ENF Extraction of Bangladesh Grid and Identification Efficiency Comparison

Riyasat Ohib, Samin Yeasar Arnob, MD. Muhtady Muhaisin, Tanzil Bin Hassan

Department of EEE, Islamic University of Technology
Board Bazar, Gazipur-1704, Bangladesh.

Abstract— **The Electric Network Frequency (ENF) is the supply frequency of power distribution networks, which can be captured by multimedia signals recorded near electrical activities. It normally fluctuates slightly over time from its nominal value of 50 Hz/60 Hz. The ENF remain consistent across the entire power grid. This has led to the emergence of multiple forensic application like estimating the recording location and validating the time of recording. Recently an ENF based Machine Learning system was proposed which infers the region of recording from the power grids ENF signal extracted from the recorded multimedia signal, with the help of relevant features. In this work, we report ENF data extraction process from Power Recording data of Bangladesh National Grid and used the ENF data to compare available power recordings data from online database [10]. We used a set of features from [7] which serve as identifying characteristics for detecting the region of origin of the multimedia recording. We used those characteristics in a multiclass machine learning implementation based on MATLAB which is able to identify the grid of the recorded signal.**

Keywords—Electric Network Frequency; Grid; Forensics; Estimation; SVM; Fine Gaussian; Machine Learning.

I. INTRODUCTION (*HEADING 1*)

The supply frequency of the power distribution network i.e. ENF has a nominal value of 50/60 Hz. The instantaneous frequency fluctuates about the nominal value due to the load control mechanisms and the changes in the load demands within the power grid. The fluctuations of the ENF although random, are unique within a particular electrical network [1]. The tendency of these variations, at a particular time, are almost same throughout the same grid. These variations of the ENF over time is defined as the ENF signal.

The use of ENF signal to identify modified audio recordings was proposed by Grigoras [2-5] presently being used in multimedia forensics applications, as it gets embedded in the multimedia recordings made in the vicinity of electrical activity. The ENF signals can also be used to identify and classify video signals as well [6]. The audio recordings can pick up the signals due to the mechanical or acoustic hums or electromagnetic interferences from the power lines. The clean power recordings can be extracted using an audio recorder connected directly to the power mains via a step-down transformer [7]. Applying a Band-Pass Filter around the nominal frequency and employing a frequency estimation algorithm, the dominant frequency surrounding the nominal frequency can be estimated frame-by-frame, thus forming the ENF signal.

The ENF signals from a particular grid although random, vary from those of other grids in their nature and manner of variations. In our work, we have extracted Power recordings with the help of sensing circuit from National Power Grid of Bangladesh and used it alongside the other grids available [10]. Using statistical features derived from the ENF signals, already discussed in [8], the region of recording of the media signals were differentiated based on a Machine Learning approach.

II. ENF EXTRACTION AND DATABASE DESCRIPTION

The In this section we describe the database of Power recordings. Subsequently, we discuss the procedure and methodology adopted, to extract the ENF signal from the recordings. We end this section by analysing the similarities and dissimilarities between the extracted ENF signals, and discussing the statistical features that might be used as discerning features for the classification of the ENF signals.

A. Database Description and Feature Extraction

Media recordings of 12 different grid from all over the world were available on an online database [10] published by the authors of [8] which is subset of the data used in their work and the whole dataset was not published by the authors. However the ENF extraction code [24] used by the authors of [8] was available, therefore using their extraction algorithm and code we performed feature extraction process from our ENF data.

B. Hardware Implementation

A circuit has been built to collect the power recordings from the Bangladeshi power grid for analyzing the ENF variation present. The setup was built with least cost possible while not compromising the efficiency and accuracy. After completing the circuit, more than ten hours of reference power recording were collected at different times of the day in different days of a week. With the help of a voltage divider, 6V was converted into 200mv (p-p). We avoided the use of passive circuit elements while building the setup, so that unnecessary frequency fluctuation can be avoided. This setup was connected with the sound card of a Computer. Thus the originally transmitted electrical signal was recorded with an audio recorder software for ENF extraction using our developed program.

