

## Research Article

# Power Grid Estimation Using Electric Network Frequency Signals

**Woorim Bang and Ji Won Yoon** 

*Graduate School of Information Security, Korea University, Seoul, Republic of Korea*

Correspondence should be addressed to Ji Won Yoon; [jiwon\\_yoon@korea.ac.kr](mailto:jiwon_yoon@korea.ac.kr)

Received 7 April 2019; Revised 12 July 2019; Accepted 25 August 2019; Published 24 September 2019

Guest Editor: Sebastian Schrittwieser

Copyright © 2019 Woorim Bang and Ji Won Yoon. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The electric network frequency (ENF) has a statistical uniqueness according to time and location. The ENF signal is always slightly fluctuating for the load balance of the power grid around the fundamental frequency. The ENF signals can be obtained from the power line using a frequency disturbance recorder (FDR). The ENF signal can also be extracted from video files or audio files because the ENF signal is also saved due to the influence of the electromagnetic field when video files or audio files are recorded. In this paper, we propose a method to find power grid from ENF signals collected from various time and area. We analyzed ENF signals from the distribution level of the power system and online uploaded video files. Moreover, a hybrid feature extraction approach, which employs several features, is proposed to infer the location of the signal belongs regardless of the time that the signal was collected. Employing our suggested feature extraction methods, the signal which extracted from the power line can be classified 95.21% and 99.07% correctly when ENF signals have 480 and 1920 data points, respectively. In the case of ENF signals extracted from multimedia, the accuracy varies greatly according to the recorded environment such as network status and microphone quality. When constructing a feature vector from 120 data points of ENF signals, we could identify the power grid had an average of 94.17% accuracy from multimedia.

## 1. Introduction

A supply frequency of electrical power in power distribution networks is called electric network frequency (ENF). In most countries, the ENF has a nominal value of 50 or 60 Hz. It is essential to maintain a constant frequency value to prevent a total collapse of the electric system and to keep the grid stable. In order to balance between aggregate generation and aggregate demand across the interconnection, the ENF signals have small variations every moment [1]. This change is a normal phenomenon for stable operation of the power grid and is acceptable if the value of ENF signal is less than the threshold. Because of these characteristics of the ENF, the ENF has uniqueness according to time and location.

There are two main methods for obtaining the ENF signals. One is to collect the ENF signals utilizing special equipment such as frequency disturbance recorders (FDRs). The other is to extract them from multimedia. When recording audio files or video files, the ENF signals are also being stored together because of the electromagnetic

influences from the power line [2]. In other words, the ENF signals can be included in multimedia. Thus, the ENF signals in video files or audio files can be applied to digital forensics to determine whether the file is modified or not [3, 4]. In order to demonstrate the authenticity of the file, reference database is required to store the value of the ENF signals extracted from the same time and the same area as the file is required. Since most forensics is done after a series of events, a database of the ENF signals requires to be established for almost all regions and times for the investigation. This makes it difficult to use the ENF signal for digital forensics. If we can suggest which region an arbitrary ENF signal belongs to, it will increase the value of applying the ENF signal to the digital forensics.

In most countries, the location privacy is regarded as one of the important personal information. Numerous people around the world freely upload and download numerous photos, audio files, and video files through social network services, video sharing web services, and live webcam services such as Facebook [5], Instagram [6], YouTube [7],

Earthcam [8], Skyline Webcams [9], and explore [10]. To comply with personal privacy regulations, web service providers remove GPS information from uploaded photo or video file's metadata before posting on online or obtain user's agreement to release their location information. However, there is a research that not only the location information exists in the metadata of audio files or video files but also the location could be inferred from ENF signals in audio files or video files [11, 12]. Kim et al. [12] conducted a study on the construction of a national-scale ENF map using multimedia recorded in several areas. Jeon et al. [11] suggested intragrid location estimation of audio files and video files from Voice over Internet Protocol (VoIP) applications and streaming services using ENF map. Both studies [11, 12] require ENF signals collected at the same for inferring location.

In this paper, we introduce the model that extracts features from ENF signals and identifies the power grid where ENF signals are extracted from. Using our proposed method, we can estimate the location even if ENF signals are not obtained from close time to the train dataset because we extract features by power grids. We evaluated our model on three power grids in the United States and the Eastern, the Western, and the Texas power grid. We collected ENF signals from these power grids. We constructed two types of ENF signal datasets. The first is composed of ENF signals at the distribution level of the power systems. Another is extracted from video files and audio files provided by worldwide live webcam services, such as Skyline Webcam [9], Earthcam [8], and explore [10] and video sharing web service YouTube [7] on the Internet.

To extract enough features to classify the power grid, we mainly merge features extracted from three different methods which consist of the autoregressive coefficients using the Yule-Walker method, the Shannon entropy on terminal nodes of maximal overlap discrete wavelet packet transform, and the multiscale variance of maximal overlap discrete wavelet transform. Finally, a feature vector is constructed by adding the variance of ENF signals into the merged feature from three main feature extraction algorithms. The support vector machine (SVM) and XGBoost [13] are compared as a method of classifying the power grid using extracted features. We finally evaluated the experimental results of our proposed method depending on the number of ENF signals. Through our proposed method, we have been able to identify power grids with high accuracy.

Our main contributions are summarized as follows:

- (i) We can estimate the power grid of ENF signals extracted from the power line and online multimedia. By analyzing ENF signals, the power grid can be efficiently identified without reference data which are composed of ENF signals extracted from the same time and the same region of recorded files.
- (ii) Using our proposed feature extraction algorithm, we can perform analysis on much less data than the number of raw signals. We propose to construct a feature vector from the three main feature extraction methods.

- (iii) We studied the factors that influence the estimation of the power grid through ENF signals. First of all, we have investigated the power grid identification accuracy according to the number of ENF signals. In addition, we analyzed the power grid estimation accuracy conducted on the harmonic components used in extracting ENF signals from online multimedia.

The structure of this paper is as follows. In Section 2, we introduce the related work of estimating location using ENF signals. Moreover, we present a related research that extracts ENF signals from a multimedia file with sound recorded such as audio files or video files. Section 3 shows the two main background knowledge required in this paper. In Section 4, we explain the two types of the ENF signal dataset used in our proposed method. We propose our algorithm to estimate the power grid using ENF signals in Section 5. We evaluate our proposed technique using the accuracy and F1 score in Section 6. In Section 7, we discuss the importance of our experiments and the limitations of our work. Finally, we conclude this paper in Section 8.

## 2. Related Work

In this section, we present the related research of location estimation using ENF signals and the study of extracting ENF signals from multimedia.

