Journal of Engineering Sciences
Vol 13, Issue 06, June/2022
ISSN NO:0377-9254

INTELLIGENT SYSTEM FOR THE ANALYSIS OF AUDIO

[1]Dr. Mohammad Riyaz [2]BelgaumEdiga [3]Divya Bathula Dharani [4]Yedlapalli Alekya [5]Dayala Jyothirmai

[1]Guide Assosciate Professor [2,3,4,5] U.G. SCHOLAR

G PULLAIAH COLLEGE OF ENGINEERING AND TECHNOLOGY KURNOOL

ABSTRACT

With the continuous rise in ingenious forgery, a wide range of digital audio authentication applications are emerging as a preventive and detective control in real-world circumstances, such as forged evidence, breach of copyright protection, and unauthorized data access. To investigate and verify, this paper presents a novel automatic authentication system that differentiates between the forged and original audio. The design philosophy of the proposed system is primarily based on three psychoacoustic principles of hearing, which are implemented to simulate the human sound perception system. Moreover, the proposed system is able to classify between the audio of different environments recorded with the same microphone. To authenticate the audio and environment classification, the computed features based on the psychoacoustic principles of hearing are dangled to the Gaussian mixture model to make automatic decisions. It is worth mentioning that the proposed system authenticates an unknown speaker irrespective of the audio content i.e., independent of narrator and text. To evaluate the performance of the proposed system, audios in multi environments are forged in such a way that a human cannot recognize them.

I. INTRODUCTION

Scope of the Project

Industrial accidents are quite fatal and can cause quite a loss. Those that occur in the workplace can cause harm to employees, environment and damage to the equipment. Industrial related accidents, injuries and fatality data demonstrate that continued efforts and effective measures are necessary to reduce the number of industrial accidents, illnesses and fatalities. A worker dies of occupational injury every three minutes and about every second at least four workers get injured according to the International Labor Organization (ILO).

Objective

The main objective is With the continuous rise in ingenious forgery, a wide range of

digital audio authentication applications are emerging as a preventive and detective control in real-world circumstances, such as forged evidence, breach of copyright protection, and unauthorized data access. To investigate and verify, this paper presents a novel automatic authentication system that differentiates between the forged and original audio.

About the project

With the recent unprecedented proliferation of smart devices such as mobile phones and advancements in various technologies (e.g., mobile and wireless networks), digital multimedia is becoming an indispensable part of our lives and the fabric of our society. For example, unauthentic and forged multimedia can in_uence the decisions of courts as it is admissible evidence. With continuous advancements in ingenious forgery, the authentication of digital multimedia (i.e., image, audio and video) is an emerging challenge. Despite reasonable advancements in image and video, digital audio authentication is still in its infancy. Digital authentication and forensics involve the verification and investigation of an audio to determine its originality (i.e., detect forgery, if any) and have a wide range of applications. For example,the voice recording of an authorized user can be replayed or manipulated to gain access to secret data. Moreover, it can be used for copyright applications such as to detect fake MP3 audio. Audio forgery can be accomplished by copy-move,deletion, insertion, substitution and splicing . The applications of copy-move forgery are limited compared with other methods as it involves moving a part of the audio at other location in the same liaison. On the other hand, the deletion, insertion, substitution and splicing of forged audio may involve merging recordings of different devices, speakers and environments. This paper deals with a splicing forgery (i.e., insertion of one or more segments to the end or middle), which is more challenging. The primary objective of the proposed system is to address the following issues with high accuracy and a good classification rate: _ Differentiate between original and tampered audio generated by splicing recordings with the same microphone and different environments. Environment classification of original and forged audio generated through splicing. Identify forged audio irrespective of content (i.e., text) and speaker. _ Reliable authentication with forged audio of a very short duration (i.e., _5 seconds). In the past,

audio authentication has been achieved by applying various algorithms .One of the basic approaches is the visual investigation of the waveform of an audio to identify irregularities and discontinuities. For example, the analysis of spectrograms may reveal irregularities in the frequency component during the investigation. Similarly, listening to audio may also disclose abrupt changes and the appearance of unusual noise. These methods may help to decide whether the audio is original or tampered. However, one of the prime limitations of These approaches is that they are human-dependent, where judgement errors cannot be ignored. Moreover, the availability of sophisticated manipulation tools makes it convenient to manipulate audio without introducing any abnormalities. Consequently, it becomes very difficult to identify those abnormalities

II. LITERATURE SURVEY

2.1 An Overview on Image Forensics

AUTHORS: A.Piva

ABSTRACT: The aim of this survey is to provide a comprehensive overview of the state of the art in the area of image forensics. These techniques have been designed to identify the source of a digital image or to determine whether the content is authentic or modified, without the knowledge of any prior information about the image under analysis (and thus are defined as passive). All these tools work by detecting the presence, the absence, or the incongruence of some traces intrinsically tied to the digital image by the acquisition device and by any other operation after its creation. The paper has been organized by classifying the tools according to the position in the history of the digital image in which the relative footprint is left: acquisition-based methods, coding-based methods, and editing-based schemes.

