Location Forensics of Media Recordings Utilizing Power Signatures and Pole-Matching

Mohammad Ariful Haque, Sayeed Shafayet Chowdhury, Md Billal Hossain, Jayanta Dey, Ashraful Islam, Rakib Hyder, Samzid bin Hafiz, Ratul Khan, Munif Ishad Mujib, Tauhiduzzaman Khan Himel, Uday Saha

Executive Summary

Electrical network frequency (ENF) analysis has become a very promising tool for multimedia forensics and security in recent years. ENF is supply frequency of power distribution grids that has a nominal value of 50/60 Hz. Due to load variations and control mechanisms in a grid, ENF normally fluctuates around its nominal value. This fluctuation trend is almost identical in different locations covered by the same grid but significantly different from grids. This ENF can be picked up by digital audio recording due to electromagnetic interference and acoustic vibration while conducting recording near any power source. Therefore, ENF analysis can be used for the identification of location of recording of digital multimedia recording.

In this work, we have developed a novel grid-of-origin identification system from media recording. The system consists of a number of support vector machine (SVM) and pole-matching (PM) classifiers. First, we classify the signal on the basis of nominal frequency and audio or power type. Then an SVM classifier, trained for a particular data type, narrow down the list of possible grids on the basis of different discriminating features extracted from the ENF signal. The result is then passed to the pole matching classifier and the final grid is detected based on the minimum distance between the estimated poles of test and training grids. For "no grid" detection, we have used four single-class-SVM classifiers that determine, using outlier detection method, whether a given test sample is "no grid" or not.

The developed technique gives very good accuracy on the practice data with high confidence margin. Our classification system can correctly classify 48 practice samples and we have got 96% accuracy by the feedback system. We expect that the proposed technique will give good accuracy on the test data set too.

- *Report on practice data set*

AHCFF,BGIND,AFNDC,INNAE,HBBAD,CGHGB,DDCHG,EAIHI,EHECF,FNGEI

Accuracy (obtained from the feedback system): 96%

- *Report on test data set*

NDDND,HHDAF,ANGBG,BFCEH,GHHHG,NFDAI,DNFHI,IECBD,EIIBE,FGNAG,IIIIG,HAEFC,
CCFDG,CECGI,EICDN,BEBHA,DIHCG,AABIH,CNDBA,GBFBB

Affiliation: Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, Dhaka-1205, Bangladesh.
Team ID: 24546, Team Name: Resonance_1011

We have designed and implemented a circuit for capturing the power data. Using this circuit, we have recorded the power data of Bangladesh grid for a total duration of 9-10 hours. Audio data is captured using a battery powered mobile device disconnected from the power lines. Audio is captured simultaneously with the power data, near electrical activity. The data were collected at different times: weekdays and weekend, days and night.

We have compared the ENF extracted from the data acquired of our native Bangladesh grid with the ENFs extracted from the training datasets. Since the nominal frequency of our grid is 50 Hz, our analysis focuses on comparison with the other 50 Hz given grids. In this regard, the analysis is concentrated from two points of views. Firstly, a visual analysis of the ENFs is provided. Next the features extracted from the ENFs are used to compare the similarity or distinction of our grid with those of the others. It is found that the characteristics of grids B and H has the highest match with those of our Bangladesh grid.

I. INTRODUCTION

LOCATION forensics of media recordings enables us to detect the location where a media recording was originally captured. It is important because it provides us strong technologies for the analysis of audiovisual data in order to combat kidnapping, human trafficking, child exploitation cases which have become a potential threat to the global community. Moreover, it finds its application in automatic geo-tagging of multimedia recordings uploaded on YouTube, Facebook and others social media.

We can detect the location of a media recording by extracting the power signature which is also known as the Electric Network Frequency (ENF) signal. The ENF is the nominal frequency of local power grid which varies due to change of load on power grid and its internal mechanism. It is a unique and intrinsic characteristic of a power grid. This ENF can be picked up by digital audio recording due to electromagnetic interference and acoustic vibration while conducting recording near any power source [1]. After extracting the ENF signal from a media recording, we can detect the location where the audio was captured by comparing the extracted ENF signal with the reference ENF database. For this purpose, advanced machine learning technique has already been proposed in the literature [2]. Statistical and wavelet features along with AR coefficients of ENF signal is used to train the machine learning system. Thus, such a system identifies the grid (region) in which a multimedia signal was recorded.

The identification of location of media recording is not a trivial task. The recorded signal may contain human speech, acoustic noise, and other interfering signals. The amounts of noise and distortions can be different even within the same signal due to change of recording conditions. This can distort the original ENF signal and create confusions for the machine learning system when defining the class boundaries and could lead to errors in identification. To overcome the problem, features of ENF signal extracted from clean power recordings and audio recordings are incorporated to train the machine learning system that achieves a relatively better accuracy on identifying the region/grid-of-origin of audio recordings [2].

In this work, we have developed a novel grid-of-origin identification system consisting of support vector machine (SVM) and pole-matching (PM) classifiers. The SVM classifiers depend on the extracted features from the ENF signal, whereas the proposed PM classifier takes the decision by calculating the mean distance between the estimated poles of test and training grids. The poles of a grid are calculated from the auto-regressive (AR) parameters of the given raw data. Now we briefly describe how our classification system works. First, we classify the given test sample by the basis of nominal frequency and audio or power type. Thus we get four different types of data: 50 Hz power, 60 Hz power, 50 Hz audio, and 60 Hz audio. Then a single class SVM classifier of the corresponding data-type determines, using outlier detection method, whether the given test sample is an “out of grid” data or not. If the test sample is not an “out of grid” data, an SVM multiclass classifier of the

corresponding data-type is applied, using radial bias function (RBF) kernel, to narrow down the list of probable grids. The result is then passed to the pole matching classifier and the final grid is detected based on the minimum distance between the estimated poles of test and training grids. The proposed technique gives very good accuracy on the practice data with high confidence margin. The number of correctly detected grids is 48 (out of 50 samples) and thus we get 96% accuracy on the practice data set. We expect that the proposed technique will give good accuracy on the test data set too.