C. *Methodology*

To obtain the power data of the grid, 220 V was stepped down to 6 V with a transformer. Then, to bring the voltage level down to the low acceptable range of the sound card of the computer, we resorted to a simple voltage divider circuit. The 6 V transformer output was converted into 200 mV (p-p), and connected to the soundcard of a computer. Bearing in mind the change of load conditions at different times of the day, we collected power recordings during mornings, afternoon, midnight and early mornings of different days of the week. The components those are being use are:

- One 220V-6V step-down transformer
- Resistors of value 100kΩ & 10kΩ
- One 3.5mm headphone jack
- Jumpers
- Sound Card of a laptop

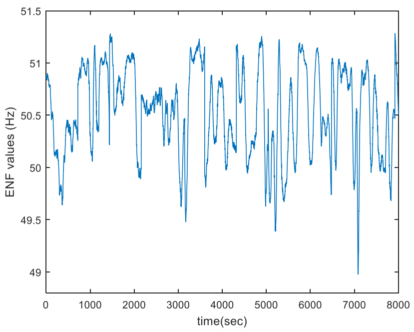

Fig. 1. ENF variation of collected signal through our sensing hardware.

We have collected more than ten hours of power recordings from the Electric mains from Gazipur District of Bangladesh. Main recording of each power file was at first divided into segments of ten minutes to create substantial number of datasets so that we can train our program properly for future prediction. Each of the segments were divided into non-overlapping frames of 5 seconds duration & then Spectrum Combining method [11] was used for estimating the dominant frequency component of each time frame, thus generating the ENF signal segment of length S = 120 samples.

D. *Analysis of Recorded ENF signal*

After extracting the ENF signals from the power recordings using our program, we noticed some similarities and dissimilarities of the signal from the ENF signals of the power recordings of Grids provided to us. We immediately notice the high dynamic range of the ENF signal collected from Bangladesh in Figure 1. The ENF signal fluctuated from around 49.5Hz to over 51Hz for most of its duration Bangladeshi Grid displays consistently high variations. This is because of the poor control mechanism and high load variation [1] in the National Grid of Bangladesh.

The number of available examples for Power recordings of each grid, nominal frequencies corresponding to each grid and their range of variation are provided in Table-1. The Database contained variable numbers of power recordings across all grids.

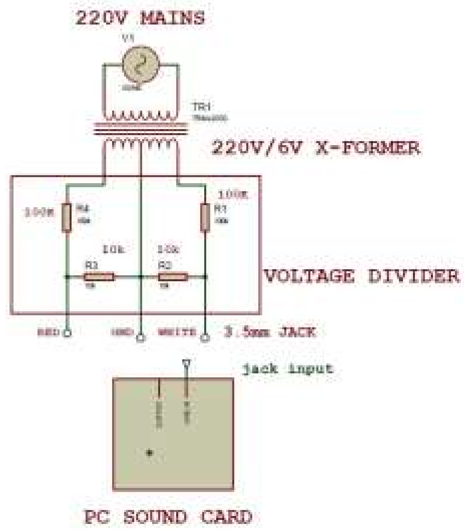

Fig. 2. Schematic Diagram of Circuit

III. Trained Classification System

While training the Classification System the normalization parameters were stored, and later used to normalize the Testing Data. We resorted to a similar technique as the one reported in [8]. In this work, a multiclass classifier based on the Error-correcting output codes (ECOC) multiclass model using support vector machine (SVM) binary learners on the MATLAB platform [19-22]. This employs a one vs one model. A Gaussian Radial Basis kernel function with automatic kernel scaling was used. When the Kernel scale mode is set to 'Auto', MATLAB uses a heuristic procedure to select the scale value; the heuristic procedure uses subsampling [23]. Both the LIBSVM, used in the previous work and MATLAB's ECOC multiclass system uses a one vs one multiclass model, weighted SVM, and works in a similar way [20, 21, 25].

IV. Result

For the classification system trained only on power recording we achieved an overall accuracy of 77.7% with a 20-fold Cross-Validation scheme selecting all the features. To closely understand the results of our approach the confusion matrix is presented in Table 4. Here, the efficiency of our system in predicting various grids are presented.

