Transactions on Information Forensics and Security

Location Forensics From Media Recordings Using Pole-matching Classifier

Journal:	*Transactions on Information Forensics & Security*
Manuscript ID	T-IFS-07089-2017
Manuscript Type:	Regular Paper
Date Submitted by the Author:	26-Jan-2017
Complete List of Authors:	Chowdhury, Sayeed Shafayet; Bangladesh University of Engineering and Technology, Electrical and Electronics Engineering Dey, Jayanta; Bangladesh University of Engineering and Technology, Electrical and Electronics Engineering Hyder, Rakib ; Bangladesh University of Engineering and Technology, Electrical and Electronics Engineering Islam, Ashraful ; Bangladesh University of Engineering and Technology, Electrical and Electronics Engineering Haque, Mohammad; Bangladesh University of Engineering and Technology, Electrical and Electronics Engineering
EDICS:	MMF-PROC-AU-Audio forensics < MMF-MULTIMEDIA FORENSICS

Location Forensics From Media Recordings Using Pole-matching Classifier

Sayeed Shafayet Chowdhury, *Student Member, IEEE,* Jayanta Dey, *Student Member, IEEE,* Rakib Hyder, *Student Member, IEEE,* Ashraful Islam, *Student Member, IEEE,* and Mohammad Ariful Haque, *Member, IEEE*

***Abstract*—Information regarding the location of power distribution grid can be extracted from multimedia signals (e.g., audio, video data) recorded near electrical activities. This implicit mechanism of identifying the origin-of-recording can be a very promising tool for multimedia forensics and security applications. Generally, the location of a media recording is estimated by exploiting the Electric Network Frequency (ENF) signal which is the instantaneous frequency of power distribution grids. However, there is no unique way to obtain ENF signal from the raw data and therefore the performance of the grid-location detection system becomes dependent on the accuracy of the ENF extraction method. In this paper, we propose a novel approach for location forensics that estimates pole features from the raw data without calculating ENF signal. We use these features to develop a distance-based multiclass classifier that is able to identify the region-of-recording with high accuracy. We demonstrate through experimental results and relevant analysis that the proposed method surpasses the state-of-the-art technique in terms of accuracy, robustness, and computational efficiency.**

***Index Terms*—Location forensics, pole estimation, pole-matching classifier, classification accuracy, computational time.**

I. Introduction

Location forensics has become an important area of research in the 21st century. With the proliferation of terrorism, child pornography [1] or abuse on women, there is an increasing demand for reliable identification of location of media recordings for ease of law enforcing agencies. Success in identifying such locations properly can ease the process of getting hold of the criminals involved. Furthermore, retrieval of location information embedded in the media recordings might pave the way for automatic tagging of geo-location of huge amount of digital data which are being uploaded every moment on social media platforms such as YouTube and Facebook [2]. Therefore, a consistent as well as fast and computationally light method of region-of-recording recognition is highly necessary.

A potential route to obtain the information of recording location from media recordings is to extract the fingerprint left on that recording by the power grid of that location. Due to electromagnetic radiation, every grid acts like an antenna that radiates an electromagnetic wave containing corresponding grid frequency. It has a nominal value of 60 Hz in North America and 50 Hz in most other places of the world. This radiation introduces a weak interference in audio or video clips recorded near places where there is electrical activity. The instantaneous frequency of a power grid fluctuates over time around its nominal value due to control mechanism and load variation. The variation pattern of the Electric Network Frequency (ENF) over time for any grid is defined as an ENF signal. ENF can vary with time over a certain range for a certain grid. For example, in Lebanon, it varies by about 1 Hz around 50 Hz nominal frequency, whereas in China, this variation is about 0.1 Hz. Though this variation of frequency can be considered a random process, it follows a particular statistical characteristics. The frequency variation is almost identical in all places of the same grid [3] and depends on load variation pattern of that grid. The load variation has different statistical distribution for different grids and follows a specific pattern for a particular grid. So, it can be inferred that the electromagnetic interference contains information about the grid. If such interference pattern can be reliably extracted from the recording, we can use this information to correctly identify the location of the recording.

The ENF signal extracted from the media recording is generally utilized for grid-location identification. However, accurate and reliable estimation of ENF signal from the media recording is not a trivial task. The frequency of the eletromagnetic inteference signal needs to be estimated from short segments of media and sensor recordings that 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 which can heavily distort the actual ENF signal pattern. A variety of ENF extraction techniques have been reported in the literature. Initial work in this area was proposed by Grigoras and Cooper [4]- [9]. Their study reveals that a reference ENF measurement obtained from anywhere in the grid can then be used to determine the time at which the recording was made and to detect tempering. Some of the recent works that have extended the initial research includes the extraction of ENF from video flicker in [10], refining the Fourier approaches in [11], using dynamic programming and a feed forward spectral estimator in [12] and incorporating power signal harmonics in [13]. 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

The authors are with the Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, Dhaka-1205, Bangladesh. E-mail: arifulhoque@eee.buet.ac.bd

[13]. Again, Garg *et al.* proposed a half-plane intersection method based on highly quantized information from the correlation coefficient between the processed ENF signals across different locations [2]. But the method provides satisfactory results only for power recordings since the audio data do not have high correlation. Additionally, order of estimation constraints increases quadratically with the number of anchor regions, making this method [2] quite complex for higher number of locations.