II. Extraction of ENF signals

A variety of ENF extraction techniques have been reported in the literature. Most of the techniques are based on either time or frequency based methods or variations of these techniques. Both the approaches have practical applications. Time-based techniques such as zero-crossing method have proven to be very useful in order to record the ENF directly from the power line. Such methods are used by power suppliers, which are obliged to keep the ENF within a given tolerance and thus need to record the ENF time history to validate that. Zero crossing method is, however, not suitable to extract ENF from real-world speech or music audio content. Frequency or short-time-Fourier-transform (STFT) based methods, in contrast, are suitable for this, and are commonly applied for this purpose [3]. We used frequency based method in our work.

A. ENF extraction method

First of all we determine the nominal frequency of the given test sample. It is done simply by detecting the peak-frequency of the magnitude spectrum around 50 Hz, 60 Hz, and their harmonic bands. The frequency having the greatest peak indicates the nominal frequency of the grid-of-origin. Then we calculate whether the signal is audio or power. For this, we compare the "summation of spectral magnitude in the nominal frequency band and its harmonics" (S_p) to the "summation of spectral magnitude in rest of the spectrum" (S_r). If S_r is much larger compare to S_p, it is treated as an audio signal.

After deciding whether the signal is audio or power, we use different versions of ENF extraction method for each type. For audio signal, we use spectrum combining technique (similar to [3] with some modifications). The method in [3] includes harmonic analysis which is particularly useful for extracting weak ENF signal from audio data. However, for power data we extract the ENF signal directly from the fundamental nominal frequency band as in [4]. Since the power signal is clean and the nominal frequency signal is always present in the data, we do not need to consider the harmonic bands in this case.

1) ENF extraction from audio recording

For ENF extraction, the total data points (of each recording) are divided into a number of overlapping frames. In our work, the time-duration of each frame is 5s and the overlap between successive frames is 3s. Thus we get estimated ENF values at 2s intervals for the audio signal. In order to obtain the combined spectrum for each frame, we compress and shift the spectral components at different harmonic bands to the nominal base range, $[f_0- f_B, f_0+ f_B]$, and then combine them together as follows (as in [3]):

$$S(f) = \sum_{k=1}^{L} w_k P_{B,k}(kf).$$

where, w_k is the combining weight of the kth harmonic band and $P_{B,k}(f)$ is the spectral band around the kth harmonic frequency. As in [3], w_k is obtained by estimating the SNR in the kth harmonic band. The ENF can vary over a wide range across the globe (for example, Laos has $f_0 \pm f_B = 50 \pm 8$ Hz [5]). Therefore, we have chosen three different bandwidths, $f_B = 1, 3, 8$ for SNR calculation. For each value of f_B, we get a different w_k and a different combined spectrum S(f). Thus we get three different probable estimates of ENF by searching the maximum in each S(f) for a given time-frame. Finally, we

choose the ENF signal that is the least varying among the successive frames. After estimating instantaneous ENF, we perform IIR Hampel filtering followed by smoothing to remove outliers in ENF signals.

2) ENF extraction from power recording

For each recording, total data points are divided into a number of non-overlapping frames of duration 2s. Then we use an STFT based quadratic interpolation technique to determine the ENF. Choosing the frequency directly from the STFT may give wrong estimation of ENF because Fast Fourier Transform (FFT) is computed for a finite number of discrete frequencies. So, actual frequency with maximum Energy may not be among them. We can increase number of FFT points for better frequency resolution but higher FFT points will be computationally complex. To overcome this limitation, a quadratic interpolation scheme escribed in [4] is used in our work.

For each STFT frame, log power spectrum is calculated at discrete frequencies. Then the peak of the spectrum is identified in the frequency range between 46 Hz to 64 Hz. Let, k represents the sample index of the peak-spectrum and α, β, γ denote the spectrum values at (k-1), k and (k+1). Now according to quadratic interpolation technique, the location of true peak can be estimated as

$$k_{true} = k + p$$

where,

$$p = \frac{1}{2}\frac{\alpha-\gamma}{\alpha-2\beta+\gamma}.$$

The ENF in Hz is obtained by $(F_s * k_{true})/N$, where F_s is sampling frequency and N is the number of FFT point used.

B. Examples of extracted ENF signals

Fig. 2.1 shows some sample ENFs extracted from power data using our adopted method for different 60 Hz and 50 Hz grids. It is observed that the ENF of a particular grid is significantly different from the other. Thus the ENF signal qualifies to be the power signature of a particular grid. Though the nominal frequency of A, C, and I grids are same, the mean and variance of the ENF will be different. The same observation is valid for B, D, and H grids. The minimum-to-maximum range of ENF signal is significantly different for the 50 Hz grids. The maximum rate of change of frequency is also different among all the grids. Fig. 2.2 shows some sample ENFs extracted from audio data using our adopted method for different 60 Hz and 50 Hz grids. Comparing Figs. 2.1 and 2.2, it is observed that there are similarities between the power and audio ENF signals of the same grid. Therefore, it seems that grid identification results from audio data can be improved by training the classifiers using both power and audio data.