2.2 Authentication of Scalable Video Streams With Low Communication Overhead

AUTHORS: K. Mokhtarian and M. Hefeeda,

ABSTRACT: The large prevalence of multimedia systems in recent years makes the security of multimedia communications an important and critical issue. We study the problem of securing the delivery of scalable video streams so that receivers can ensure the authenticity of the video content. Our focus is on recent scalable video coding (SVC) techniques, such as H.264/SVC, which can provide three scalability types at the same time: temporal, spatial, and visual quality. This three-dimensional scalability offers a great flexibility that enables customizing video streams for a wide range of heterogeneous receivers and network conditions. This flexibility, however, is not supported by current stream authentication schemes in the literature. We propose an efficient and secure authentication scheme that accounts for the full scalability of video streams, and enables verification of all possible substreams that can be extracted from the original stream. In addition, we propose an algorithm for minimizing the amount of authentication information that need to be attached to streams. The proposed authentication scheme supports end-to-end authentication, in which any third-party entity involved in the content delivery process, such as stream adaptation proxies and caches, does not have to understand the authentication mechanism. Our simulation study with real video traces shows that the proposed authentication scheme is robust against packet losses, incurs low computational cost for receivers, has short delay, and adds low communication overhead. Finally, we implement the proposed authentication scheme as an open source library called svcAuth,

which can be used as a transparent add-on by any multimedia streaming application.

2.3 Current Developments and Future Trends in Audio Authentication

AUTHORS: S. Gupta, S. Cho, and C. C. J. Kuo

ABSTRACT: Recent developments in the audio-authentication field include basic, preliminary audio analysis and advanced audio- authentication techniques that exploit audio recording conditions and compressed audio features. Multimedia is involved in every aspect of our lives, and many of us rely on various websites for information on events taking place all over the world. We form opinions based on the content of camera footage, phone conversations, and other recordings. Although some sources give us authentic information, others contain forged content. Concerns regarding how to authenticate multimedia data have led to research in forgery detection, but studies in audio are generally still limited compared to those in image and video.

2.4 Copy-move detection of audio recording with pitch similarity

AUTHORS: Q. Yan, R. Yang, and J. Huang

ABSTRACT: The widespread availability of audio editing software has made it very easy to create forgeries without perceptual trace. Copy-move is one of popular audio forgeries. It is very important to identify audio recording with duplicated segments. However, copy-move detection in digital audio with sample by sample comparison is invalid due to post-processing after forgeries. In this paper we present a method based on pitch similarity to detect copy-

move forgeries. We use a robust pitch tracking method to extract the pitch of every syllable and calculate the similarities of these pitch sequences. Then we can use the similarities to detect copy-move forgeries of digital audio recording. Experimental result shows that our method is feasible and efficient.

2.5 Detecting splicing in digital audios using local noise level estimation

AUTHORS: X. Pan, X. Zhang, and S. Lyu

ABSTRACT: One common form of tampering in digital audio signals is known as splicing, where sections from one audio is inserted to another audio. In this paper, we propose an effective splicing detection method for audios. Our method achieves this by detecting abnormal differences in the local noise levels in an audio signal. This estimation of local noise levels is based on an observed property of audio signals that they tend to have kurtosis close to a constant in the band-pass filtered domain. We demonstrate the efficacy and robustness of the proposed method using both synthetic and realistic audio splicing forgeries.

III. SYSTEM ANALYSIS

3.1 EXISTING SYSTEM:

- In the past, audio authentication has been achieved by applying various algorithms. One of the basic approaches is the visual investigation of the waveform of an audio to identify irregularities and discontinuities

  . For example, the analysis of spectrograms may reveal irregularities in the frequency component during the investigation. Similarly, listening to audio may also disclose abrupt changes and the appearance of unusual noise. These methods may help to

Vol 13, Issue 06, June/2022
ISSN NO:0377-9254

decide whether the audio is original or tampered. However, one of the prime limitations of These approaches is that they are human-dependent, where judgment errors cannot be ignored. Moreover, the availability of sophisticated manipulation tools makes it convenient to manipulate audio without introducing any abnormalities. Consequently, it becomes very difficult to identify those abnormalities.