*2.1. Location Estimation Using ENF Signals.* Over the past few years, researches have been consistently proposed to classify ENF signals by location [2, 14–16]. Yao et al. [15] suggested the wavelet-based signature extraction and feed-forward artificial neural network-based machine learning. They required a minimum length of 15 minutes of uninterrupted frequency data (9000 continuous data points) using the FNET/GridEye system. In real-world environments, it is not easy to gather data at every 0.1 second intervals for 15 minutes due to temporal equipment failures, network failures, or server load. When we collect data using the same system, the FNET/GridEye system, data loss often occurred. In Section 4.1, we introduce a data preprocessing technique that can use data even if it involves some data loss.

Hajj-Ahmad et al. [2, 14] and Šarić et al. [16] suggested a classification method for ENF signals using several feature extraction algorithms. Šarić et al. [16] used between 305 and 480 minutes of power signals and 60 minutes of audio recording to classify the power grid. Hajj-Ahmad et al. [2, 14] suggested analyzing 96 samples of ENF signals from power signals and audio recordings. In previous studies [2, 14, 16], there was a lack of research on estimation power grid accuracy according to the number of ENF data points. In Section 6, we evaluated the power grid identification accuracy for various ENF signal length.

*2.2. Extraction of ENF Signals from Multimedia.* Kim et al. [12] and Hajj-Ahmad et al. [17] have studied methods for extracting ENF signals from multimedia recordings. Hajj-

Ahmad et al. [17] have tested with extracting ENF signals from recordings using Multiple Signal Classification (MUSIC) and Estimation of Signal Parameters via Rotational Invariance Techniques (ESPRIT) methods. Kim et al. [12] used the Quadratically Interpolated Fast-Fourier Transform (QIFFT) method [18] and multitone harmonics technique to obtain ENF signals from multimedia data. In this paper, we extracted the ENF signals using QIFFT and multitone harmonics in the video files uploaded on the Internet. Moreover, we examined the relationship between the number of harmonic frequency bands which are used to extract ENF signals and power grid identification accuracy.

### 3. Background

In this section, we describe existing techniques that identify the power grids and methods for extracting ENF signals from online multimedia.

**3.1. Electric Network Frequency (ENF).** The electric network frequency (ENF) is the supply frequency of power grids. The ENF signals exist in nominal value at either 50 or 60 Hz depending on geographic location [19]. If the value of ENF signals exceeds the threshold value, it causes all equipment operate using electricity. Therefore, the value of the supply frequency should always be constant. However, the ENF signals fluctuate a little over around the dominant value because of the imbalance of the supply and demand of power grids. Thus, the ENF signals have uniqueness by time and location.

**3.2. Extraction of ENF Signals from Online Multimedia.** Kim et al. [12] applied multitone harmonics method and QIFFT (Quadratic Interpolated Fast-Fourier Transform) method to the audio of multimedia data to obtain ENF signals. When extracting ENF signals from multimedia data, the most important task is how to reduce a noise. Therefore, they applied multitone harmonics as a technique to reduce noise and obtain more accurate ENF signals. The ENF signal is generally present at 50/60 Hz as a fundamental frequency as well as in harmonic frequency bands which are integer multiplication of fundamental frequency (Figure 1). Kim et al. [12] selectively used harmonic frequency bands and applied QIFFT to extract the maximum value of ENF signals from each window.

### 4. Dataset Description

This section explains the two main ENF datasets that were used to evaluate the ENF signal classification. In general, there are two approaches to acquire ENF signals. One is obtained from the distribution level of the power system. The other is extracted from audio files and video files.

**4.1. ENF Signal Extracted from Power Grid Distribution Level.** We acquired the ENF signal extracted from the distribution level of the power system using the FNET/GridEye system [20]. A wide-area frequency monitoring network (FNET)

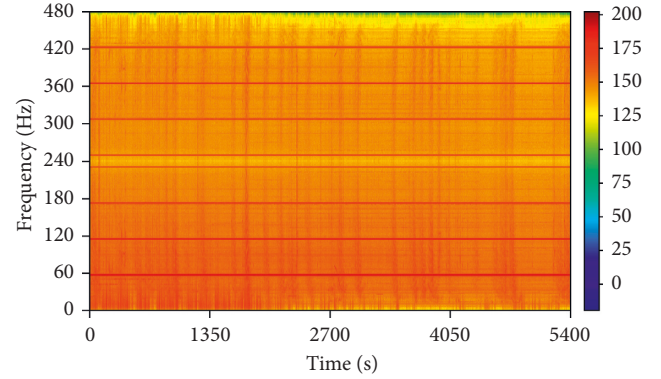


FIGURE 1: ENF signals at fundamental frequency (60 Hz) and harmonic frequency bands (around 120, 180, 240, 300, 360, 420, and 480 Hz) in multimedia file.

consists of the frequency disturbance recorder (FDR) data collected from the three interconnections in North America [21] and the Eastern power grid, the Western power grid, and the Texas power grid. The FDRs, members of the Phasor Measurement Unit (PMU) family, measure the voltage phase angle, amplitude, and frequency from a single-phase voltage source at 120 V distribution level with GPS synchronized. 129 units of FDR are installed in the Eastern Interconnection, 7 units in the Texas Interconnection, and 45 units in the Western Interconnection in the United States. The FNET had provided a web service which dynamically updates the frequency data every 4 seconds received from the FDRs until April 30, 2018 [22]. We collected electric network frequency data for more than 180 United States regions provided by FNET web service every 4 seconds for 49 days.

Despite of the high accuracy of the FDR ( $\pm 0.0005$  Hz) [21], we find that there often exist invalid or missing data in the collected data due to hardware failures such as GPS signal loss, aging of equipment or network interruption and web server failure. Therefore, data preprocessing is required for enhancing the accuracy of the experiment because the time interval for collecting data during 49 days is not always constant at 4 seconds. First of all, we selectively choose areas which collected over 99% for data accuracy; 52 regions in the Eastern power grid, 19 regions in the Western power grid and 2 regions in the Texas power grid. For each selected region, linear interpolation is employed every second based on the collected data to fill the missing data. After linear interpolation, we subsampled every 15 seconds for 49 days not only to reduce the features of the data but also to avoid distortion due to data loss. Accordingly, it means that there are 4 data points per minute and 5760 data points per day for each FDR installed area.

**4.2. ENF Signal Extracted from Online Multimedia.** To extract ENF signals from multimedia data, we gathered audio files and video files from online web services. To classify the location, we collected videos from four web services: Skyline Webcams [9], EarthCam [8], explore [10], and YouTube [7]. Skyline Webcams [9], EarthCam [8], and explore [10] are worldwide live webcam services which stably provide videos



enough to extract the ENF signals as well as including location information about videos. In addition, we collected multimedia data from YouTube, which is used by many users worldwide for video sharing. In YouTube which does not provide accurate location information, we selectively gathered videos such as concerts, festivals, and studio broadcasts of local broadcasting stations which are certain of the location where the video was taken. When we gathered multimedia, we only need data from the sound to obtain the ENF signals by analyzing the frequency bands. Therefore, we set the sampling rate to 1000 Hz and stored only sounds to utilize storage space efficiently.