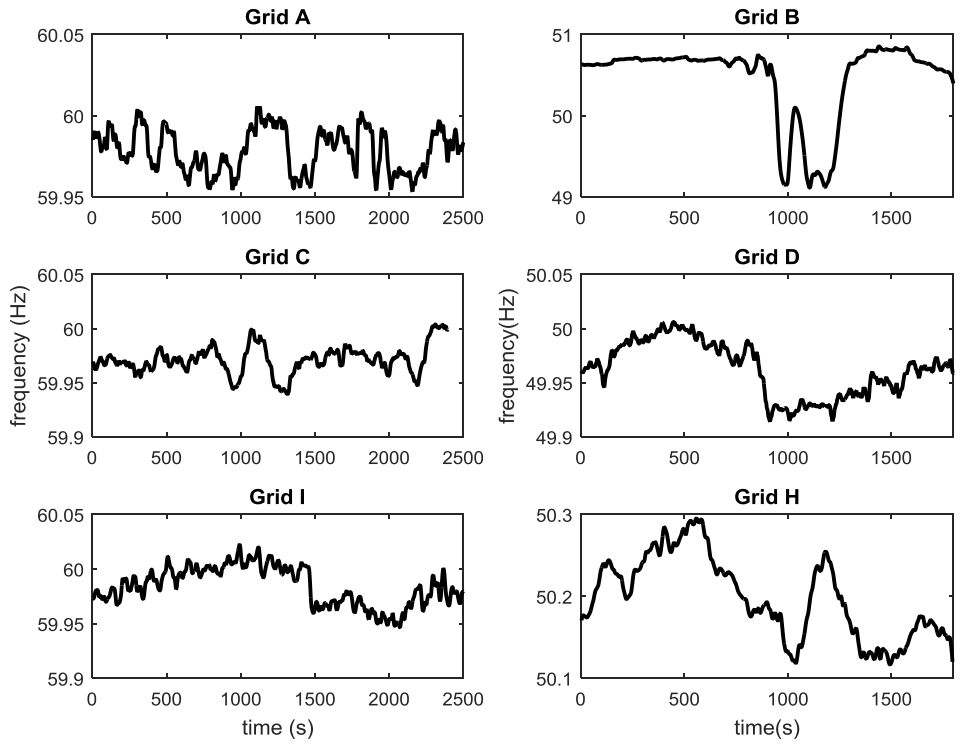

Fig 2.1: ENF signals of different grids extracted from the power data (Train_Grid_AP1, Train_Grid_CP1, Train_Grid_IP1, Train_Grid_BP1, Train_Grid_DP1, Train_Grid_HP1).

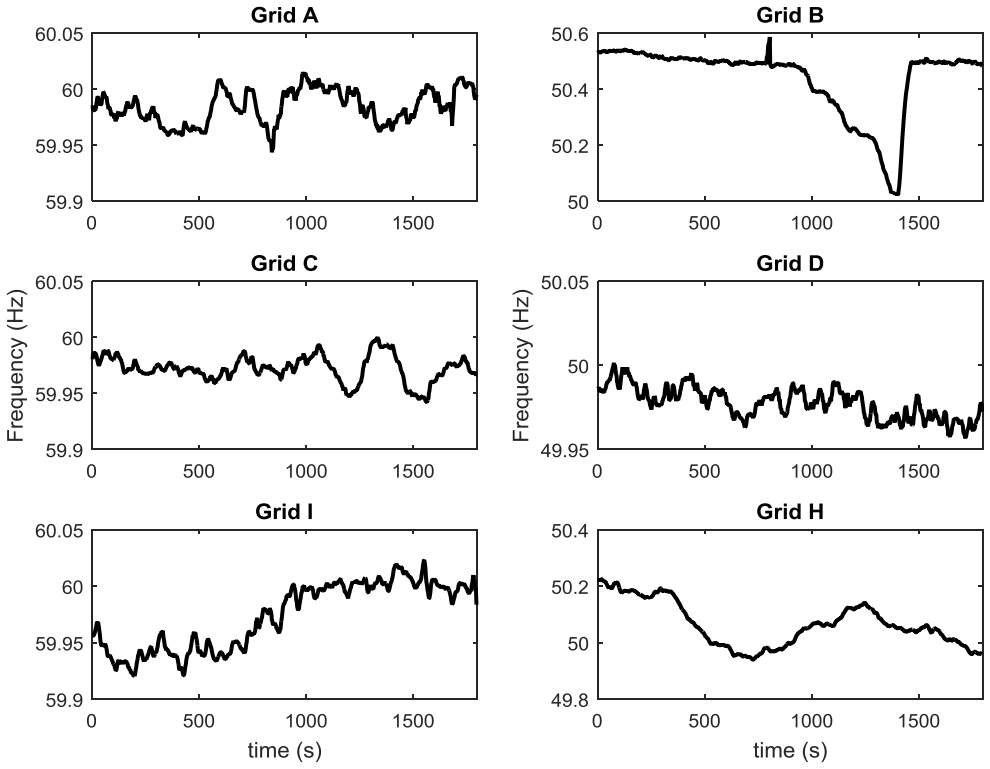

Fig 2.2: ENF signals of different grids extracted from the audio data (Train_Grid_AA2, Train_Grid_CA2, Train_Grid_I A2, Train_Grid_BA2, Train_Grid_DA2, Train_Grid_HA2).

III. Location Classification/Identification System

A. *Description of the classification system*

A simplified block diagram of the proposed classifier is presented in Fig. 3.1.