3.1.1 DISADVANTAGES OF EXISTING SYSTEM:

- The visual inspection of the waveform and spectrogram of the tampered audio depicted does not provide any clue of irregularity and hearing is also quite normal.

3.2 PROPOSED SYSTEM:

The proposed system is able to classify between the audio of different environments recorded with the same microphone. To authenticate the audio and environment classification, the computed features based on the psychoacoustic principles of hearing are dangled to the Gaussian mixture model to make automatic decisions. It is worth mentioning that the proposed system authenticates an unknown speaker irrespective of the audio content i.e., independent of narrator and text. To evaluate the performance of the proposed system, audios in multi environments are forged in such a way that a human cannot recognize them. Subjective evaluation by three human evaluators is performed to verify the quality of the generated

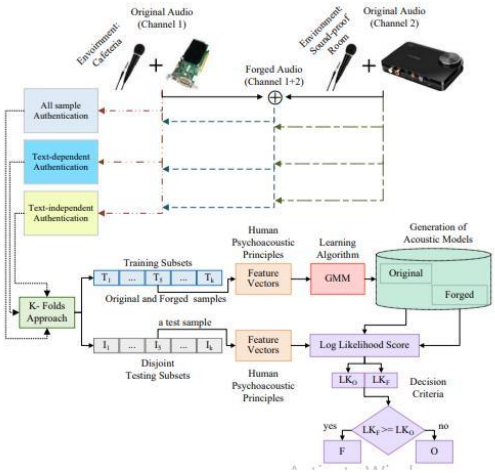
forged audio. The proposed system provides a classification accuracy of 99.2% ± 2.6. Furthermore, the obtained accuracy for the other scenarios, such as text-dependent and text-independent audio authentication, is 100% by using the proposed system.

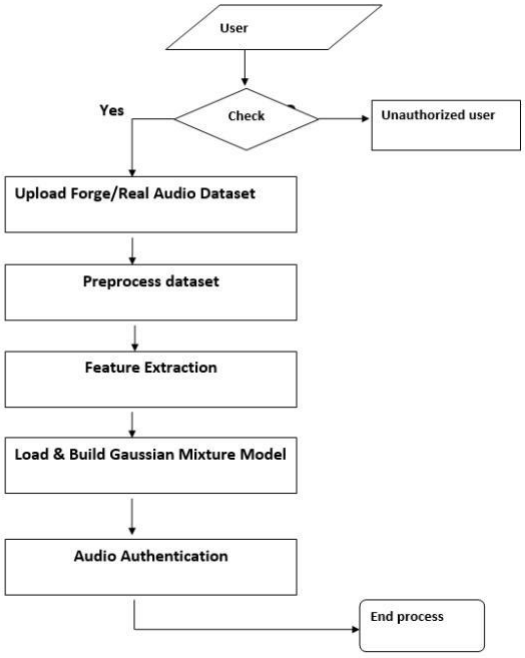
3.2.1 ADVANTAGES OF PROPOSED SYSTEM:

- The system is also evaluated by using each recording of the developed forged.

IV. SYSTEM DESIGN

4.1 SYSTEM ARCHITECTURE:





V. Modules

Upload Forge/Real Audio Dataset

Using this model user Upload dataset

Preprocess dataset

Using this module dataset is preprocess here.

Feature Extraction

Using this module data is normalized here.

Load & Build Gaussian Mixture Model

Using this module data is loaded and Build Gaussian Mixture Model.

Audio Authentication

Using this module audio authentication here.