In contrast to the ENF signals obtained at the distribution levels, which only have the fundamental frequency (50 or 60 Hz), multimedia files contain a wide range of frequencies as well as fundamental frequency. Thus, unlike the ENF signals collected at the distribution level, additional steps are required to obtain the ENF signals from the multimedia. We used multitone harmonics method and QIFFT method to extract ENF signals from videos proposed by Kim et al. [12]. Some files have ENF signals in either fundamental frequency or harmonics frequency bands. However, some of the files are included in neither the fundamental frequency nor the harmonic frequency bands. For a file including ENF signals at least one frequency band, the number of frequency bands used to extract the ENF signal is shown in Table 1. When ENF signals are extracted using all harmonic frequencies including fundamental frequency, they are extracted using a total of 8 frequency bands. When the sampling rate is 1000 Hz and the window size for QIFFT is set to 296000, about 80% of the 374 files have ENF signals in all frequency bands. The distribution of the number of harmonic bands used and the harmonic band frequency used is shown in Figure 2. The extraction of the ENF signal using eight harmonic band frequencies is much more than the extraction of the ENF signal using two to seven harmonic band frequencies. Thus, it displays the employed frequency information corresponding to using 2 ~ 7 frequency bands from the multimedia files to extract ENF signals.

We conducted experiments on 374 video files which were taken in the United States of America and include ENF signals. The videos were collected for about two years, and the length of the video recorded is varied from 15 minutes to over 4 hours (see details of the videos in Table 2).

## 5. Proposed Approaches

In this section, we introduce how to construct a feature vector which are composed of three different feature extraction methods for the ENF signals and a classification method to the power grid.

**5.1. Feature Extraction on ENF Signals.** We present three different methods for extracting features from ENF signals. To obtain the envelope of the signal spectrum feature, we employ the autoregressive model in Section 5.1.1. An approach for extracting time-invariant properties from a signal is given in Section 5.1.2. The maximal overlap discrete wavelet packet transform was used to determine the local

TABLE 1: Number of frequency bands employed to extract ENF signals in videos (the videos are limited to including ENF signals).

Number of frequency bands	1	2	3	4	5	6	7	8
Number of multimedia	0	8	7	12	9	12	28	298

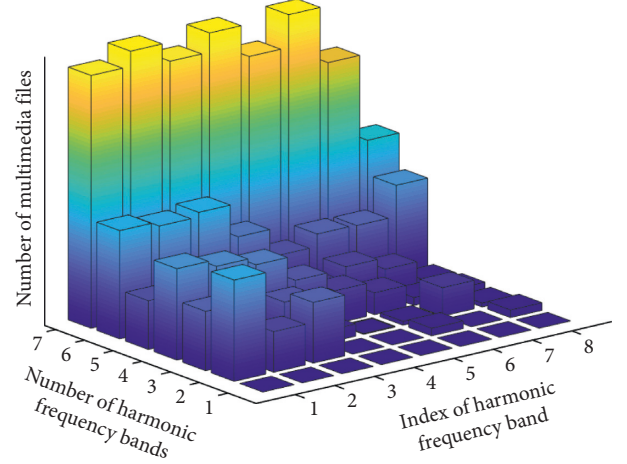


FIGURE 2: The distribution of the number of harmonic bands and the information of extracted harmonic band frequency used to restore the ENF signals from the multimedia files.

TABLE 2: Number of videos by recorded length.

Time (hour)	0.25 ~	0.5 ~	1 ~	2 ~	3 ~	4 ~
Number of multimedia	374	199	20	3	2	1

characteristics in Section 5.1.3. Using the results obtained by three feature extraction methods from ENF signals, we were able to construct a feature vector.

**5.1.1. Autoregressive (AR) Coefficients.** The basic idea of the autoregressive model is that the value at time  $t$ ,  $x_t$ , depends linearly on its own  $p$  previous values,  $x_{t-1}, x_{t-2}, \dots, x_{t-p}$ , with a white noise  $\varepsilon_t$  at time  $t$ . The notation  $AR(p)$  indicates an autoregressive model of order  $p$ .  $AR(p)$  defined as

$$x_t = \sum_{i=1}^p \varphi_i x_{t-i} + \varepsilon_t, \quad (1)$$

where  $\varphi_i$  is a parameter of the model and  $\varepsilon_t$  is a white noise. To calculate the parameters, we used the Yule-Walker equation which is also called the autocorrelation method. The Yule-Walker equation for the autoregressive model is based on minimizing the forward prediction error in the least-square sense. The Yule-Walker equation can be represented by a matrix form as follows:

$$\begin{pmatrix} r_0 & r_1 & r_2 & \cdots & r_{p-1} \\ r_1 & r_0 & r_1 & \cdots & r_{p-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{p-1} & r_{p-2} & r_{p-3} & \cdots & r_0 \end{pmatrix} \begin{pmatrix} \varphi_1 \\ \varphi_2 \\ \vdots \\ \varphi_p \end{pmatrix} = \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_p \end{pmatrix}, \quad (2)$$

where  $r_d$  is the autocorrelation coefficient at delay  $d$  and the diagonal value  $r_0$  sets to 1. As a result, the Yule–Walker equation gives the same number of parameters as the order of the autoregressive model. In our experiment, the order of the autoregressive model,  $p$ , was set to 12. Therefore, 12 features can be obtained per ENF segment.

**5.1.2. Multiscale Variance of Maximal Overlap Discrete Wavelet Transform.** In order to find the time-invariant property, we used the maximal overlap discrete wavelet transform (MODWT) decomposition which retains all possible times at each time scale [23]. Similar to the discrete wavelet transform (DWT), the MODWT can be used in a multiresolution analysis. Therefore, we perform the MODWT of the signals using the Daubechies wavelet with 2 of vanishing moments (db2) down to the maximum level. The wavelet and scaling filter coefficients at level  $j$  result from the inverse discrete Fourier transform (DFT) of a product of DFTs. The DFTs in the product are the signal's DFT and the DFT of the  $j^{\text{th}}$  level wavelet or scaling filter. Let MODWT wavelet and scaling filters of the length  $N$  DFTs be  $H_k$  and  $G_k$ . The  $j^{\text{th}}$  level wavelet filter is defined as [24]

$$\frac{1}{N} \sum_{k=0}^{N-1} H_{j,k} e^{(i2\pi nk)/N}, \quad (3)$$

where

$$H_{j,k} = H_{2^{j-1}k \bmod N} \prod_{m=0}^{j-2} G_{2^m k \bmod N}. \quad (4)$$

The  $j^{\text{th}}$  level scaling filter is defined by [24]

$$\frac{1}{N} \sum_{k=0}^{N-1} G_{j,k} e^{(i2\pi nk)/N}, \quad (5)$$

where

$$G_{j,k} = \prod_{m=0}^{j-1} G_{2^m k \bmod N}. \quad (6)$$

After MODWT, we got the unbiased estimates of the wavelet variance by scale as features. The wavelet variance is computed as follows:

$$\hat{v}_j^2 = \frac{1}{N_j} \sum_{t=0}^{N_j-1} W_{j,t}^2, \quad (7)$$

where  $N_j$  is the number of coefficients at level  $j$  and  $W_{j,k}$  is the wavelet coefficients of level  $j$  at  $k^{\text{th}}$  node [23].