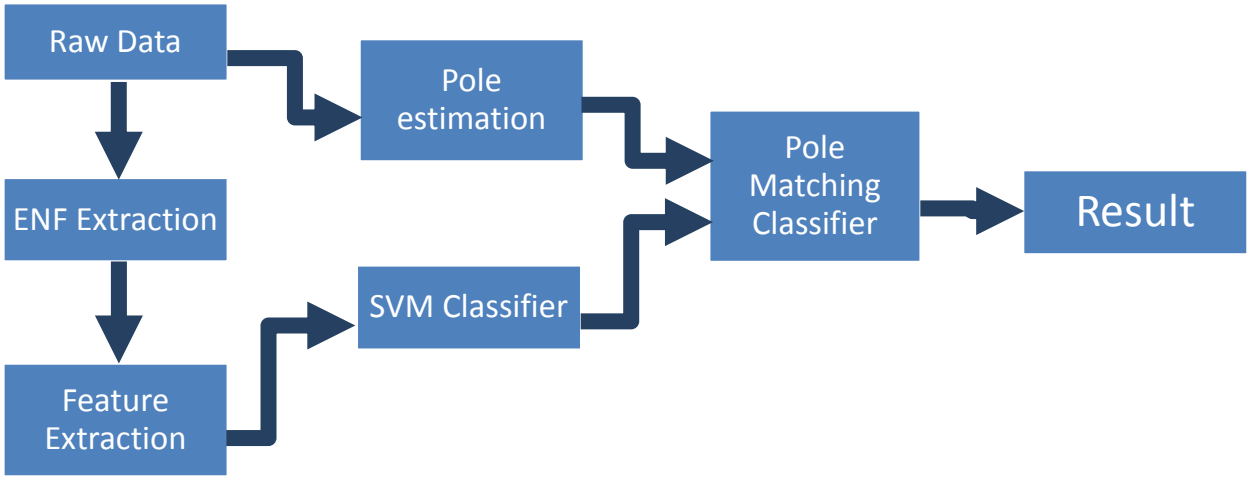

Fig. 3.1: Block diagram of the proposed classification system

We now describe the functions of different blocks in some details.

1) ENF extraction

The ENF signal is extracted from the raw data using the methods described in Section II. As stated earlier, we have two different techniques of ENF extraction for power and audio data. For power data, we estimate the ENF directly from the spectral band of the nominal frequency. For audio data, we consider all the different harmonic bands of the nominal frequency in order to extract the ENF signal.

2) Feature extraction

The extracted ENF signal is divided into non-overlapping segments having 32 samples in each block. Then, the following features were extracted from each segment:

Table 3.1: List of extracted features from the ENF signals

Name of the Feature	Feature No.
Variance	1
Mean	2
Average of max. three frequency fluctuation	3
AR Coefficient	4-5
Wavelet Coefficients (5 level)	6-37
Range of each segment	38

3) SVM classifier

We have used four different SVM classifiers for each of the different data-types (50Hz-power/60Hz-power/50Hz-audio/60Hz-audio). Different classifiers for different data-types help us to reduce the overlapping between different features. Moreover, to make the classifier free of the "curse of dimensionality", the features are analyzed using the "sparse logistic regression with Bayesian

regularization" [6] and only the significant features are taken. The tolerance level of feature selection was adjusted to 0.6 by applying it on the practice dataset and observing the highest cross validation and the result accuracy. The list of significant features for different data-type is presented in Table 3.2. The order of the features implies the level of significance. The most significant feature is placed in the first position.

Table 3.1: List of selected features for different data types

Data type	Selected feature nos.
60 Hz power	1, 38, 5, 3, 13, 4, 20, 26, 22, 34, 30, 6, 37, 27, 19, 31, 36, 32, 33, 18, 14
50 Hz power	6, 3, 1, 2, 5, 26, 13, 33, 18, 4, 12, 11, 38, 24, 31, 32, 23, 34, 22, 36, 21, 30, 25, 28, 20, 29
60 Hz audio	3, 1, 4, 25, 12, 33, 30, 28, 37, 21
50 Hz audio	2, 1, 3, 26, 37, 4, 6, 11, 22, 13, 28, 29, 5, 35, 10, 31

We now describe how our SVM classifier works. First, a single class SVM classifier trained for a particular data-type is used to identify whether the given sample is a "no grid" data or not. If it is not a "no grid" data, an SVM multiclass classifier trained for the corresponding data type is applied using RBF kernel in order to narrow down the list of probable grids. The decision is then passed to the pole matching classifier to take final decision.

4) Pole estimation

The raw signal is divided into non-overlapping segments of duration 10s. We estimate the AR parameters from each of the data segments. The roots of the AR-parameter polynomial give us the poles of the system. Fig. 3.2 shows the distribution of poles (in the z-plane) for different 60 Hz grids obtained from the training power data. It is observed that poles are significantly isolated for different grids.