There is no unique way to extract the ENF signal and different extraction methods may give different ENF estimations for the same media recording. In this paper, we propose a novel grid-location identification method that does not require extraction of ENF signal, rather we only use the features obtained directly from the raw data. In our approach, we first estimate the AR parameters of short data segments of the training data set. The roots of the AR-parameter polynomial provide the poles of the AR model. In testing phase, poles are extracted from the testing data. Now a distance-based classifier is used to match the poles of the testing data with those of different grids of training data set. The grid with minimum average distance between the training and testing poles is estimated as the identified grid. We have tested our system using the data set provided for the IEEE Signal Processing Cup, 2016 [14] on “Location Forensic of Media Recording”. The results show that the proposed method can classify power/audio recordings with accuracy of 97.78% on the development set and 94.03% on the test set. The accuracy obtained by our method outperforms the published results based on ENF-based classifiers [15], [16]. Moreover, the analyses of results have been carried out considering various conditions that confirm the robustness of our method. A comparative study of the computational overhead between an ENF-based classifier and the pole-matching classifier shows that the later is much faster than the former.

The organization of rest of the paper is as follows. Section II defines the problem statement and provides a concise overview on the dataset. Section III presents the factors that restrict the application of ENF based classifiers for the problem at hand. The following section gives a brief account of the poles of a given signal as a distinguishing feature for grid identification. The proposed method is described in detail in Section V. Section VI shows the experimental results and comparison with state-of-the-art methods. The paper is concluded with some remarks in Section VII.

II. Data Description and Problem Statement

In this work, the data provided for the IEEE Signal Processing Cup, 2016 [14] have been used. Two types of signals namely audio and power signals are provided for training purpose from 12 different grids and the grids were labeled as A to L. Table I depicts the location of the grids of the dataset along with their corresponding labels. The data consist of 8 grids with nominal frequency of 50 Hz and 4 grids with nominal frequency of 60 Hz. Power recordings were taken by a signal recorder connected to a power outlet using a step-down transformer. However, the audio signal is simply one made using an audio recorder not connected to anything. For each grid, 2 audio recordings of duration 30 min each were provided. The power data were provided for varying amounts of time (5 to 7 hours) for different grids. However, in the development and test data set, both audio and power signals were of 10 min duration. Development data consisted of 45 recordings and testing was done on 134 recordings from the 12 grids.

TABLE I
Location of the grids in the dataset.

Grid Label	Location	Nominal Frequency (Hz)
A	Texas	60
B	Lebanon	50
C	Eastern U.S.	60
D	Turkey	50
E	Ireland	50
F	France	50
G	Tenerife	50
H	India (Agra)	50
I	Western U.S.	60
J	Brazil	60
K	Norway	50
L	Australia	50

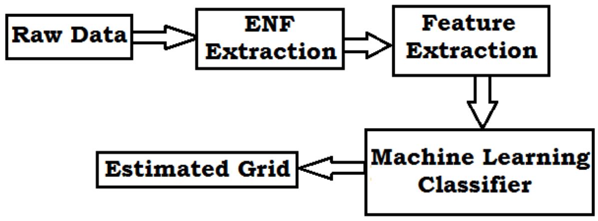


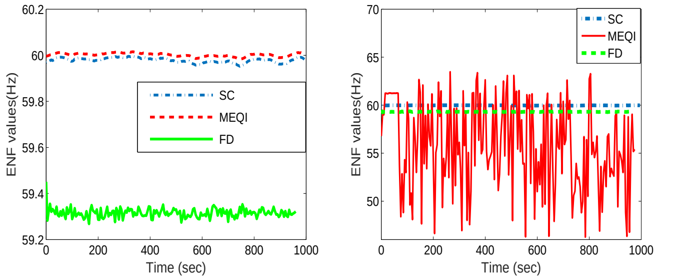
Fig. 1. Block diagram of a ENF based grid identification system.

The given power data are sinusoidal signals with clear periodicity. However, the audio recordings are extremely noisy and thus obtaining the time varying power frequency component from such a signal is very challenging. The problem is to build an efficient and accurate classifier system that can detect the region-of-recording for both power and audio recordings.

III. Limitations of ENF based Location Forensics

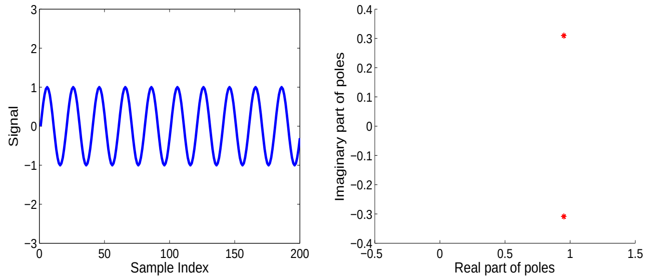
The state-of-the art methods for region-of-recording classification using media recordings are based on the ENF pattern of the corresponding grid [15]. Figure 1 shows the block diagram of such a method. At the onset, the ENF is extracted from the power/audio signal and then suitable features are selected from the obtained ENFs to train a machine learning classifier, such as support vector machines (SVM) as reported in [15]. Then the system is utilized to classify a given input signal.

Though the ENF is successfully applied for location forensics, ENF based approach has some intrinsic limitations. First of all, since ENF is considered as a power signature of the grid, it should have minimal dependency on the ENF extraction algorithm. But as it turns out, this does not hold true. Fig. 2a presents the ENFs extracted for the same power data using three different algorithms, namely, spectral combining (SC) [13], maximum energy with quadratic interpolation (MEQI) [17], frequency demodulation (FD) [18]. Since power signals are rather clean, we anticipate the ENFs to be identical. However, as we can see, ENF extracted using FD differs from the other two greatly. Again, although the patterns of the two



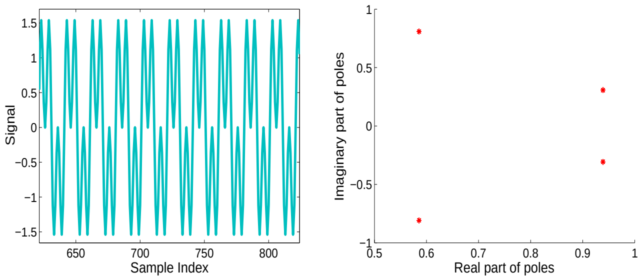
(a) ENF of a power signal (b) ENF of an audio signal

Fig. 2. ENF estimation results on a training data segment of the same signal using different algorithms (SC= spectral combining, MEQI= maximum energy with quadratic interpolation, FD=frequency demodulation).



