at Koneru Lakshmaiah Education Foundation (Deemed to be University), Vaddeswaram, Guntur (dist), Andhra Pradesh, India. His area of interest is signal processing and machine learning and deep learning.

Kovi Vamsi Krishna, is a Under Graduate student, currently Studying B-tech final year in the stream Electronics and communication engineering at Koneru Lakshmaiah Education Foundation (Deemed to be University), Vaddeswaram, Guntur (dist), Andhra Pradesh, India. His area of interest is signal processing and machine learning and artificial neural networks.

Kodali Rama Krishna, is a Under Graduate student, currently Studying B-tech final year in the stream Electronics and communication engineering at Koneru Lakshmaiah Education Foundation (Deemed to be University), Vaddeswaram, Guntur (dist), Andhra Pradesh, India. His area of interest is signal processing and audio Signal processing and pattern recognition.

.