For system trained on Power files, the (M) Bangladesh has efficiency of identification of 100% even though we have relatively very bad identification accuracy for other grids. This is for two reason:

TABLE I. **Description of the database**

Grid	No. of Power Examples	Nominal Frequency (Hz)	Maximum Frequency (Hz)	Minimum Frequency (Hz)	Frequency Range (Hz)
(A) Texas	33	60	60.032	59.651	0.067
(B) Lebanon	38	50	50.580	49.142	1.707
(C) Eastern U.S	34	60	60.044	59.962	0.082
(D) Turkey	37	50	50.057	49.942	0.115
(E) Ireland	38	50	50.077	49.925	0.152
(F) France	26	50	50.778	49.942	0.134
(G) Tenerife	36	50	50.142	49.801	0.341
(H) India (Agra)	41	50	50.316	49.716	0.600
(I) Western U.S	37	60	60.047	59.951	0.095
(J) Brazil	35	60	60.0655	59.8662	0.1993
(K) Norway	35	50	50.1069	49.8734	0.2335
(L) Australia	36	50	50.0963	49.8701	0.2262
(M) Bangladesh	58	50	51.2637	49.9932	2.2705

TABLE II .**Confusion Matrix**

Testing Classes	No. of Examples	Grid A	Grid B	Grid C	Grid D	Grid E	Grid F	Grid G	Grid H	Grid I	Grid J	Grid K	Grid L	Grid M
(A) Texas	33	56%	-	-	-	-	-	-	-	-	-	-	-	-
(B) Lebanon	38	-	74%	-	-		-	-	-	-	-	-	-	26%
(C) Eastern U.S	34		-	73%	-	-	-	-	-		-	-	-	-
(D) Turkey	37	-	-	-	53%				-	-	-	-	-	-
(E) Ireland	38	-	-	-		68%		-	-	-	-	-	-	-
(F) France	26	-	-	-	-		30%	-	-	-	-	-	-	-
(G) Tenerife	36							62%	-	-	-	-	-	-
(H) India (Agra)	41	-	-	-	-	-	-	-	85%	-	-	-	-	2%
(I) Western U.S	37	-	-	-	-	-	-	-	-	88%	-	-	-	2%
(J) Brazil	35	-	-	-	-	-	-	-	-	-	97%	-	-	-
(K) Norway	35	-	-	-	-	-	-	-	-	-		92%		-
(L) Australia	36	-	-	-	-	-	-	-	-	-	-	-	94%	3%
(M) Bangladesh	58	-	-	-	-	-	-	-	-	-	-	-	-	100%

Firstly, other grids have less amount of data to train. The lack of data in comparison to the work in [8] explains the somewhat reduced accuracies of this program while using the same features, as more training data generally leads to better approximations and accuracy of the machine learning system [26]. Apparently Grid – B has high false positive result due to the presence of Bangladesh grid and it happened to have a wide range of ENF as well thus at times it falsely identifies as Bangladesh grid

Secondly, from ENF data it's evident that ENF collected from Bangladesh grid (M) fluctuates in wider range than any other grids mentioned. We know this is due to the instantaneous frequency fluctuates about the nominal value due to the load control mechanisms and the changes in the load demands within the power grid. [1] Due to poor load control produced ENF fluctuates a lot in Bangladesh Grid (M).

Apparently we have found few grids with nominal 50 Hz frequency to be classified as False Positive for Bangladesh Grid (M). If we had more ENF for other grids, we believe percentage of false positive data would decrease.

For the classification system trained only on power recording we achieved an overall accuracy of 96.03% with a 20-fold Cross-Validation scheme selecting all the features. To closely understand the results of our approach the confusion matrix is presented in Table 4. Here, the efficiency of our system in predicting various grids are presented.

V. Conclusion

In this work, we have described our method to extract ENF using Power Recording from Bangladesh National Grid. We analyzed our ENF comparing with 12 other grids around the world using machine learning. From confusion matrix we found Bangladesh Grid to be classified with 100% efficiency. Which is a great news as ENF data can be used for multiple forensic application like estimating the recording location. As we mostly compared with data from grids with advanced load control mechanism, wide fluctuation range of our grid was easily identified. This work also demonstrates the applicability of MATLABs ECOC Multiclass Classification System in classifying the region of origin of multimedia recordings through the analysis of ENF signals embedded in them.

REFERENCES

[1] Bollen, M.H. and I. Gu, Signal processing of power quality disturbances. Vol. 30. 2006: John Wiley & Sons.

[2] Grigoras, C., Digital audio recording analysis–the electric network frequency criterion. International Journal of Speech Language and the Law, 2005. 12(1): p. 63-76.

[3] Grigoras, C., Applications of ENF criterion in forensic audio, video, computer and telecommunication analysis. Forensic Science International, 2007. 167(2): p. 136-145.

[4] Grigoras, C. Applications of ENF Analysis Method in Forensic Authentication of Digital Audio and Video Recordings. in Audio Engineering Society Convention 123. 2007: Audio Engineering Society.