In order to obtain the unbiased estimates of the wavelet variance by scale as features, we have to perform the maximal overlap discrete wavelet transform (MODWT) of the signals. Figure 3 displays the MODWT of the signals using the Daubechies wavelet with 2 of vanishing moments down to the maximum level when the 12-hour ENF signals are given. After performing MODWT, we estimated the wavelet variance. In our experiment, we could obtain less than  $\lceil \log_2 N + 1 \rceil$  features when the ENF segment consists of  $N$  data samples.

**5.1.3. Shannon Entropy on the Terminal Nodes of Maximal Overlap Discrete Wavelet Packet Transform.** The maximal overlap discrete wavelet packet transform (MODWPT) does not perform downsampling unlike discrete wavelet transform and has time-invariant properties and shifting invariant properties. Moreover, the coefficients of the MODWPT can include the local characteristics information of the signal. In spite of these properties, it is difficult to directly use it as a feature for classification because of the large number of coefficients [25]. In order to analyze the relevance of the coefficients, we employed the Shannon entropy. As a result, we calculate the Shannon entropy from the terminal node of the MODWPT as a feature.

We denote the ENF time series to be MODWPT,  $x_t$  ( $t = 0, 1, \dots, N-1$ ), by  $\mathbf{x}$ , where  $N$  is the length of the series. The sequence of MODWPT coefficients at level ( $j = 0, 1, \dots, J$ ) and frequency index ( $n = 0, 1, \dots, 2^j - 1$ ) represents as  $W_{j,n} = \{W_{j,n,t}, t = 0, 1, \dots, N-1\}$ . To compute the maximal overlap wavelet packet coefficients for levels, we can recursively filter the wavelet packet coefficients at the previous stage. Let  $\mathbf{W}_{0,0} \equiv \mathbf{x}$  and given the series  $W_{j-1,n/2,t}$  of length  $N$ , we can derive  $W_{j,n,t}$  using [26]

$$W_{j,n,t} \equiv \sum_{l=0}^{L-1} \tilde{f}_{n,l} W_{j-1,n/2,(t-2^{j-1}l) \bmod N}, \quad (8)$$

$$\tilde{f}_{n,l} = \begin{cases} \tilde{g}_l, & \text{if } n \bmod 4 = 0 \text{ or } 3, \\ \tilde{h}_l, & \text{if } n \bmod 4 = 1 \text{ or } 2, \end{cases}$$

where  $L \leq N$ ,  $\tilde{g}_l = g_l/\sqrt{2}$ , and  $\tilde{h}_l = h_l/\sqrt{2}$ . The scaling (low-pass) filter is denoted as  $g_l$  and the wavelet (high-pass) filter is represented as  $h_l$ .

The total wavelet energy of the  $n^{\text{th}}$  node at level  $j$  is defined as

$$E_{j,n} = \sum_{t=0}^{N-1} \|W_{j,n,t}\|^2. \quad (9)$$

The probability of the  $t^{\text{th}}$  coefficient represents as

$$p_{j,n,k} = \frac{\|W_{j,n,t}\|^2}{E_{j,n}}. \quad (10)$$

That is,  $p_{j,n,k}$  are the normalized squares of the wavelet packet coefficients in the  $n^{\text{th}}$  node at level  $j$ . Finally, the Shannon entropy which is an measure of information entropy is given by

$$\text{SE}_{j,n} = - \sum_{t=0}^{N-1} p_{j,n,k} \times \log p_{j,n,t}. \quad (11)$$

In order to calculate the Shannon entropy, we perform a level 2 decomposition of the raw ENF signals using the MODWPT. By decomposing the MODWPT to level 2, there are four nodes at level 2:  $W_{2,0}$ ,  $W_{2,1}$ ,  $W_{2,2}$ , and  $W_{2,3}$ . The results of MODWPT for one ENF segment can be seen in Figure 4. After performing MODWPT up to level 2, Shannon entropy was calculated for each of the four terminal nodes ( $W_{2,0}$ ,  $W_{2,1}$ ,  $W_{2,2}$ , and  $W_{2,3}$ ).

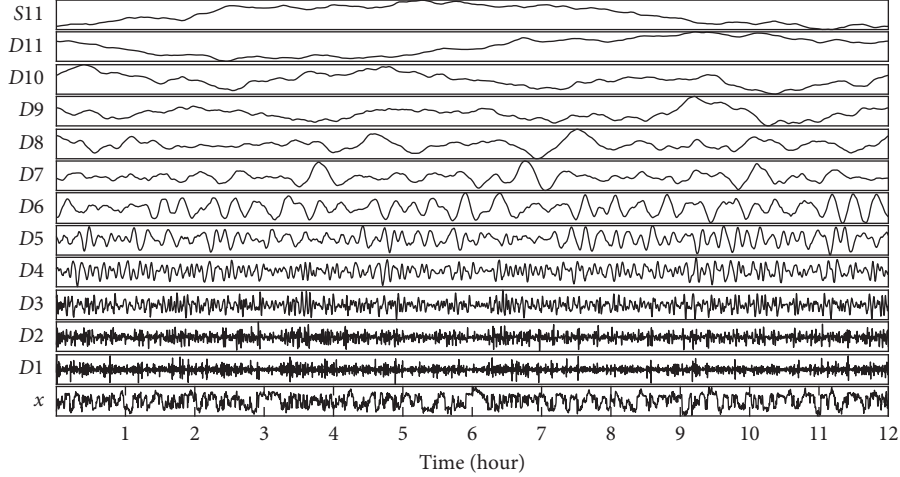


FIGURE 3: The maximal overlap discrete wavelet transform decomposition of the raw ENF signals during 12 hours ( $x$ ). The different frequency components of the raw ENF signals displayed as the orthogonal components ( $D1, D2, \dots, D11$ ). The smoothed components are plotted in  $S11$ .