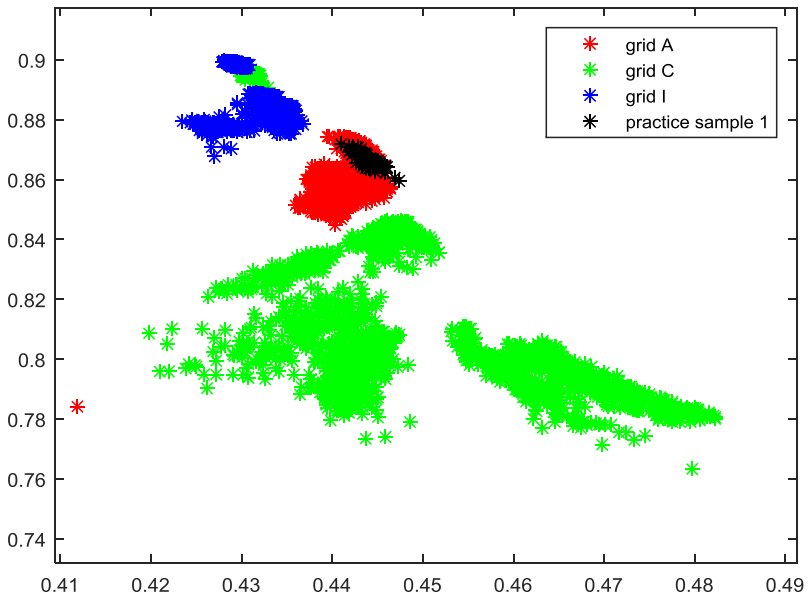

Fig. 3.2: Distribution of poles for different 60 Hz grids obtained from the training power data

5) Pole matching classifier

The SVM classifier often gives confused decisions on grid-of-origin detection. Therefore we have used the SVM classifier only to narrow down the list of probable classifiers. The result of the SVM classifier is passed to the pole matching classifier that determines the mean distance between the test and training poles. The (short-listed) training grid for which we get the minimum mean distance is identified to be the grid-of-origin. Fig. 3.2 superimposes the poles obtained from practice data 1 on the poles of the training dataset. It is clearly understood that the test poles have the minimum distance with those of grid A and hence it is identified as grid A.

B. Classification result on practice and test data

1) Report on practice data set

AHCFF,BGIND,AFNDC,INNAE,HBBAD,CGHGB,DDCHG,EAIHI,EHECF,FNGEI

Accuracy (obtained from the feedback system): 96%

2) Report on test data set

NDDND,HHDAF,ANGBG,BFCEH,GHHHG,NFDAI,DNFHI,IECBD,EIIBE,FGNAG,IIIIG,HAEFC,
CCFDG,CECGI,EICDN,BEBHA,DIHCG,AABIH,CNDBA,GBFBB

C. Explanation of the submitted code

We have used "LIBSVM" for developing and testing our classification system [7].

We have developed a Matlab GUI-based Application to facilitate the verification of our classification results. A separate README file is attached with that application folder. Please see the README file to use the developed Matlab application.

IV. Circuit Design and Data Analysis for ENF Acquisition

A. Description of the Hardware and data acquisition technique

1) Circuit description

The diagram of the circuit that we built for the acquisition of power data is given in Fig. 4.1.

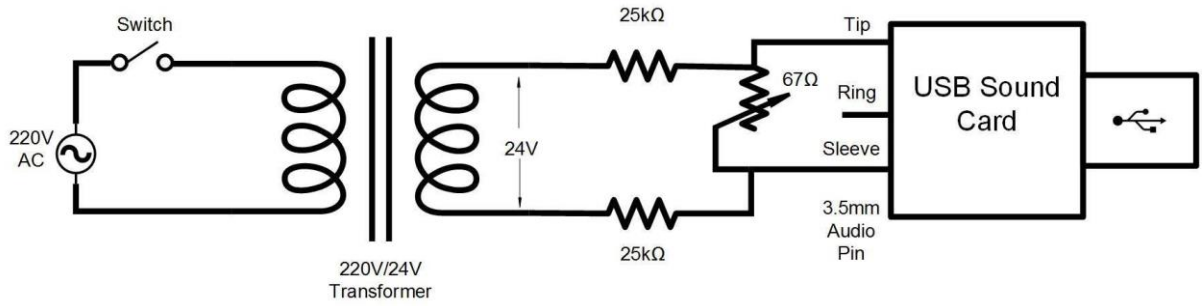

Fig. 4.1: Circuit diagram of the data acquisition hardware

The circuit consists of a step-down transformer, fixed and variable resistors, switch and USB sound card. The primary terminal of a 220V/24V, 600mA transformer is connected to the grid power supply

through an SPDT switch. The secondary of the transformer is connected to a resistive voltage divider. The divider comprises of two 25k resistors in series with a 10k variable resistor. The two-terminals of the variable resistor go to the tip and sleeve of a standard 3.5mm audio jack. The other end of the audio connector is connected to a USB sound card.

2) Capturing of power data

The functional block diagram of the hardware is shown in Fig. 4.2.

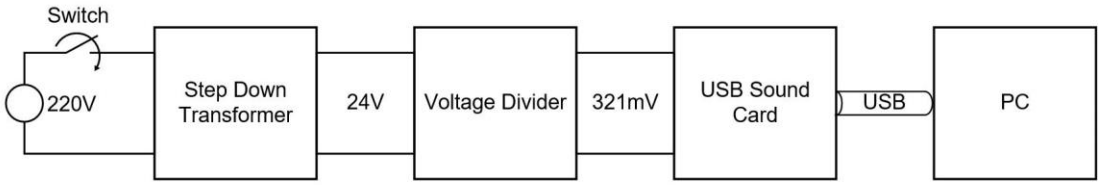

Fig. 4.2: Functional block diagram of the hardware setup.

220V AC input is fed in from the mains through an SPDT switch to a step-down transformer. The transformer steps the voltage signal down from 220V to 24V. This voltage is then applied to a resistive voltage divider. The output signal voltage of the divider is 32.1mV. This signal is fed into a sound card using a standard 3-pin 3.5mm audio connector. The sound card is connected to a PC using a USB connection. The sound card converts the analog voltage input into a digital data signal which is then captured through software on the PC. The sampling rate is 8KHz.