(a) A portion of a 50 Hz sinusoidal signal (b) Pole-plot of the signal in (a)

Fig. 3. A simple sinusoid and its corresponding pole-plot.



(a) A portion of a signal containing a 50 and a 150 Hz sinusoid (b) Pole-plot of the signal in (a)

Fig. 4. A summation signal of two sinusoids and its corresponding pole-plot.

ENFs estimated using SC and MEQI are close, they will differ in features such as mean and range of ENF etc. For audio recordings, ENF extraction becomes all the more challenging since large noise is present in the signals and obtaining the actual ENF engulfed in such noisy conditions becomes quite difficult. Fig. 2b demonstrates the ENFs extracted using the above mentioned three techniques. Here, unlike power data the ENF of MEQI method differs significantly from those of the other two. Since MEQI method seeks to find the ENF based on maximum energy locations from the spectrogram, the ENF fluctuates a great deal in the case of an audio signal where we have much stronger spectral components even outside the nominal frequency and its harmonics. On the other hand, the SC method calculates the ENF by searching the peak in a combined harmonic frequency band and exhibits less fluctuations.

The ENF is not a physically acquired signal, rather a computed one from the raw data. The ENF signal may vary significantly depending on the selection of ENF estimation algorithm as well as parameter setting of the selected ENF extraction method. As a result, there is no unique set of features for the same raw input data and the performance of the classifier may degrade due to improper selection and parameter setting of the ENF extraction algorithm. On the contrary, if the classification can be made using the features extracted from the raw data itself, we can get rid of the dependency on the ENF extraction algorithm and it will also reduce the complexity of the classification system. This induced us to explore a novel algorithm to extract features from the given raw media recordings and build a classifier based on these primary features.

IV. Concept of Pole as a Distinguishing Feature

In this section, we explain the concept of pole extracted from a sinusoidal signal and show its relevance to capturing the frequency information directly from the raw data. Let us consider a simple 50 Hz sinusoidal signal depicted in Fig. 3a. If the signal is considered to be the output of an autoregressive (AR) system, we can estimate the complex poles associated with this model. Using the sampling frequency $f_s = 1000$Hz, the location of the poles can be obtained as $0.95 \pm j0.31$ in the z-plane. We now divide the signal into non-overlapping short-duration segments and estimate poles from each of these segments. All the poles estimated from different segments are superimposed on the z-plane as depicted in Fig. 3b. It is seen from the figure that the estimated poles form dense clusters around the theoretical locations. We also consider another signal which is the combination of two sinusoids having 50 and 150 Hz frequencies. Figs. 4a and 4b show the time-domain signal and superimposed pole plot, respectively. Once again, the poles from short-duration segments are clustered around four theoretically calculated points $(0.95 \pm j0.31)$ and $(0.59 \pm j0.81)$. Therefore, the frequency content is easily decipherable from the pole-plot as distinct frequency containing signal forms distinct clusters of poles, as evident from the above figures. This triggered our idea of developing a classifier based on the pole-locations of the grid signals.

We now examine the pole diagrams obtained from different grids with identical nominal frequency. The control mechanism and load variation pattern of a grid cause its instantaneous

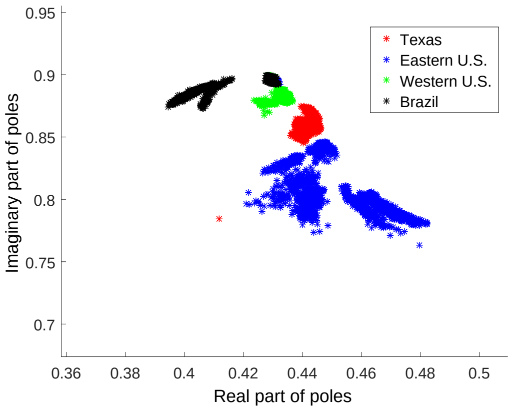


Fig. 5. Scatter plot of the estimated poles for a number of 60 Hz grids.

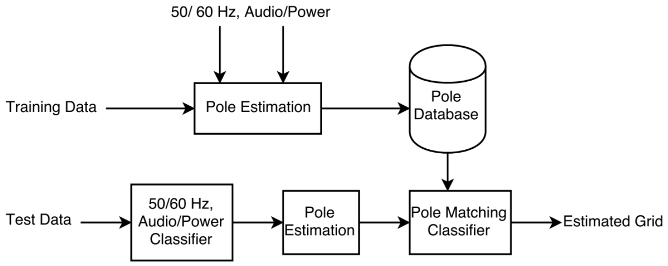
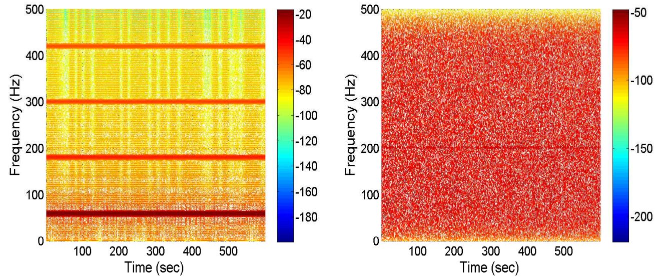


Fig. 6. Block diagram of the proposed pole-matching based region-of-recording classification system.