VI. Algorithms

Generation of Forged Audio Corpus

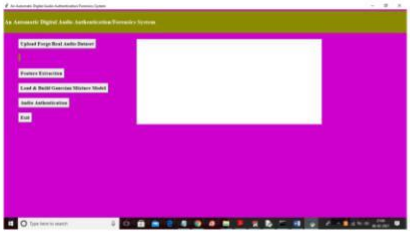
The generation of tempered audio, in a way that a human evaluator cannot guess it is so, is a big challenge and one of the crucial steps towards the development of the proposed system. Forged and normal audio samples are generated by using the King Saud University Arabic Speech Database (KSU-ASD) [24]. The reason for selecting the KSU-ASD is its diversity in recorded text, recording environments and equipment [25, 26]. To the best of our knowledge, none of the existing publicly available databases serves our purpose. The KSU-ASD is publicly available through the Linguistic Data Consortium, which is hosted by the University of Pennsylvania, Philadelphia, USA. Although the language of the KSU-ASD is Arabic, the proposed system will work for any language

Generation of Forged Audio by Splicing

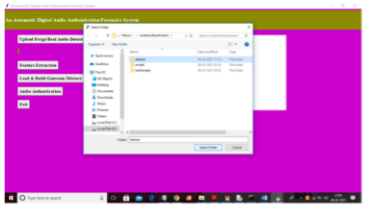
The KSU speech database was recorded in three different environments i.e., office (normal), cafeteria (noisy) and soundproof room (quiet). In this study, two very different environments, cafeteria and sound-proof room, are mixed to generate the forged audio. Mixing the sound-proof room with the cafeteria is the worst case scenario, where the former represents an absolutely quiet environment and the latter represents a noisy environment containing background noise. Audio is forged by mixing the speech of two different recording settings: 1. Recording of digits in the cafeteria with a microphone (Sony F-V220) attached to a built-in sound card on the desktop

(OptiPlex 760) through an audio-in jack. This is denoted as CDMB (Cafeteria, Digits, Microphone, Builtin sound card). 2. Recording of digits in the sound-proof room with a microphone (Sony F-V220) connected to an external sound card (Sound Blaster X-Fi Surround 5.1 Pro) through the USB port of the desktop (OptiPlex 760). This is represented by SDME (Sound-proof room, Digits, Microphone, External sound card). Although it is ideal to forge an audio recording through a mobile phone because a person is unaware of the recording in such a scenario, his/her speech can be used for any purpose. However, the amplitude of mobile phone recordings is low in the KSU-ASD compared with the microphones, and through visualization, it is easy to identify that the audio is forged. Therefore, the recording of the mobile phone is not used to generate forged samples

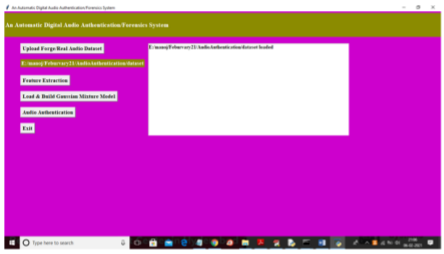
VII. RESULTS



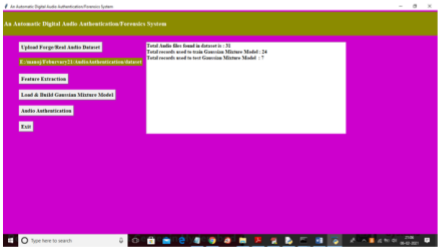
In above screen click on 'Upload Forge/Real Audio Dataset' button and then upload dataset folder



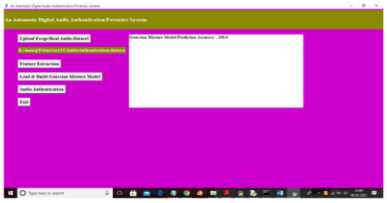
In above screen selecting and uploading 'dataset' folder and then click on 'Select Folder' button to load dataset and to get below screen



In above screen dataset loaded and then click on 'Feature Extraction' buttor to read audio files and to extract features



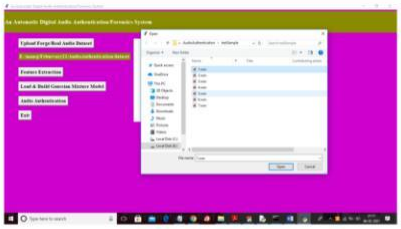
In above screen all audio files features extracted and then application find total 31 audio files and the using 24 files to train GMM and 7 to test GMM. Now dataset ready with train and test records and now click on 'Load & Build Gaussian Mixture Model' button to train GMM model and calculate prediction accuracy.



In above screen GMM model generated and its prediction accuracy is 100% on test data and now click on 'Audio Authentication' button to upload new test file and perform prediction

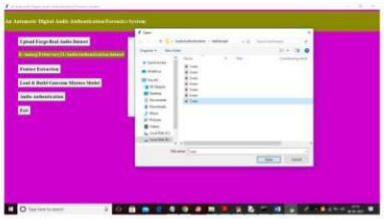
In above screen text area we can see predicted result as 'uploaded file contains REAL audio' and you can see difference in above 2 graphs for forge and real. In forge graph due to tamper lots of fluctuation is there in red line and in second graph many fluctuations not there.