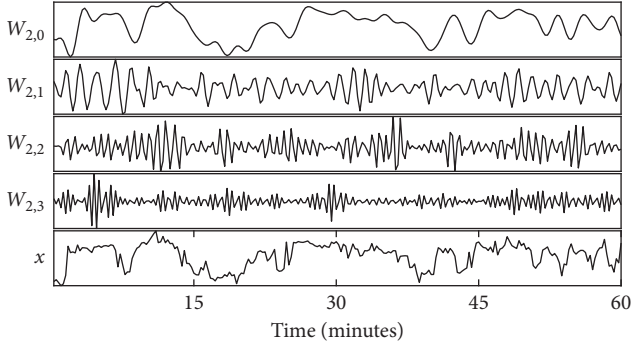


FIGURE 4: Raw ENF signal,  $x$ , and two-level maximal overlap discrete wavelet packet transform for the raw signals.

**5.1.4. Feature Vector Configuration through Integration of Feature Extraction Methods.** The ENF signal fluctuates randomly over time around 60 Hz, and it is difficult to predict the pattern [4]. Therefore, it is hard to obtain enough features to classify grid in ENF signals by using only one feature extraction technique. As a result, we combined (1) the log variance of signals, (2) the AR coefficients, (3) the Shannon entropy on the terminal nodes of maximal overlap, and (4) multiscale variance of maximal overlap as a feature vector to prevent bias caused by lost information and to extract sufficient characteristics of ENF signals. The details of the feature vector are shown in Table 3.

**5.2. Signal Classification on Extracted Features.** In order to find the power grid based on the features extracted from the section, we experimented two different types of classifiers. One is support vector machine-based approach, and the other is XGBoost based classifier.

**5.2.1. Support Vector Machines (SVMs).** In order to classify the power grids, we applied support vector machines (SVMs) to the feature extracted from Section 5.1. The SVM is a

TABLE 3: Feature vector components for ENF segment.

Index	Extracted features
1	$\log(\text{variance})$ of ENF segment
2 ~ 13	AR(12) model parameter
14 ~ 17	Shannon entropy on the MODWPT at level 2
18 ~ (18 + $\lfloor \log_2 N + 1 \rfloor$ )	$\log(\text{multiscale variance})$ of MODWT

$N$ , the number of data samples.

supervised learning model for classification, regression, and novelty detection. One of the SVM's notable characteristics is that the model parameters are determined by a convex optimization, so any local solution is also a global optimum. Given labeled training data, the SVM algorithm finds the maximal margin hyperplane which determines the classification of the new examples. Before training the classifier, we standardized the feature vectors which described in Table 3 and then put into as SVM input vectors. Basically, the SVM is a 2 class classifier. However, we need to classify 3 classes (Eastern, Western, and Texas power grids) more than 2 classes. The values of the regularization item in SVM is set to 1. Accordingly, we employed the one versus all multiclass SVM which considers all possible class pairs to learn. Furthermore, to effectively perform nonlinear classification, the polynomial kernel function with order  $q$  was applied as  $\text{kern}(x, y) = (1 + x^T y)^q$ , where  $x$  and  $y$  are the  $p$  dimensional vectors. The polynomial kernel of order 2 was chosen in our experiment.

**5.2.2. XGBoost.** We experimented with another classifier besides SVM to classify the ENF signals. We employed XGBoost [13] which is a scalable end-to-end tree boosting system. The XGBoost, which is an open-source library, is an optimized distributed gradient boosting library. Moreover, the XGBoost proposes a parallel tree learning. The XGBoost provides various parameters for training the model. In this paper, we set the maximum tree depth for base learners as 3,



the value of gamma and alpha as 0, the lambda as 1, and boosting learning rate as 0.1.

## 6. Implementation and Evaluation

In this section, we describe the results of our proposed approaches with ENF signals which are obtained from power line and multimedia. For both type of ENF signal datasets, we measured the accuracy and F1 score of the power grid estimation according to the number of data points. Moreover, we compared the performance of two types of classifiers: support vector machines and XGBoost.

**6.1. Software and Hardware Specifications.** All of our experiments were implemented using two PCs. One of two PCs is equipped with an Intel Core i7-7700 CPU (3.60 GHz) with 8 cores, 16 GB memory, and Ubuntu 16.04 LTS (64 bit). The other PC consists of an Intel(R) Core(TM) i5-3230M CPU (2.60 GHz) with 2 cores, 8 GB (1600 MHz DDR3) memory, 250 GB SATA hard drive, and macOS 10.14. We used Python to gather ENF data and to implement “XGBoost” [13]. Additionally, we applied MATLAB for data preprocessing, feature extraction, and SVM. The FFmpeg [27] was employed to extract audio data from videos.

**6.2. Evaluation of ENF Signals Extracted from Distribution Level.** In Section 5.1, we suggested a feature extraction algorithm for ENF signals. The purpose of this research is not just to extract the features but to classify them into the appropriate power grid: the Eastern, the Western, or the Texas Interconnection. In order to train the model and test the unlabeled data using the introduced classifiers, support vector machines and XGBoost, we randomly divided the datasets into test and train datasets to evaluate the proposed algorithms. When dividing the datasets into training and test datasets, we separated them equally by class so that they maintain the same ratio as the total dataset. We used ENF signal extracted from the distribution level which is described in Section 4.1 as dataset.

We trained 70% and tested 30% of our dataset. For the accurate validity of the results, we repeated these experiments for 20 rounds. In Table 4, the number of data points means the length of data points of the raw signal before extracting the feature. We examined the accuracy conducting on the ENF signal length. As can be seen in Table 4, the longer the signal is used as an input data, the more accurate it is in our proposed method. Furthermore, if the time of the raw ENF signals is longer than one hour, we can accurately predict the power grid more than 88.39%. In spite of accuracy, our proposed method with SVM classifier is little bit higher than the XGBoost classifier in all analyzed time lengths. Based on the F1 score in Table 5, SVM is not always better than XGBoost. However, the F1 score difference between SVM and XGBoost is almost similar to 0.01 or less (Table 5).

**6.3. Evaluation of ENF Signals Extracted from Online Multimedia.** In order to evaluate the location estimation in multimedia, the multimedia data which have the ENF signals

were selected to configure the dataset (Section 4.2). The total number of video files used in the dataset was 374. When dividing the train dataset and the test dataset, we took into account the ratio of each class in the entire dataset. In case of multimedia files, the ENF signals are obtained by using QIFFT and multitone harmonic methods. Afterwards, a feature vector was constructed using the proposed method in Section 5.1. In order to evaluate our suggested feature extraction algorithm and two different classifiers, SVM and XGBoost, we calculate the accuracy of location estimation based on the number of data points used in multimedia (Figure 5 and Algorithm 1).