(a) STFT magnitude spectrum of a power signal

(b) STFT magnitude spectrum of an audio signal

Fig. 7. Spectrogram estimation using short-time Fourier transform (STFT) of media recordings.

frequency to vary over a small range. Consequently, the pole plots from the grid power data will not be very highly localized, rather a more scattered clustering is expected. Fig. 5 presents the plot of the estimated poles extracted from the power data of a number of 60 Hz grids. As expected, the poles of a grid signal are spread out over a somewhat wider region, however, their specific locations and pattern of clustering vary significantly depending upon the source grid. As a result, the pole features extracted from the raw data can easily be exploited to identify an unknown grid and we do not need the ENF signal for feature extraction.

V. Proposed Method

In this section, we describe the sequence of signal processing tasks carried out to classify a media recording and estimate its grid location. Fig. 6 depicts the functional block diagram of the proposed scheme. Since we have four types of signals, namely 50 or 60 Hz, audio or power, at the very onset, we classify the given signal in to one of these four categories. Then the poles are estimated from the training data and they are stored. In the testing phase, once again, the poles are estimated and then the pole-matching classifier is used to estimate a distance score between the test and pre-calculated training poles. The pre-defined training grid that gives the minimum distance score is identified as the grid label of the test data. In what follows, the proposed method is described in details.

A. Data Type and Nominal Frequency Detection

For any given data, before estimating the poles, it is passed through a nominal frequency (50/60 Hz) and data type (power/audio) detection step. This is done to enhance the inter-grid separability and thus improve the grid classification performance. The data type of a given signal can be easily detected in the frequency domain. Figs. 7a and 7b present the spectrogram of a power and an audio signal, respectively. As we can see, for the power signal, the spectrum is sparse and energy of the signal is mostly concentrated on the nominal and harmonic frequencies. However, for the audio signal, the energy of the signal is spread over the entire spectrum. As a result, we can easily detect the data type by observing the sparsity of the given signal in the frequency domain. The figures also reveal that the detection of nominal frequency is a trivial task for the power signal. However, for the audio signal, the ENF signal is engulfed in the noise and hence nominal frequency estimation becomes a challenge.

In order to determine the nominal frequency, we take the magnitude of the Fourier transform (FT) of the given signal. We split the spectrum into two parts: S_n and S_r. The S_n contains the frequency components that are located in the narrow spectral band around the two nominal frequencies and their second harmonics and S_r is the rest of the spectrum. Then we find the frequency of the maximum magnitude in S_n that is denoted by F_p. We now calculate two distance parameters for two different nominal frequencies as

$$D_{50} = \min(|F_p - 50|, |F_p - 100|) \tag{1}$$

and

$$D_{60} = \min(|F_p - 60|, |F_p - 120|). \tag{2}$$

If $D_{50} < D_{60}$, then 50 Hz is chosen as nominal, otherwise 60 Hz is chosen.

Again, for audio or power type detection, we compute the summation of frequency components of S_n and S_r that are denoted by P_n and P_r, respectively . To calculate P_r, we only consider the frequency components up to 125 Hz. Now, if the ratio of P_r to P_n is greater than a certain predefined threshold, T, the signal is classified as an audio signal. On the other hand, if this ratio is small meaning that the power is highly concentrated on the nominal frequency and its harmonics, then it is detected as a power signal.

B. Pole Estimation

Having identified the data type and its nominal frequency, next we estimate the pole locations of the data. For this purpose, we model the given raw data, $x(n)$, as the output of an autoregressive (AR) system as [19]

$$x(n) = \sum_{k=1}^{N} a_k x(n-k) + e(n) \tag{3}$$

where, N is the order of the AR system, a_k represent the AR coefficients and e(n) is the random excitation signal. Applying z-transform on both side of (3), we can get the transfer function, $H(z)$, of an all-pole AR system as

$$H(z) = \frac{X(z)}{E(z)} = \frac{1}{1 - \sum_{k=1}^{N} a_k z^{-k}} \tag{4}$$

The roots of the denominator polynomial of (4) provides us the poles of the given data $x(n)$. However, we need to evaluate the values of AR coefficients, a_k, in order to estimate the poles.

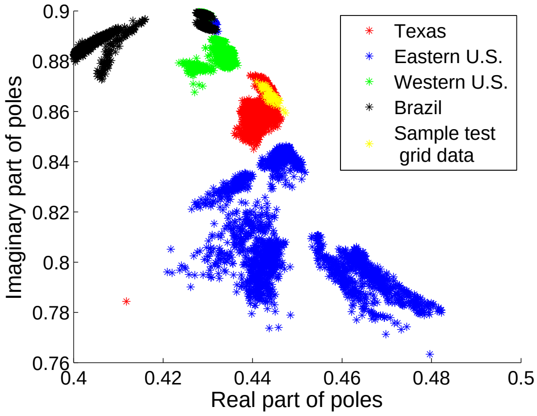


Fig. 8. Plot of the location of poles of a testing set superimposed on the pole plot of the training grids.

