International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 4, Issue 10, ISSN No. 2455-2143, Pages 216-220
Published Online February 2020 in IJEAST (http://www.ijeast.com)

CHALLENGES AND ISSUES OF SOUND ARCHIVES FOR ENVIRONMENTAL SOUND CLASSIFICATION

Dr. S. Veena
Professor,
Department of CSE
S.A. Engineering College,
Thiruverkadu, Chennai,
Tamil Nadu, India.

Nerisai.M.V
UG Student,
Department of CSE
S.A. Engineering College,
Thiruverkadu, Chennai,
Tamil Nadu, India.

Remya.J.V
UG Student,
Department of CSE
S.A. Engineering College,
Thiruverkadu, Chennai,
Tamil Nadu, India.

Sai Tejah.S
UG Student,
Department of CSE
S.A. Engineering College,
Thiruverkadu, Chennai,
Tamil Nadu, India.

ABSTRACT:- **Sound is the most important product of natural activities. Each sound has variations and uniqueness of its own. But separating them from the mixed sound environment is a difficult task. The present paper discusses the activities of the researches to identify and classify these sounds by using many techniques. Analysis of these help to motivate the researchers to find new ways and means for identification of each sound, which may help in various fields such as hearing impairment treatment, criminal activities prevention, forensic science, humanoid robots.**

Keywords: Sound, Classification, Archives, Data sets

I. INTRODUCTION

Every day we hear sounds or noises around us. The sounds that we hear are basically classified into two types: Indoor sound and Outdoor sound. Under each of these categories, there are large different numbers of subsets that are used for classification. Here, in this paper, we are going to overview the available sound archives that is going to help classify the detected sound. Though music and sound are vast categories that have multiple variations on each class, the sound archives have managed to take the sound parameters which give utmost precision when used for classification. Every sound archive mentioned in this paper has its uniqueness as not all archives work efficiently on all the algorithms. Also, it can be viewed that few archives are soulfully used for research purposes and few are dedicated for surveillance purposes.

II. EXISTING DATASETS

Antonio Greco, Alessia Saggese, Mario Vento and Vincenzo Vigilante[1] did their project using Mivia audio events archives. It is the only benchmarking sound archive used for surveillance application purposes. It composes of .wav extension files whose overall duration is 30 hours. The audio clips were recorded with the help of "Axis P8221 Audio Module" and an "Axis T83 omnidirectional microphone". This Mivia audio event archive was used to extract about 83,000 spectrogram images in which 50,000 of them were used for training and 33,000 for testing.

Takumi Kobayashi and Jiaxing Ye[2] have used RWCP as their sound archive since it contains enough classes of sounds that can give around 98.62% of precision as well as robustness to noise and low computation time. It contains 9,722 sound clips and each of its signal strength is 48 KHz with 16 bit resolution. The RWCP sound archive is recorded in clean setting suppressing noise. As the environmental audio was neither stationery nor well structured, RWCP sound archive proved to overcome this obstacle by providing accurate results.

Dan Stowell and Mark D. Plumbley[3] aimed to reflect the general audio archive and then group the rest of the collection as "field recording". There were 10 subsets for each of the classifier and they were: water, voice, nature, people, bird song, train, city, etc., this states that for each fold, the authors have used nine of the subsets as training data for the classifier. Hence, the presence/absence of the subset was the binary attribute to be learned and then to be tested by the classifier using the audio from that one remaining subset.

J. Salamon, C. Jacoby, and J. P. Bello[4] chose Urban Sound archive as it contained urban sounds with high frequency and often appear in noise complaints in the urban environment. It holds audio of duration 27 hours out of which18.5 hours have annotated sound events in the 10 classes used. The sound archive contains a subset called UrbanSound8k. With this archive, it is easy for the subjects to recognize the environmental sounds in 4 seconds in a listening test with 82% accuracy.