3) Capturing of audio data

Audio data is captured using a battery powered mobile device disconnected from the power lines. Audio is captured simultaneously with the power data, near electrical activity. The sampling rate is 8KHz.

4) Data description

We have captured both audio and power data of duration more than 9 hours. The data were collected at different times encompassing weekdays and weekend, day and night. Table 4.1 shows the list of collected data along with the corresponding time-stamp.

Table 4.1: List of collected Bangladesh grid data

Name of the audio data	**Name of the power data**	**Time stamp**
abd1	pbd1	Dec1, 2:20am
abd2	pbd2	Jan3, 11am
abd3	pbd3	Jan7, 1pm
abd4	pbd4	Nov7,1am
abd5	pbd5	Dec8,1:10am
abd6	pbd6	Jan8,11am
abd7	pbd7	Jan8,12am
abd8	pbd8	Jan9,1am
abd9	pbd9	Dec21,1am
abd10	pbd10	Dec21,9pm
abd11	pbd11	Jan10,12:20am
abd12	pbd12	Dec23,9am

B. Comparison of our own grid ENF with the given ENF sets on training

Here we would like to compare the ENF extracted from the data acquired of our native Bangladesh grid with the ENFs extracted from the training datasets. Since the nominal frequency of our grid is 50 Hz, our analysis focuses on comparison with the other 50 Hz given grids. In this regard, the analysis is concentrated from two points of views. Firstly, a visual analysis of the ENFs is provided. Next the features extracted from the ENFs are used to compare the similarity or distinction of our grid with those of the others.

1) Visual analysis of the ENFs

Fig. 4.3 presents sample ENF signals extracted from the training power data along with a sample ENF signal of Bangladesh grid extracted from the acquired data using our hardware setup.

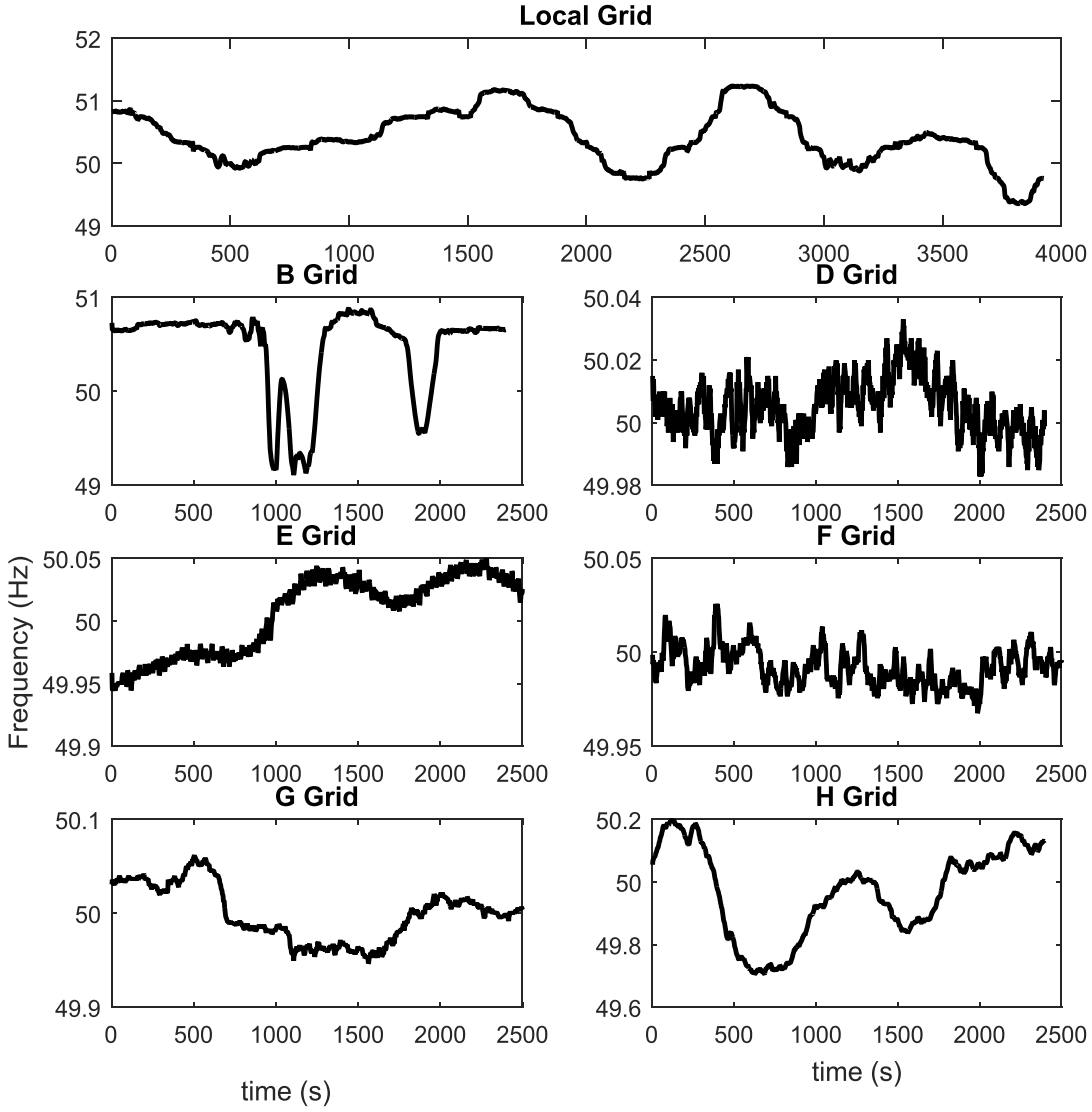

Fig. 4.3: ENF signals of different grids extracted from the power data (Bangladesh grid_pbd11, Train_Grid_BP1, Train_Grid_DP3, Train_Grid_EP1, Train_Grid_FP4, Train_Grid_GP10,Train_Grid_HP2).