For this, we multiply both sides of (3) by $x(n-j)$ and take the expectation operation which gives

$$E[x(n)x(n-j)] = \sum_{k=1}^{N} a_k E[x(n-k)x(n-j)] + E[e(n)x(n-j)] \tag{5}$$

where $E[.]$ represent expectation operation. Now for the optimal model coefficients, the random excitation $e(n)$ is orthogonal to the past samples, and (5) gives

$$r_{xx}(j) = \sum_{k=1}^{N} a_k r_{xx}(j-k). \tag{6}$$

Given $N+1$ correlation values, (6) can be solved to obtain the AR coefficients a_k. Since the data is non-stationary, i.e., the grid frequency varies with time, it is not useful to compute the correlation values for the whole data. Therefore, we divide the given data into non-overlapping small segments assuming that the grid frequency remains stationary during that short period. Now for each segment, the correlation values are estimated that give the AR coefficients and finally we get the estimated poles. The poles extracted from the training dataset are stored as reference to be compared with those obtained from the test data for grid identification.

C. Pole-matching Classifier

First we explain the concept of our pole-matching classifier through an illustrative example as shown in Fig. 8. Here, the pole locations of a test signal is plotted along with four different training poles. The different training poles are marked with different colors and they are extracted from Texas, Eastern U.S., Western U.S., and Brazil grids. Clearly, the test poles (in yellow) match mostly with the poles of the Texas grid (marked as red). So, we can easily decide that this test signal belongs to the Texas grid. The proposed pole-matching classifier is based on this visual comprehension, i.e., we quantitatively identify the training grid that has the closest match with the test poles.

The test signal is first segmented into short duration non-overlapping blocks and the poles are estimated from each blocks. After that, the classification process is initiated by measuring distances between the pole points of test signal and those of the training dataset. Let, p_i is the i^{th} pole point of the test data and $\mathbf{g}^k$ is the vector containing all the pole points of the training data of grid k. Then, a distance function is calculated as follows

$$d_{i,j}^k = ||p_i - g_j^k||, \tag{7}$$

where $||.||$ denote the l_2 norm, $i = 1, 2, ...U$, $j = 1, 2, ...V_k$ and $k = 1, 2, ...R$. Here, U, V_k and R are total no. of testing poles, total no. of training poles of grid k and total no. of training grids, respectively. For each testing pole, only X no. of minimum distances are considered and stored in a vector. Thus, a vector $\mathbf{d}^k$ of length XU is obtained for the grid k considering all the test poles. We then calculate the average distance o^k from $\mathbf{d}^k$ as

$$o^k = \frac{1}{XU} \sum_{l=1}^{XU} d^k(l). \tag{8}$$

The final grid label for the test data is determined from the grid that gives the minimum distance. Mathematically, the identified grid is

$$\widehat{grid} = \arg\min_k (o^k). \tag{9}$$

VI. Experimental Results

In this section we present results to validate the effectiveness of our grid identification system, perform accuracy analysis and provide comparison with state-of-the-art techniques reported in the literature.

A. Parameter Settings

First, we mention the specified values of different parameters used in the experiment. The threshold, T, for power or audio data type detection was set to 3. The data was segmented into 10 sec blocks for pole estimations. The order of the AR system was 8 and 12 for power and audio data, respectively. The pole matching distance for a certain grid was calculated using $X = 2$, i.e., the mean distance of the two closest training poles from each test pole was used for measuring pole-matching distance.

B. Performance Measurement

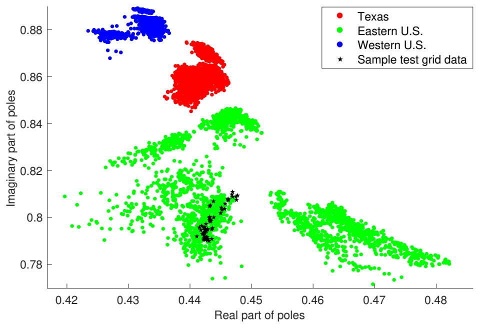
The performance of different grid detection algorithms are evaluated using the accuracy criterion defined as

$$Accuracy(\%) = \frac{No.\ of\ correctly\ classified\ grids}{Total\ no.\ of\ grids\ classified} \tag{10}$$

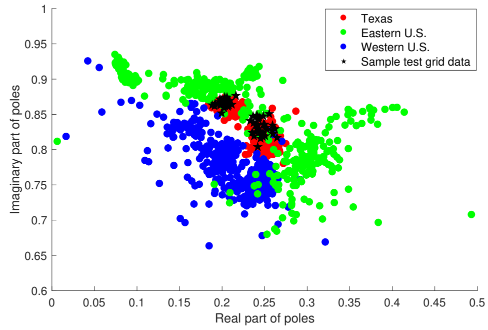
We also consider the computational time for classification purpose in evaluating the algorithms.

C. Verification of grid identification results

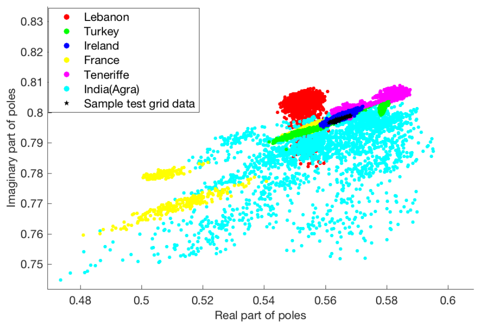
Here, we verify the effectiveness of our proposed grid identification algorithm using numerical examples. We consider four test samples representing power and audio data type and 50 or 60 Hz nominal frequencies. The meta data associated with the test samples are presented in Table II. First we evaluate the data type and nominal frequency classification results by our proposed method. Different parameters extracted from the test samples for this purpose are shown in Table III. It is seen that data type and nominal frequency can be easily determined from the extracted parameters. We now verify the



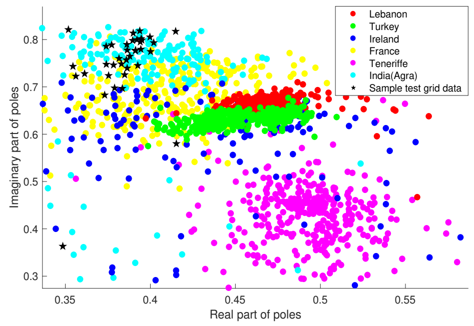
(a) Pole-plot of test sample 1.