When calculating the accuracy according to the experimented data points, the accuracy was measured up to 15 minutes, 30 minutes, and an hour unit, because there were less than three video files recorded over 2 hours (Table 2). Each row of Table 6 represents the accuracy and F1 score of  $S_j = 60$  for 374 multimedia files,  $S_j = 120$  for 199 multimedia files, and  $S_j = 240$  for 20 multimedia files. In Table 6, the performance of XGBoost classifier is better than SVM in the entire length of multimedia files. In Table 4, both classifiers display that the longer the length of the ENF signal analyzed from the data collected from the power line level, the more stable performance was obtained. However, when the power grid is estimated using SVM in the multimedia dataset, the accuracy and the F1 score of the multimedia files, which are recorded over 1 hour, were suddenly decreased to 68.82%. Since the XGBoost has increased accuracy and F1 score for the same data, XGBoost is more efficient than SVM in classifying ENF signals.

We analyzed the relationship between the number of harmonic frequencies used to extract ENF signals in multimedia and their accuracy. Moreover, the relationship between the accuracy and frequency band used for the restoration of the ENF signal is analyzed. When we obtained ENF signals from multimedia, we used the multitone harmonics method. Therefore, some multimedia files extracted ENF signals in both fundamental frequency and harmonic frequency bands, totally 8 harmonic bands are employed to extract ENF signals. However, the ENF signals could be extracted from the fundamental frequency and some harmonic frequency bands, only the fundamental frequency or only from the harmonic frequency bands. As shown in Table 1, we extract the ENF signals from multimedia files using 2 to 8 harmonics frequency bands.

The result of the accuracy using XGBoost is shown according to the number of harmonic frequency bands used for ENF signal extraction in Table 7. As can be seen in Table 7, the accuracy is not low and not high according to the number of harmonic bands used for ENF signal. Comprehensively considering Table 7 and Figure 2, the accuracy does not have a meaningful relationship depending on whether signal extraction of a specific frequency band is applied or not. Therefore, it is difficult to conclude that the number of frequency bands extracted from multimedia files or participation of a specific frequency band affects the accuracy of experiment.

The accuracy of experiment is more affected by whether data loss occurs when collecting multimedia data,

TABLE 4: The accuracy of power grid estimation on power line signal (trained 70%, averaged over 20 rounds).

Time (hour)	0.25	0.5	1	2	3	4	6	8	12	24
Number of data points	60	120	240	480	720	960	1440	1920	2880	5760
Proposed with SVM (%)	<b>67.14</b>	<b>76.74</b>	<b>88.39</b>	<b>95.21</b>	<b>96.47</b>	<b>98.07</b>	<b>99.01</b>	<b>99.07</b>	<b>99.44</b>	<b>99.78</b>
Proposed with XGBoost (%)	66.36	78	87.67	94.42	95.88	97.06	98.28	98.61	99.55	98.89

TABLE 5: The F1 Score of power grid estimation on power line signal (trained 70%, averaged over 20 rounds).

Time (hour)	0.25	0.5	1	2	3	4	6	8	12	24
Number of data points	60	120	240	480	720	960	1440	1920	2880	5760
Proposed with SVM	<b>0.66</b>	0.76	<b>0.88</b>	<b>0.95</b>	<b>0.96</b>	<b>0.98</b>	<b>0.99</b>	<b>0.99</b>	<b>0.99</b>	<b>0.99</b>
Proposed with XGBoost	<b>0.66</b>	<b>0.77</b>	0.87	0.94	0.95	0.97	0.98	<b>0.99</b>	<b>0.99</b>	0.98

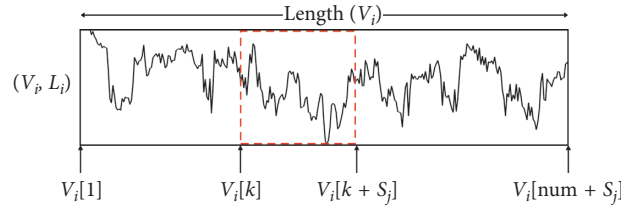


FIGURE 5: Notation for ENF signals extracted from multimedia where  $\text{length}(V_i) \geq S_j$ . The ENF signals obtained from one multimedia is represented by  $V_i$ . The label of the area where multimedia is recorded is indicated by  $L_i$ .  $S_j$  ( $j = 1, 2, \dots, J$ ) is the number of data samples to estimate the location. The *num* is equal to  $\text{length}(V_i) - S_j + 1$ .

**INPUT:**

$V = \{V_i\}$  consists of ENF signals from videos ( $i = 1, 2, \dots, I$ )  
 $L = \{L_i\}$  is the real label of videos ( $i = 1, 2, \dots, I$ )  
 $S = \{S_j\}$  is the number of data points to analyze ( $j = 1, 2, \dots, J$ )

**OUTPUT:**

$A = \{A_j\}$  is the location estimation accuracy according to analyzed data points ( $j = 1, 2, \dots, J$ )

```

(1) for all  $S_j \in S$  do
(2)    $\text{cnt\_num} \leftarrow 0$ 
(3)    $\text{cnt\_correct} \leftarrow 0$ 
(4)    $\text{cnt\_video} \leftarrow 0$ 
(5)   for all  $V_i \in V$  do
(6)     if  $\text{length}(V_i) < S_j$  then
(7)       break
(8)     end if
(9)      $\text{num} \leftarrow \text{length}(V_i) - S_j + 1$ 
(10)     $\text{cnt\_num} \leftarrow \text{cnt\_num} + \text{num}$ 
(11)     $\text{cnt\_video} \leftarrow \text{cnt\_video} + 1$ 
(12)    for  $k = 1$  to  $\text{num}$  do
(13)       $\text{predLabel} \leftarrow \text{predict}(v_i[k : k + S_j])$ 
(14)      if  $\text{predLabel} == L_i$  then
(15)         $\text{cnt\_correct} \leftarrow \text{cnt\_correct} + 1$ 
(16)      end if
(17)    end for
(18)  end for
(19)   $A_j \leftarrow 100 \times (\text{cnt\_correct} / \text{cnt\_num})$ 
(20) end for

```

ALGORITHM 1: Calculate accuracy from multimedia.

rather than how many frequency bands can be extracted from the multimedia data. Most of the multimedia data we collected are live webcam service's data. Therefore, those video files were uploaded to the server in real time and we downloaded the files. Considering these characteristics of

the video files, data loss could have occurred due to delays caused by the network or external environment. If data loss occurs when downloading a file, our proposed method which analyzes the time and frequency domains is hard to predict the correct result. Consequently, in order



TABLE 6: The accuracy and F1 score of power grid estimation on multimedia data (trained 70%, averaged over 20 rounds).