International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 4, Issue 10, ISSN No. 2455-2143, Pages 216-220
Published Online February 2020 in IJEAST (http://www.ijeast.com)

Edurado Fonseca, Jordi Pons, Xavier Favory, Frederic Font, Dmitry Bogdanov, Andres Ferraro, Sergio Oramas, Alastair Porter and Xavier Serra[5] have used free sound as it is an open audio archive available in online that is transparent, organized and sustained. The transparency of the sound archive helps with the workflow of creating the archive process as the users of the sound archive are aware of it. Having a dynamic character for sound archive will be useful as the community can contribute and update the archive easily.

Peter Foster, Siddharth Sigtia, Sacha Krstulovic, Jon Barker and Mark D Plumbley[6] used Chime-Home sound archive in order to recognize the sounds. This archive contains Domestic environment sounds of 6.8 hours duration and consists of muti-labeled annotations of sound recordings. The sound archives are refined by perpetuating the chunks where two or more annotators have been allocated a given label for all the labels that are considered. The refined sound archive and raw multiple annotators are made available in this sound archive.

The authors of Google Inc Research Team[7] created the Audio set sound archive manually. It considers all types of sound events rather than a constrained environment. The main objective of this sound archive according to the team is that to provide comprehensive coverage of real world sounds at image Net scale. The events were segregated and structured using the abstraction hierarchy. As the events were categorized, they were collectively grouped and named as Ontology.

According to the authorK. J. Piczak[8] the compiled sound archive used is made of three parts and one of them is the main sound archive that contains 50 different environmental sound classes. The ESC-50 sound archive holds over 2000 environmental sounds that are categorised in 50 classes. The sound archive can also be expanded if needed.

ESC-10 is a proof-of-concept sound archive which expects higher accuracy and achieved accuracy of 95.7% (human classification). And based on the 50 classes in ESC-50, the archive are separated in 10 major ways. This dataset provides common and distinct sounds. It maintains quality control with the help of CrowdFlower's procedures.

Xiaohu Zhang, Yuexian Zou and Wei Shi[9] has used CICESE sound archive which is to classify environmental sound among the indoors. One of the freely available archive specially used for research needs. It is a compact, lightweight sound class representation with good generalization properties. To some extent it is able to identify sound mixes without modelling background/foreground noises. The final representation of sound does not depend on the length of the input sound.

III. COMPARISON OF DATA SETS

S.NO	TITLE	DATASET USED	ADVANTAGES	DISADVANTAGES
1	SoReNet: A novel deep network for audio surveillance applications	Mivia Audio Events	It achieves accuracy of 88.9% when Mivia is used for Fine Tuning.	It cannot find the difference between screaming and communication between people.
2	Acoustic Feature Extraction by Statistics based local Binary pattern for Environmental Sound Classification	RWCP	Using RWCP, the proposed method exhibited the state-of-the-art performance compared to other methods, demonstrating the robustness to noise and low computation time.	Since there are many categories of the same sound, classification is quite challenging.
3	An open dataset for research on audio field recording archives: freefield1010	FreeField 1010	This dataset is used in field recording of audio archives related to data mining. For the given audio content, this dataset is of sufficient size to makes a relative predictability of mentioned tags (class).	In the crowd sourced Freesound archive, amplitude levels are uncontrollable that may cause trouble for listening tests.