(b) Pole-plot of test sample 2.



(c) Pole-plot of test sample 3.



(d) Pole-plot of test sample 4.

Fig. 9. Plot of location of poles of four test samples superimposed on the pole plot of the training grids.

TABLE II
META DATA OF FOUR TEST SAMPLES USED TO VERIFY THE GRID DETECTION RESULTS.

Sample No.	Data type	Nominal frequency	Region of recording
1	Power	60	Eastern US
2	Audio	60	Texus
3	Power	50	Ireland
4	Audio	50	Agra-India

TABLE III
NUMERICAL RESULTS OF DETECTION OF DATA TYPE AND NOMINAL FREQUENCY FOR THE TEST SAMPLES.

Sample No.	Data type		Nominal frequency		
	P_r/P_n	Data type	D_{50}	D_{60}	Nominal freq.(Hz)
1	0.375	Power	10.005	0.035	60
2	7.620	Audio	19.965	0.035	60
3	0.390	Power	0.052	10.052	50
4	10.741	Audio	0.323	19.677	50

proposed grid identification method through visual observation. Fig. 9 presents the location of poles extracted from the four test samples superimposed on the pole plot of the training grids. It is clearly observed that test poles (marked in black) has the maximum match with their corresponding training poles in all cases. Next we present quantitative performance analysis of the pole matching classifier. Table IV presents the calculated pole matching distance between the test poles and available training grids. The results show that the calculated distance is minimum for the actual grid and we are able to identify the region of recording in all the test cases.

D. Classification Accuracy

In this section, we provide grid identification results and accuracy comparison with state-of-the-art techniques. Among the 45 development data samples, the proposed algorithm is able to correctly identify 44 grids, which corresponds to 97.78% accuracy. On the 134 unknown test datasets, pole-matching is able to provide accurate estimation for 126 signals, resulting in an accuracy of 94.03%. The detailed grid-wise classification results on the testset are given in Table V. For comparison purpose, we evaluated the results on the same testset using the state-of-the art ENF-based SVM classifier, as described in [15]. The features were selected as given in [15] and the results are presented in Table V. The overall obtained accuracy for this method is 73.13%.

To further compare the results obtained using pole-matching with those achieved using an ENF-based machine learning classifier, we compare our results with the reported results in [16]. Saric *et al.* reported the results on the same database using five different ENF-based machine learning classifiers [16]. The classification results are given in Table VI. The different classifiers used were K$^{\text{th}}$ Nearest Neighbour (KNN), Random Forests (RF), Linear Perceptron (LP), Neural Network (NN) and Support Vector Machines (SVM). Table VI presents the performance of each of the classifiers on both power and audio signals. Notably, the accuracy reported here for the SVM classifier is higher than the one given in Table V ; this improved results were obtained using some extra features. As we can observe, the proposed pole-matching classifier gives higher overall accuracy compared to other classifiers. In terms of overall accuracy, only the RF classifier gives

TABLE IV
POLE MATCHING DISTANCE BETWEEN TEST POLES AND AVAILABLE TRAINING GRIDS.

Test sample	Available grids						Detected grid
60 Hz	A	C	I				
1	0.0373	0.0026	0.0268				C (Eastern US)
2	0.0024	0.0123	0.0124				A (Texus)
50 Hz	B	D	E	F	G	H	
3	0.0036	0.0014	9.13e-5	0.002	0.0013	0.0278	E (Ireland)
4	0.0817	0.0846	0.0244	0.0356	0.0882	0.0079	H (Agra-India)

TABLE V
CLASSIFICATION ACCURACY ON DIFFERENT GRIDS USING POLE-MATCHING AND ENF-BASED SVM CLASSIFIER [15] FOR THE TESTING SET.

Grid label	Pole matching			ENF-based SVM		
	Power # correct (# total)	Audio # correct (# total)	Accuracy(%)	Power # correct (# total)	Audio # correct (# total)	Accuracy(%)
A	9(9)	3(3)	100	9(9)	2(3)	92
B	8(8)	6(6)	100	8(8)	6(6)	100
C	7(8)	7(7)	93	5(8)	3(7)	53
D	6(7)	6(7)	86	6(7)	7(7)	93
E	7(8)	3(5)	77	7(8)	2(5)	69
F	8(8)	3(3)	100	6(8)	3(3)	82
G	9(9)	6(6)	100	6(9)	2(6)	53
H	8(8)	5(5)	100	8(8)	4(5)	92
I	7(8)	6(7)	87	6(8)	7(7)	87
J	4(4)	-	100	0(4)	-	0
K	4(4)	-	100	0(4)	-	0
L	4(4)	-	100	0(4)	-	0

TABLE VI
CLASSIFICATION ACCURACY (%) ON DIFFERENT GRIDS FOR DIFFERENT ENF-BASED CLASSIFIERS [16] ALONG WITH POLE-MATCHING RESULTS