Time (hour)	Number of data points ( $S_j$ )	Number of multimedia	Accuracy		F1 score	
			Proposed with SVM (%)	Proposed with XGBoost (%)	Proposed with SVM	Proposed with XGBoost
0.25 ~	60	374	83.23	<b>85</b>	0.75	<b>0.79</b>
0.5 ~	120	199	82.92	<b>94.17</b>	0.75	<b>0.93</b>
1 ~	240	20	68.82	<b>99.92</b>	0.56	<b>0.99</b>

TABLE 7: The accuracy according to the number of harmonic bands extracted from multimedia files with XGBoost (trained 70%, averaged over 20 rounds).

Time (hour)	Accuracy by the number of harmonic bands (%)								Average
	1	2	3	4	5	6	7	8	
0.25 ~	—	<b>99.07</b>	<b>98.91</b>	<b>99.25</b>	49.25	59.86	61.84	<b>88.21</b>	85
0.5 ~	—	<b>99.74</b>	<b>99.95</b>	<b>100</b>	73.85	85.63	87.79	<b>95.22</b>	94.17
1 ~	—	<b>99.88</b>	—	—	<b>100</b>	<b>100</b>	99.68	<b>99.97</b>	99.92

to estimate a location using multimedia files, the data loss should be small and the file should not be modified.

## 7. Discussions

In this paper, we have studied about estimating the location using feature extraction and classification methods from ENF signals. The ENF signals used in this paper are extracted from the power system distribution level and online multimedia. In this section, we describe our research can be applied to discussion of location privacy issues and digital forensics. Furthermore, we discuss the ways to reduce the influence of ENF signals when recording videos or audios. We also explain about the limitations and future work of our research.

**7.1. Location Privacy.** The location information is one of the most important personal information. Therefore, most countries recommend that the web service provider should have user's consent to gather user's location information. We showed that we could find out the power grid of recorded video files which are uploaded in three worldwide live webcam services, such as Skyline Webcams [9], Earthcam [8], and explore [10] and one global video sharing web service YouTube [7]. If the video files or audio files are taken in a good environment such as using low noise environment and high quality equipment, the location can be found with a higher probability. Even if the user does not agree with the location information when uploading video files or audio files, the location of the recorded files can be estimated by our proposed method. This is a matter of caution as video sharing services and personal broadcasts using the Internet are increasing worldwide.

**7.2. Digital Forensics.** Since the ENF signals have unique properties in location and time, it can be used for digital forensics. In particular, the applicability of ENF signals has been continuously raised in determining the authenticity of digital audio recordings (Cooper [28], Grigoros [3], and

Grigoros [4]). However, in those studies, a reference ENF database which consists of ENF signals extracted in the same area and the same time as the digital audio recording was required to determine the authenticity. Since the authenticity of a file is generally progressed after recording, it is difficult to determine the modification of the file if there are no reference data. This means that, in order to actually apply these researches to digital forensics, reference data must always be constructed. This is not a simple matter in terms of time and economy. However, even if there are no ENF signals extracted at the same time of recording an audio file, we can determine the modification of multimedia files by applying our algorithm. By using the feature extraction techniques and the multiclassification technique proposed in this paper, the ENF signals can be more efficiently applied to the digital forensics.

**7.3. Countermeasure.** In terms of information security, the fact that the ENF signal is stored when recording video files or audio files can cause a problem of privacy leakage of users. The quality of the ENF signals stored in the files is highly affected by the surroundings and equipment at the time of recording. Hajj-Ahmad et al. [29], Fechner and Kirchner [30], Brixen [31], and Chai et al. [32] studied about the factors that capture the ENF signals in audio files or video files. They could capture the ENF signals when recording with dynamic microphones, but not with electret microphones because of the electromagnetic fields. In addition, they stated that the factors that capture ENF signals are related to the internal compression of the recorder. If the recorder internal compression is strong, it can reduce the ENF signals in videos or audios. Furthermore, if the recorded video files or audio files are modified such as deleting or inserting fake signals, it will be hard to obtain the information.

**7.4. Limitations and Future Work.** We collected data and experimented only on power grids in the United States. Since there are only three power grids in the United States, it is

necessary to carry out our experiments on more power grids in order to confirm our proposed methods. In addition, we need to supplement the technique of classifying the areas in the same power grid because wide areas are made up of a single power grid. Although the ENF signals have originality in terms of time and location, many related studies about classifying ENF signals are limited to classifying locations, including our studies. Therefore, we will also study on techniques for extracting features and classifying ENF signals over time. If the features of the ENF signals can be extracted by time, it is possible to infer both the time and region where an arbitrary signal is extracted. This will further enhance the value of the ENF signals as digital forensics.

In future work, we will collect more video files and audio files from the Internet and examine our proposed algorithms. In this paper, we examined 374 video files collected from web services. In order to analyze the accuracy according to the harmonic frequency bands, the number of videos was insufficient and it was difficult to generalize. Therefore, we will gather more video files and audio files from the Internet and then improve our algorithms. Furthermore, the research about divide and conquer technique can help to detect the modification of the multimedia file and improve our algorithm. The divide and conquer method is an approach to tough problems divide into several problems. It is widely used to solve the classification or forecasting problems [33–35].

## 8. Conclusion

In this paper, we propose an algorithm to identify the power grid using the ENF signals extracted from the power system's distribution level and online multimedia such as video files and audio recordings. In order to find out the power grid from ENF signals, we extracted the feature from the ENF signal using three main feature extraction algorithms and found out the power grid using the classifier. We used two types of datasets to evaluate the validity of our proposed algorithm. The first dataset is ENF signals extracted using the FDR equipment at the distribution level of the power system. Another dataset is the ENF signal from the video files and audio files uploaded on the Internet. For each dataset, the accuracy of the proposed algorithm was measured according to various time units such as 15 minutes, 30 minutes, and an hour. In case of the ENF signals extracted from the power line and multimedia files, the more the number of analyzed samples, the higher the accuracy of estimation results when employing XGBoost classifier. Consequently, this paper is notable in that it can estimate the power grid from video files and audio files uploaded on the Internet as well as the ENF signals collected from the power line.

## Data Availability

All data used in the paper come from cited references or are reported in the paper

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This work was supported by the Institute for Information & Communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (no. R7117-16-0161; Anomaly Detection Framework for Autonomous Vehicles).