[5] Grigoras, C., Applications of ENF analysis in forensic authentication of digital audio and video recordings. Journal of the Audio Engineering Society, 2009. 57(9): p. 643-661.

[6] Garg, R., A.L. Varna, and M. Wu. Seeing ENF: natural time stamp for digital video via optical sensing and signal processing. in Proceedings of the 19th ACM international conference on Multimedia. 2011: ACM.

[7] Maher, R.C., Audio forensic examination. Signal Processing Magazine, IEEE, 2009. 26(2): p. 84-94.

[8] Hajj-Ahmad, A., R. Garg, and M. Wu, ENF-Based Region-of-Recording Identification for Media Signals. Information Forensics and Security, IEEE Transactions on, 2015. 10(6): p. 1125-1136.

[9] Cooper, A.J. The electric network frequency (ENF) as an aid to authenticating forensic digital audio recordings–an automated approach. in Audio Engineering Society Conference: 33rd International Conference: Audio Forensics-Theory and Practice. 2008: Audio Engineering Society.

[10] Adi Hajj-Ahmad, M.W. IEEE Signal Processing Cup 2016, DATASETS 2016; Available from: http://spcup16.ece.umd.edu/datasets.php.

[11] Hajj-Ahmad, A., R. Garg, and M. Wu, Spectrum combining for ENF signal estimation. Signal Processing Letters, IEEE, 2013. 20(9): p. 885-888.

[12] Neumaier, A. and T. Schneider, Estimation of parameters and eigenmodes of multivariate autoregressive models. ACM Transactions on Mathematical Software (TOMS), 2001. 27(1): p. 27-57.

[13] Schneider, T. and A. Neumaier, Algorithm 808: ARfit—A Matlab package for the estimation of parameters and eigenmodes of multivariate autoregressive models. ACM Transactions on Mathematical Software (TOMS), 2001. 27(1): p. 58-65.

[14] Arrillaga, J. and N.R. Watson, Power system harmonics. 2004: John Wiley & Sons.

[15] 61000-2-1, I., Electromagnetic Compatibility (EMC)—Part 2: Environment—Section 1: Description of the Environment–Electromagnetic Environment for Low-Frequency Conducted Disturbances and Signalling in Public Power Supply Systems. IEC, Geneva., 1990.

[16] Telegraph, I.T.U.I. and T.C. Committee, Directives Concerning the Protection of Telecommunication Lines Against Harmful Effects from Electricity Lines: Text Approved by the 2nd Plenary Assembly of the International Telegraph and Telephone Consultative Committee CCITT, New Dehli 1960. 1963: International Telecommunication Union.

[17] C57.110-1998, I.A., IEEE recommendation practice for establishing transformer capability when supplying nonsinusoidal load currents. IEEE, New York.

[18] IEEE/ANSI Std 18–2002 (Reaff 1991), IEEE Standard for Shunt Power Capacitors. IEEE, New York.

[19] Allwein, E.L., R.E. Schapire, and Y. Singer, Reducing multiclass to binary: A unifying approach for margin classifiers. Journal of machine learning research, 2000. 1(Dec): p. 113-141.

[20] Escalera, S., O. Pujol, and P. Radeva, Separability of ternary codes for sparse designs of error-correcting output codes. Pattern Recognition Letters, 2009. 30(3): p. 285-297.

[21] Escalera, S., O. Pujol, and P. Radeva, On the decoding process in ternary error-correcting output codes. IEEE transactions on pattern analysis and machine intelligence, 2010. 32(1): p. 120-134.

[22] Fürnkranz, J., Round robin classification. The Journal of Machine Learning Research, 2002. 2: p. 721-747.

[23] MathWorks, M. Choose Classifier Options. 2016 [cited 2017 31 March]; Available from: https://www.mathworks.com/help/stats/choose-a-classifier.html#bunt0n0-1.

[24] Adi Hajj-Ahmad, R., MinWu,. Spectrum Combining for ENF Signal Estimation - Dataset & Codes,. 2013 [cited 2017 March 03]; Available from: http://cwwongpatch2.no-ip.org/mast_new/research.php?t=enf.

[25] Chang, C.-C. and C.-J. Lin, LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology (TIST), 2011. 2(3): p. 27.

[26] Banko, M. and E. Brill. Scaling to very very large corpora for natural language disambiguation. in Proceedings of the 39th annual meeting on association for computational linguistics. 2001: Association for Computational Linguistics.