Classifier	Power Data	Audio Data	Overall Accuracy
Pole-matching	96.02	94.43	94.97
KNN	83.29	70.58	78.31
RF	96.67	87	92.54
LP	92.31	66	82.08
NN	96	78.31	88.91
SVM	93	70	84.02

comparable performance. However, it has been reported that this classifier requires considerably long time to train and test the signal among all the ones given [16]. As a result, the authors themselves suggested to use NN or SVM instead of RF to speed up the process. In a later subsection, we analyze that our proposed algorithm is computationally much faster than the SVM classifier. So, clearly the pole-matching provides a better option as a region-of-recording classifier compared to the state-of-the-art ENF-based methods. Interestingly, the results reported in [16] are the best possible results obtained by tuning the parameters for all 4 types [16]. A consistent set of features was not used to obtain these results since in that case, accuracy dropped quite heavily [16]. However since pole-matching requires just two features for classification, namely the x and y coordinates of the poles in a 2D plane, the results obtained in this case are less sensitive to parameter tuning and desired accuracy is achieved without the need of extensive feature optimization for each class of signals which can be cumbersome as well as difficult for larger number of grids. Another point to note is, pole-matching provides significantly improved results for audio data. Since in real-life scenario where simultaneous power recording are available and actual classification has to be made on media recordings solely, clearly the proposed classifier outperforms the state-of-the art ENF-based ones by a fair margin.

E. Effect of Length of Testing Data on Accuracy

In our experimental data, testing data signals were of 10 min duration. However, in real-life scenario, the length of the signal available to track a certain location may vary and length may be quite small as well. So, to further explore the efficacy of the proposed method against variation in the length of testing data, we curtailed the testing signals in various proportions and determined the classification accuracy. The results are given in Table VII. Interestingly, the results obtained using the pole-matching classifier is quite consistent even when the testing data are as small as just 1 min duration. For comparison purpose, results for the SVM classifier is also given. Notably, the pole-matching classifier is specially effective when the media clip to be tested is of very short duration. As we can see, when the testing data was taken of below 5 mins, result of SVM dropped sharply as for such a short duration, accurate feature extraction and classification becomes quite difficult. However, since even for a very short duration signal, we obtain

TABLE VII
EFFECT OF LENGTH OF TESTING DATA ON THE ACCURACY OF POLE-MATCHING AND SVM.

Duration of testing data (min)	Pole-matching Accuracy (%)	SVM Accuracy (%)
10	94.97	73.13
9	94.03	68.66
8	94.03	71.64
7	94.03	68.66
6	94.03	67.91
5	94.03	67.91
4	93.28	43.44
3	92.54	31.72
2	91.04	15.03
1	90.30	13.83

some testing poles and the algorithm matches the location of those poles with the training ones, the pole-matching algorithm does not suffer from the testing data available for a limited time. This could be very useful for crime detection purposes where an audio clip from the criminals may be of just 1 or 2 mins and in those cases, the proposed pole-matching method is still able to identify the grid location.

F. Computational Complexity

Another important parameter of evaluation is the computational load. We used a MacBook with Intel core i5, 2.7 GHz processor and 8 GB RAM. The average time taken to classify a test recording of 10 min duration is $0.7354 \pm 0.02\ s$ (mean$\pm$ SD). However, the same measure for the ENF-based SVM classifier was $4.6055 \pm 0.09\ s$. So, the proposed method is more accurate as well as faster than the ENF-based method.

G. Pole-matching Versus a Traditional ML classifier

For the classification purpose, here we have used a simple yet robust and effective distance based pole-matching classifier. However, if the training poles are used as features and a K-Nearest Neighbors (KNN) classifier is trained with these pole locations, then classification is also possible. But this gives a slightly inferior performance with an accuracy of around 90.30%. So the proposed pole-matching classifier provides better performance compared to a traditional ML classifier.

VII. Conclusion

We have developed a classification system based on pole-matching criterion that can identify the grid-of-origin of a power or audio signal. The approach makes use of the estimated poles calculated from power or audio data that are used in a distance based classification system. Whereas the conventional methods estimate ENF signal first and then extract features from ENF, our approach calculates pole features directly from raw data. The results show that the proposed approach can achieve more accurate estimates for both power and audio data than the conventional ENF based approach. We were able to achieve an average accuracy of 96.02% on grid detection from power recordings and an average accuracy of 94.43% on grid detection from audio recordings. We have shown that our approach can achieve robust results compared to conventional approaches with significantly small amount of test data and it requires much lesser amount of computation time. In our future work, we plan to propose a hybrid model based on pole-matching and ENF-based machine learning classifiers to further enhance the performance.

References

[1] *UNICEF: Child Protection From Violence, Exploitation and Abuse Child Trafficking.* [Online]. Available: http://www.unicef.org/ protection/57929_58005.html.

[2] R. Garg, A. Hajj-Ahmad, and M. Wu, “Geo-location estimation from Electrical Network Frequency signals,” In *ICASSP,* pp. 2862–2866, 2013.

[3] N. Fechner and M. Kirchner, “Background Noise as a Carrier of ENF Artifacts in Mobile Device Audio recordings”, 2014.

[4] C. Grigoras, “Digital audio recording analysis: The electric network frequency (ENF) criterion,” *Int. J. Speech Lang. Law,* vol. 12, no. 1, pp. 63-–76, 2005.

[5] C. Grigoras, “Applications of ENF criterion in forensic audio, video, computer and telecommunication analysis,” *Forensic Sci. Int.,* vol. 167, pp. 136-–145, Apr. 2007.

[6] A. J. Cooper, “The electric network frequency (ENF) as an aid to authenticating forensic digital audio recordings an automated approach,” In *33rd Audio Eng. Soc. Int. Conf. Audio Forensics-Theory and Practice,* Jun. 2008.

[7] A. J. Cooper, “An automated approach to the electric network fre- quency (ENF) criterion - theory and practice,” *Int. J. Speech Lang. Law,* vol. 16, pp. 193-–218, Nov. 2009.