## References

- [1] B. Liscouski and W. Elliot, "Final report on the August 14, 2003 blackout in the United States and Canada: causes and recommendations," vol. 40, US-Canada Power System Outage Task Force, Washington, DC, USA, 2004, Tech. Rep. 4, <https://emp.lbl.gov/publications/final-report-august-14-2003-blackout>.
- [2] A. Hajj-Ahmad, R. Garg, and M. Wu, "ENF based location classification of sensor recordings," in *Proceedings of the 2013 IEEE International Workshop on Information Forensics and Security (WIFS)*, pp. 138–143, IEEE, Guangzhou, China, November 2013.
- [3] C. Grigoros, "Digital audio recording analysis: the electric network frequency (ENF) criterion," *International Journal of Speech, Language and the Law*, vol. 12, no. 1, pp. 63–76, 2005.
- [4] C. Grigoros, "Applications of ENF criterion in forensic audio, video, computer and telecommunication analysis," *Forensic Science International*, vol. 167, no. 2–3, pp. 136–145, 2007.
- [5] Facebook, "Facebook," March 2019, <https://www.facebook.com>.
- [6] Instagram, "Instagram," March 2019, <https://www.instagram.com>.
- [7] YouTube, "YouTube," March 2019, <https://www.youtube.com>.
- [8] EarthCam, "EarthCam," March 2019, <https://www.earthcam.com>.
- [9] Skyline Webcams, "Skyline Webcams," March 2019, <https://www.skylinewebcams.com>.
- [10] Explore, "Explore," March 2019, <https://explore.org>.
- [11] Y. Jeon, M. Kim, H. Kim, H. Kim, J. H. Huh, and J. W. Yoon, "I'm listening to your location! inferring user location with acoustic side channels," in *Proceedings of the 2018 World Wide Web Conference on World Wide Web*, pp. 339–348, International World Wide Web Conferences Steering Committee, Lyon, France, April 2018.
- [12] H. Kim, Y. Jeon, and J. W. Yoon, "Construction of a national scale ENF map using online multimedia data," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pp. 19–28, ACM, Singapore, November 2017.
- [13] T. Chen and C. Guestrin, "XGBoost: a scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, ACM, San Francisco, CA, USA, August 2016.
- [14] A. Hajj-Ahmad, R. Garg, and M. Wu, "ENF-based region-of-recording identification for media signals," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 6, pp. 1125–1136, 2015.

- [15] W. Yao, J. Zhao, M. J. Till et al., "Source location identification of distribution-level electric network frequency signals at multiple geographic scales," *IEEE Access*, vol. 5, pp. 11166–11175, 2017.
- [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 *Proceedings of the 2016 IEEE 14th International Symposium on Intelligent Systems and Informatics (SISY)*, pp. 51–56, IEEE, Subotica, Serbia, August 2016.
- [17] A. Hajj-Ahmad, R. Garg, and M. Wu, "Instantaneous frequency estimation and localization for ENF signals," in *Proceedings of the 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference*, pp. 1–10, IEEE, Hollywood, CA, USA, December 2012.
- [18] M. Abe and J. O. Smith III, "Design criteria for simple sinusoidal parameter estimation based on quadratic interpolation of FFT magnitude peaks," in *Audio Engineering Society Convention 117*, Audio Engineering Society, New York, NY, USA, 2004.
- [19] G. Hua, G. Bi, and V. L. L. Thing, "On practical issues of electric network frequency based audio forensics," *IEEE Access*, vol. 5, pp. 20640–20651, 2017.
- [20] Y. Liu, S. You, W. Yao et al., "A distribution level wide area monitoring system for the electric power grid-FNET/Grid-Eye," *IEEE Access*, vol. 5, pp. 2329–2338, 2017.
- [21] Y. Zhang, P. Markham, T. Xia et al., "Wide-area frequency monitoring network (FNET) architecture and applications," *IEEE Transactions on Smart Grid*, vol. 1, no. 2, pp. 159–167, 2010.
- [22] Power Information Technology Laboratory at the University of Tennessee, "FNET/grideye web display," 2018, <http://fnetpublic.utk.edu/index.html>.
- [23] E. A. Maharaj and A. M. Alonso, "Discriminant analysis of multivariate time series: application to diagnosis based on ECG signals," *Computational Statistics & Data Analysis*, vol. 70, pp. 67–87, 2014.
- [24] D. B. Percival and H. O. Mofjeld, "Analysis of subtidal coastal sea level fluctuations using wavelets," *Journal of the American Statistical Association*, vol. 92, no. 439, pp. 868–880, 1997.
- [25] T. Li and M. Zhou, "ECG classification using wavelet packet entropy and random forests," *Entropy*, vol. 18, no. 8, p. 285, 2016.
- [26] A. T. Walden and A. C. Crisan, "The phase-corrected undecimated discrete wavelet packet transform and its application to interpreting the timing of events," *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1976, pp. 2243–2266, 1998.
- [27] FFmpeg Developers, "FFmpeg," March 2019, <https://ffmpeg.org>.
- [28] A. J. Cooper, "The electric network frequency (ENF) as an aid to authenticating forensic digital audio recordings—an automated approach," in *Proceedings of the 33rd International Conference: Audio Forensics-Theory and Practice Audio Engineering Society Conference*, Audio Engineering Society, Denver, CO, USA, June 2008.
- [29] A. Hajj-Ahmad, C.-W. Wong, S. Gambino, Q. Zhu, M. Yu, and M. Wu, "Factors affecting ENF capture in audio," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 2, pp. 277–288, 2019.
- [30] N. Fechner and M. Kirchner, "The humming hum: background noise as a carrier of ENF artifacts in mobile device audio recordings," in *Proceedings of the 2014 Eighth International Conference on IT Security Incident Management & IT Forensics*, pp. 3–13, IEEE, Münster, Germany, May 2014.
- [31] E. B. Brixen, "Techniques for the authentication of digital audio recordings," in *Audio Engineering Society Convention 122*, Audio Engineering Society, New York, NY, USA, 2007.
- [32] J. Chai, F. Liu, Z. Yuan, R. W. Conners, and Y. Liu, "Source of ENF in battery-powered digital recordings," in *Audio Engineering Society Convention 135*, Audio Engineering Society, New York, NY, USA, 2013.
- [33] Y. Zhou, T. Li, J. Shi, and Z. Qian, "A CEEMDAN and XGBOOST-based approach to forecast crude oil prices," *Complexity*, vol. 2019, Article ID 4392785, 15 pages, 2019.
- [34] Y.-X. Wu, Q.-B. Wu, and J.-Q. Zhu, "Improved EEMD-based crude oil price forecasting using LSTM networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 516, pp. 114–124, 2019.
- [35] J. Kevric and A. Subasi, "Comparison of signal decomposition methods in classification of EEG signals for motor-imagery BCI system," *Biomedical Signal Processing and Control*, vol. 31, pp. 398–406, 2017.