Similarly you can upload other files and test



In above screen selecting and uploading '1.wav' file and then click on 'Ope button to predict its content

In above screen text area we got predicted result as uploaded file contains 'FORGE' audio and then displaying audio features in graph and now test with other file



In above screen selecting and uploading '7.wav' file and then click on 'Open' button to get below result

VIII. CONCLUSION

This paper proposed an automatic audio authentication system based on three human psychoacoustic principles. These principles are applied to original and forged audio to obtain the feature vectors, and automatic authentication is performed by using the GMM. The proposed system provides 100% accuracy for the detection of forged and audio in both channels. The channels have the same recording microphone but different recording environments. Moreover, an accuracy of 99% is achieved for the classification of the three different environments. In automatic systems based on supervised learning, the audio text is vital. Therefore, both the text dependent and the text-

independent evaluation of the proposed system is performed. The maximum obtained accuracy is 100%. In all experiments, the speakers used to train and test the system are different (i.e., speaker-independent) and the obtained results are reliable, accurate and significantly outperform the subjective evaluation. The lower accuracy in the subjective evaluation also confirms that the forged audios are generated so sophisticatedly that human evaluators are unable to detect the forgery.

IX. FUTURE SCOPE

In Future Work we will study other feature selection methods combined with more machine learning algorithms applied to real-time data from IoT devices.

X. REFERENCES

Journals:

[1] B. B. Zhu, M. D. Swanson, and A. H. Tewfik, "When seeing isn't believing [multimedia authentication technologies]," IEEE Signal Processing Magazine, vol. 21, pp. 40-49, 2004.

[2] A. Piva, "An Overview on Image Forensics," ISRN Signal Processing, vol. 2013, p. 22, 2013.

[3] A. Haouzia and R. Noumeir, "Methods for image authentication: a survey," Multimedia Tools and Applications, vol. 39, pp. 1-46, 2008.

[4] K. Mokhtarian and M. Hefeeda, "Authentication of Scalable Video Streams With Low Communication Overhead," IEEE Transactions on Multimedia, vol. 12, pp. 730-742, 2010.

[5] S. Gupta, S. Cho, and C. C. J. Kuo, "Current Developments and

Future Trends in Audio Authentication," IEEE MultiMedia, vol. 19, pp. 50- 59, 2012.

[6] R. Yang, Y.-Q. Shi, and J. Huang, "Defeating fake- quality MP3," presented at the Proceedings of the 11th ACM workshop on Multimedia and security, Princeton, New Jersey, USA, 2009.

[7] Q. Yan, R. Yang, and J. Huang, "Copy-move detection of audio recording with pitch similarity," in 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2015, pp. 1782-1786.

[8] X. Pan, X. Zhang, and S. Lyu, "Detecting splicing in digital audios using local noise level estimation," in 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2012, pp. 1841-1844.

[9] A. J. Cooper, "Detecting Butt-Spliced Edits in Forensic Digital Audio Recordings," in 39th International Conference: Audio Forensics: Practices and Challenges, 2010.

[10] D. Campbell, E. Jones, and M. Glavin, "Audio quality assessment techniques—A review, and recent developments," Signal Processing, vol. 89, pp. 1489-1500, 8// 2009.

[11] R. C. Maher, "Overview of Audio Forensics," in Intelligent Multimedia Analysis for Security Applications, H. T. Sencar, S. Velastin, N. Nikolaidis, and S. Lian, Eds., ed Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 127-144.

[12] B. E. Koenig and D. S. Lacey, "Forensic Authentication of Digital Audio Recordings," Journal of Audio Engineering Society, vol. 57, pp. 662-695, 2009.

[13] Audacity Team, "Audacity(R): Free Audio Editor and Recorder. Version 2.1.2 retrieved on November 25, 2016 from

http://www.audacityteam.org/.", ed, 2016.

[14] GoldWave Inc., "GoldWave: Digital Audio Editing Software. Version 6.24 Retrived on November 25, 2016 from https://www.goldwave.com/goldwave.php," ed, 2016. [

15] C. Kraetzer, A. Oermann, J. Dittmann, and A. Lang, "Digital audio forensics: a first practical evaluation on microphone and environment classification," presented at the Proceedings of the 9th workshop on Multimedia & security, Dallas, Texas, USA, 2007.