4	A Dataset and Taxonomy for Urban Sound Research	UrbanSound 8K	Believed to be the taxonomy representation of urban sound in concise manner can be helpful in expanding and reformulating the taxonomy as we increase the scope covered by the researchers	The challenge is to identifying sound sources in the presence of (real) background noise which creates confusion due to timbre similarity (especially for noise-like continuous sounds), and sensitivity to background interference.
5	FREESOUND DATASETS: A PLATFORM FOR THE CREATION OF OPEN AUDIO DATASETS	FreeSound (FSD)	FSD includes a mixture of strongly and weakly labeled data whereas in AudioSet only weakly labeled data is provided.	Sounds in Freesound can widely vary in length, so need to filter out all samples longer than 90s, which left with a total of 268,261 candidate samples from more than 300,000.
6	CHIME-HOME: A DATASET FOR SOUND SOURCE RECOGNITION IN A DOMESTIC ENVIRONMENT	Chime	The dataset comprises multi-label annotations of audio recordings, obtained with the aid of multiple annotators	The dataset restricts to a single domestic Environment, thus does not permit evaluation of how models generalise to other environments.
7	AUDIO SET: AN ONTOLOGY AND HUMAN-LABELED DATASET FOR AUDIO EVENTS	Audioset	The data can accelerate all acoustic distinctions made by a 'typical' listener which helps in the research area of acoustic event detection just like ImageNet.	Not every dataset have been included out of632 categories, 56 areblacklisted, meaning they are not exposed to labelers because they have turned out to be obscure (e.g. Alto saxophone) or confusing (e.g.Sounds of things).Therefore the resulting dataset includes 1,789,621 segments (4,971 hours), comprising at least 100 instances for 485 audio event categories. The remaining categories are either excluded or difficult to it find using our current approaches
8	ESC: Dataset for Environmental Sound Classification	ESC – 50	All the recorded sounds are equally balanced in the classes.	This dataset is bounded to have fewer amounts of images in each of the classes.
9	ESC: Dataset for Environmental Sound	ESC - 10	The classes have less obscurity and more distinct	Since the classification of sounds is done based on

	Classification		and also it can be used for various machine learning techniques.	the human's point of view which is a little insignificant, the accuracy that is expected from the sound is high.
10	Dilated Convolution Neural Network with LeakyReLU for Environment sound classification	CICESE	It gives the state of the art results and shows minimum error.	If the choice of parameter is not given properly then there are more chances that the output can give a worst case scenario.

IV. CONCLUSION

The paper had discussed on the activities of the researches to identify and classify the various sounds by using many techniques. Classifying the sound in real-time is still a research based topic which is in its infant stage. To collect these data and to classify them itself is a huge process and needs the right amount of technical knowledge to do so. Every sound clip that is considered for a collection undergoes compression of signals in a certain bitrate and each signal are processed to find out what sound it contains.

V. ACKNOWLEDGEMENT

We would like to thank our Management, Advisor and Principal of S.A.Engineering College for constantly motivating us to do research in our work environment and encourages us to publish more research papers.

VI. REFERENCES

[1] Antonio Greco, Alessia Sagesse, Mario Vento and Vincenzo Vigilante, "SoReNet: A novel deep network for audio surveillance applications," in IEEE International Conference on Systems, Man and Cybernetics (SMC) on October 6-9 in 2019 in Bari,Italy.

[2] Takumi Kobayashi and Jiaxing Ye, "Acoustic feature extraction by statistics based local binary pattern for environmental sound classification," in Proc. ICASSP, IEEE International Conferemce on Acoustics, Speech and Signal Processing in 2014.

[3] D. Stowell and M. D. Plumbley, "An open dataset for research on audio field recording archives: freefield1010," in Proc. Audio Engineering Society 53rd Intern. Conf.: Semantic Audio,

[4] J. Salamon, C. Jacoby, and J. P. Bello, "A dataset and taxonomy for urban sound research," in Proc. 24th ACM Intern. Conf. Multimedia, 2014, pp. 1041–1044.

[5] Edurado Fonseca, Jordi Pons, Xavier Favory, Frederic Font, Dmitry Bogdanov, Andres Ferraro, Sergio Oramas, Alastair Porter and Xavier Serra, "Freesound Datasers: A Platform for the creation of Open Audio Datasets," in Proceedings of the 18th ISMIR Conference in 2017, Suzhou, China.

[6] Peter Foster, Siddharth Sigtia, Sacha Krstulovic, Jon Barker, and Mark D Plumbley. "Chime-home: A dataset for sound source recognition in a domestic environment." In Workshop on Applications of Signal Processing to Audio and Acoustics, pages 1–5. IEEE, 2015.