From the ENFs, we can see that Gird B has frequent outliers in the ENF where frequency falls by almost 1Hz or so and then returns to a more stable value, but Bangladesh grid doesn't show such characteristics. A prominent distinction between our grid's ENF and most others is the range. Bangladesh grid's ENF varies within a large range from around 49.5 to 51.5 Hz whereas grids D, E, F and G have very small range. Although grid D has a more or less consistent ENF, its value doesn't show much drift, rather returns to the nominal quite frequently. On the other hand, grid E shows higher drift although having similar small range. Bangladesh grid too shows drift in the sense that it doesn't rapidly repeat the same pattern; rather it takes quite a long time to return to the nominal. Grid F's ENF tends to have peaks once in a while but our grid doesn't show any feature like that. Grid G also has very small variance in the ENF. The ENF of grid H has a close match with that of our Bangladesh grid. Both of them vary following a smooth pattern rather than sudden jumps. The main difference here too is the range. While our grid varies between two larger ENF values, grid H has a smaller domain. But other features tend to be quite similar.

2) Analysis of ENF features

In this section we illustrate the distribution of different features obtained from the ENF signals of different grids in each time-frame. Fig. 4.4 shows the Box-Whisker plots of mean and variance of ENF signals in different grids. It is seen that Bangladesh grid has the highest variation of mean ENF among all the grids. In terms of mean and variance features, Grid B has the closest match with Bangladesh grid.

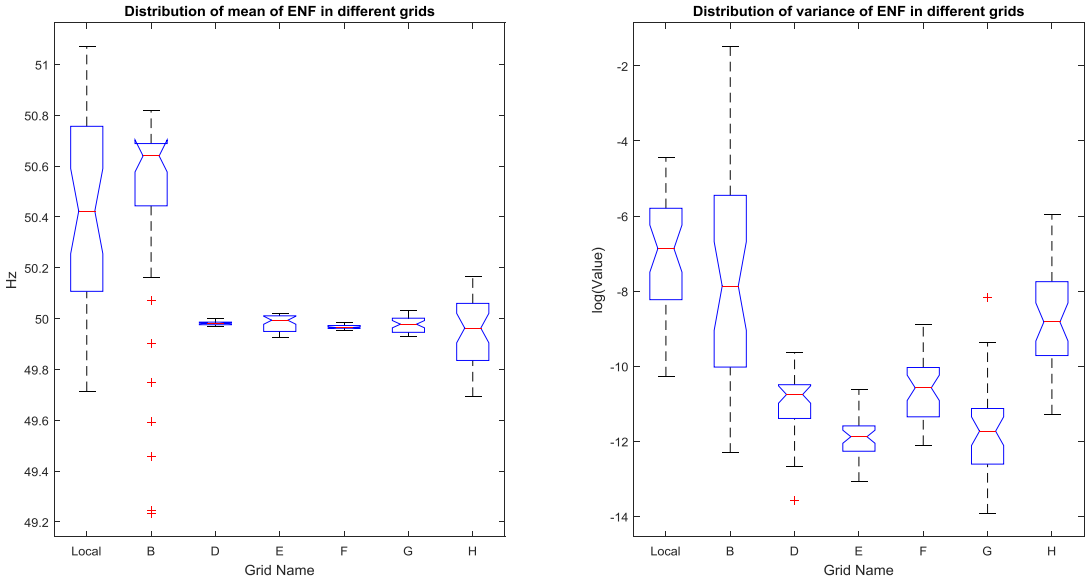

Fig. 4.4: Distributions of mean and variance of ENF signals (features no. 1 and 2 of Table 3.1) in different grids.

Fig. 4.5 shows the Box-Whisker plots ENF-fluctuations (rate of change of frequency) and 2nd AR coefficient of ENF signals in different grids. Similar to Grid B, Bangladesh grid has high ENF fluctuation. However, this fluctuation is not as abrupt as grid B. The ENF of Bangladesh grid has similar AR coefficient as B, F, and H.