[8] C. Grigoras, “Applications of ENF analysis in forensic authentication of digital audio and video recordings,” *J. Audio Eng. Soc.,* vol. 57, pp. 643-–661, Sep. 2009.

[9] A. J. Cooper, “Further considerations for the analysis of ENF data for forensic audio and video applications,” *Int. J. Speech Lang. Law,* vol. 18, pp. 99-–120, Jun. 2011.

[10] R. Garg, 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,* pp. 23–32, ACM 2011.

[11] Y. Liu, Z. Yuan, P. Markham, R. Conners, and Y. Liu, “Application of power system frequency for digital audio authentication,” *IEEE Trans. Power Del.,* vol. 27, pp. 1820—1828, Oct. 2012.

[12] O. Ojowu, J. Karlsson, J. Li, and Y. Liu, “ENF extraction from digital recordings using adaptive techniques and frequency tracking,” *IEEE Trans. Inf. Forensics Secur.,* vol. 7, pp. 1330-–1338, Aug. 2012.

[13] A. Hajj-Ahmad, R. Garg, and M. Wu, “Spectrum combining for ENF signal estimation,” *IEEE Signal Process. Lett.,* vol. 20, pp. 885—888, Sep. 2013.

[14] Signal Processing Cup 2016, database. [Online]. Available: http://sigport.org/documents/information-mast-enf-power-signature-dataset.

[15] A. Hajj-Ahmad, R. Garg, and M. Wu, “ENF-based region-of-recording identification for media signals,” *IEEE Trans. Inf. Forensics Security,* vol. 10, no. 6, pp. 1125-–1136, Jun. 2015.

[16] Ž. Šarić, A. Žunić, T. Zrnić, M. Knežević, D. Despotović, and T. Delić, “Improving location of recording classification using Electric Network Frequency (ENF) analysis,” In *Intelligent Systems and Informatics (SISY), 2016 IEEE 14th International Symposium on,* pp. 51–56. IEEE, 2016.

[17] 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.

[18] L. Dosiek, “Extracting electrical network frequency from digital recordings using frequency demodulation,” *IEEE Signal Processing Letters,* vol. 22, no. 6, pp. 691–695, Jun. 2015.

[19] A. V. Oppenheim, and R.W. Schafer, *Discrete-Time Signal Processing,* Prentice-Hall, Englewood Cliffs, NJ, 1989, pp.713–718.

Sayeed Shafayet Chowdhury was born in Bangladesh in 1993. He received the B.Sc. degree in electrical and electronic engineering from Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh, in 2016.

Currently he is working as a lecturer in the Department of Electrical and Electronic Engineering, BUET where he is also working as a Researcher with Nanotechnology Research Lab. His research interests include biomedical signal processing, pattern recognition, nanomaterials characterization.

Mr. Chowdhury has been a member of IEEE Signal Processing Society since 2014. He was a member of the team which won the 1st place in the IEEE Signal Processing Cup, 2016 as well as the 2nd place of the same competition in 2015, under the supervision of Prof. Mohammad Ariful Haque.

Jayanta Dey was born in Bangladesh in 1994. He completed his B.Sc. degree in Electrical and Electronic Engineering from Bangladesh University of Engineering and Technology, Dhaka, Bangladesh in 2017 and is currently affiliated with this department.

His current research interest includes bio-medical signal processing, machine learning, adaptive filtering, speech processing and compressed sensing. His undergrad thesis was on Blind Multi-Channel Ultrasound Deconvolution.

He participated in the IEEE Signal Processing Cup, Shanghai, China in 2016, and his team secured the first place in the competition. He has been a member of IEEE Signal Processing Society since 2015.

Rakib Hyder Rakib Hyder received the B.Sc. degree in electrical and electronic engineering from Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh, in 2016.

He has been a student member of IEEE and IEEE signal processing society since 2014. His research interest includes digital speech processing and classification, digital image processing and pattern recognition, machine learning, biomedical signal processing, and plasmonic bio-sensor design and characterization.

R. Hyder attended 40th and 41st IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) as a part of the IEEE Signal Processing Cup Competition. His team won the 1st place in IEEE Signal Processing Cup, 2016 on location forensics using electrical network frequency (ENF) signal under the supervision of Prof. M. A. Haque. He was also a member of the team which won the 2nd place in the IEEE Signal Processing Cup, 2015 on heart rate estimation from PPG signal under the same supervisor. He and his teammates won the best project award in Tensilica Xtensa Embedded DSP Design Contest, 2016 for their work on double talk detection.

Ashraful Islam was born in Bangladesh in 1993. He completed his B.Sc. degree in Electrical and Electronic Engineering from Bangladesh University of Engineering and Technology, Dhaka, Bangladesh in 2017 and is currently affiliated with this department.

His current research interest includes signal processing, machine learning, speech processing, computer vision and compressed sensing.

He participated in the IEEE Signal Processing Cup, Shanghai, China in 2016, and his team secured the first place in the competition. He has been a member of IEEE Signal Processing Society since 2015.

Mohammad Ariful Haque received the B.Sc., M.Sc., and Ph.D. degrees in electrical and electronic engineering all from the the Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh, in 2003, 2005 and 2009, respectively.

In 2003, he joined BUET as a Lecturer in the Department of Electrical and Electronic Engineering. He is currently working as Professor in the same department. He was a postdoctoral fellow at Concordia University, Montreal, QC, Canada in 2012. His research interests are digital speech processing and classification, robust speech and speaker recognition, biomedical signal processing, adaptive filtering, and medical imaging.

Prof. Haque has been the faculty mentor of one of the finalist teams of the IEEE Signal Processing Cups in 2014 to 2016 and his team came in 1st, 2nd, and 1st place the last three years.