[7] Jort F Gemmeke, Daniel PW Ellis, Dylan Freedman, Ared Jansen, Wade Lawrene, R Channing Moore, Manoj Plakal and Marvin Ritter, "Audio set: An ontology and human-labeled dataset for audio events. In proceedings of the Acoustics, Speech and Signal Processing" International Conference, 2017.

[8] K. J. Piczak, "ESC: Dataset for environmental sound classification," in Proceedings of the ACM International Conference on Multimedia. ACM, 2015, in press.

[9] Xiaohu Zhang, Yuexian Zou and Wei Shi, "Dilated Convolution Neural Network with LeakyReLU for Environment sound classification," in2017 22nd International Conference on Digital Signal Processing (DSP).

[10] Bob L Sturm. The gtzan dataset: Its contents, its faults,their effects on evaluation, and its future use. arXivpreprint: 1306.1461, 2013.

[11] S. Chu, S. Narayanan, and C.-C. Kuo. Environmentalsound recognition with time-frequency audio features.IEEE TASLP, 17(6):1142–1158, 2009.

[12] T. Bertin-Mahieux, D. P. W. Ellis, B. Whitman, andP. Lamere. "The Million Song Dataset. In Proceedingsof the 12th International Conference on Music Information Retrieval" (ISMIR-11), pages 591–596, Miami,FL, USA, Oct 2011.

[13] D. Barchiesi, D. Giannoulis, D. Stowell, and M. D. Plumbley,"Acoustic scene classification: Classifying environments fromthe sounds they produce," IEEE Signal Processing Magazine,2015.

[14] S. Chu, S. Narayanan, and C.-C. J. Kuo, "Environmentalsound recognition with time–frequency audio features," IEEETrans. Audio, Speech, and Language Processing, vol. 17, no.6, 2009.

[15] M. D. Plumbley, S. A. Abdallah, J. P. Bello, M. E. Davies,G. Monti, and M. B. Sandler, "Automatic music transcriptionand audio source separation," Cybernetics & Systems, vol. 33,no. 6, pp. 603–627, 2002.

[16] B. Defreville, F. Pachet, C. Rosin, and P. Roy, "Automatic recognition of urban sound sources," in Proc. 120th Audio Engineering Society Convention, 2006.

[17] G. Lagrange, M. Lafay, M. Rossignol, E. Benetos, andA. Roebel, "An evaluation framework for event detection using a morphological model of acoustic scenes," arXiv preprintarXiv:1502.00141, 2015.

[18] R. Radhakrishnan, A. Divakaran, and P. Smaragdis, "Audio analysis for surveillance applications." In IEEEWASPAA'05, pages 158–161, 2005.

[19] D. P. W. Ellis, X. Zeng, and J. H. McDermott, "Classifying soundtracks with audio texture features," In IEEE ICASSP, pages 5880–5883, 2011.

[20] Brian McFee, Eric Humphrey, and Julian Urbano," A plan for sustainable mir evaluation," In Proceedings of the International Society for Music Information Retrieval Conference, pages 285–291, 2016.

[21] Selina Chu, Shrikanth Narayanan, and C.-C. Jay Kuo, "Environmental sound recognition with time-frequency audio features," in Transactions on Audio, Speech, and Language Processing, 17(6):1142–1158, August 2009.

[22] Anurag Kumar and Bhiksha Raj. "Audio eventand scene recognition: A unified approach using strongly and weakly labeled data". arXiv preprint arXiv:1611.04871, 2016.

[23] Giuseppe Bandiera, Oriol Romani Picas, HiroshiTokuda, Wataru Hariya, Koji Oishi, and Xavier Serra."Good-sounds.org: A framework to explore goodness ininstrumental sounds," In International Society for Music Information Retrieval Conference, pages 414–419,2016.