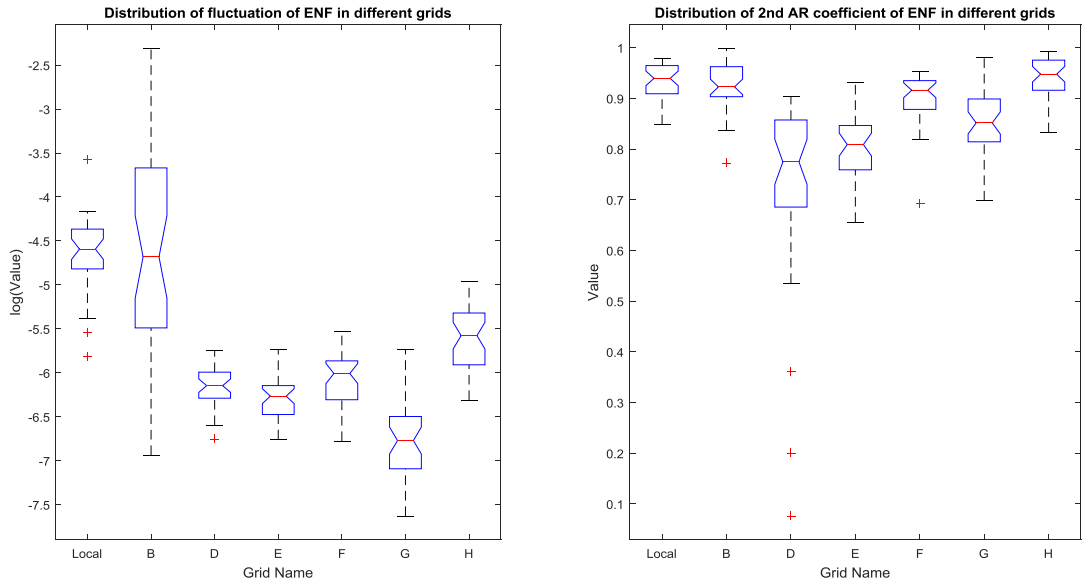

Fig. 4.5: Distributions of ENF-fluctuations and 2nd AR coefficient of ENF signals (features no. 3 and 5 of Table 3.1) in different grids.

Fig. 4.6 shows the Box-Whisker plots of 1st wavelet coefficient and range of ENF signals in different grids. Both of these features exhibit large variation in Bangladesh grid. After observing all the different Box-Whisker plots, it can be said that the grids that matches most with our Bangladesh grid are grid B and H.

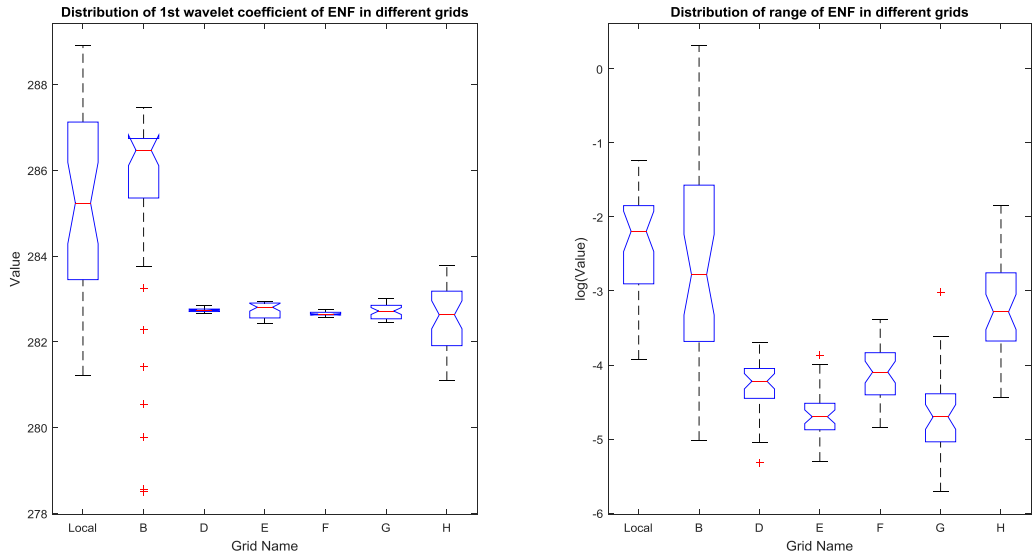

Fig. 4.6: Distributions of 1st wavelet coefficient and range of ENF signals (features no. 6 and 38 of Table 3.1) in different grids.

V. CONCLUSION

In this report, we have described our developed grid-of-origin identification system from media recording. The system consists of a number of support vector machine (SVM) and pole-matching (PM) classifiers. First, we classify the signal on the basis of nominal frequency and audio or power type. Then an SVM classifier, trained for a particular data type, narrow down the list of possible grids on the basis of different discriminating features extracted from the ENF signal. The result is then passed to the pole matching classifier and the final grid is detected based on the minimum distance between the estimated poles of test and training grids. The proposed technique has shown very good accuracy on the practice data with high confidence margin. The feedback system shows 96% accuracy on the practice data set. We expect that the proposed technique will give good accuracy on the test data set too.

REFERENCES

[1] C. Grigoras, "Applications of ENF criterion in forensics: Audio, video, computer and telecommunication analysis," Forensic Science International, vol. 167, no. 2-3, pp. 136 – 145, April, 2007.

[2] A. Hajj-Ahmad, R. Garg and M. Wu, "ENF-Based Region-of-Recording Identification for Media Signals," IEEE Trans. on Information Forensics and Security, vol. 10, no. 6, pp.1125-1136, June 2015.

[3] A. Hajj-Ahmad, R. Garg and M. Wu, "Spectrum Combining for ENF Signal Estimation," IEEE Signal Processing Letters (SPL), pp. 885-8, Vol. 20 (9), Sept. 2013.

[4] J. O. Smith and X. Serra, "PARSHL: An analysis/synthesis program for non-harmonic sounds based on a sinusoidal representation," International Computer Music Conference, 2004.

[5] http://www.hyvac.com/tech_support/electrical_sources_worldwide.pdf

[6] Gavin C. Cawley and Nicola L. C. Talbot, “Gene selection in cancer classification using sparse logistic regression with Bayesian regularization,” Bioinformatics: 22, 2348–2355

[7] Chang, Chih-Chung and Lin, Chih-Jen, “LIBSVM: A library for support vector machines,” ACM Transactions on Intelligent Systems and Technology, vol. 2, issue 3, pp. 27:1--27:27, 2011