Construction of a National Scale ENF Map using Online Multimedia Data

Hyunsoo Kim, Youngbae Jeon, and Ji Won Yoon
Center for Information Security Technologies (CIST)
Korea University
Republic of Korea
{aitch25,jyb9443,jiwon_yoon}@korea.ac.kr

ABSTRACT

The frequency of power distribution networks in a power grid is called electrical network frequency (ENF). Because it provides the spatio-temporal changes of the power grid in a particular location, ENF is used in many application domains including the prediction of grid instability and blackouts, detection of system breakup, and even digital forensics. In order to build high performing applications and systems, it is necessary to capture a large-scale nationwide or worldwide ENF map. Consequently, many studies have been conducted on the distribution of specialized physical devices that capture the ENF signals. However, this approach is not practical because it requires significant effort from design to setup, moreover, it has a limitation in its efficiency to monitor and stably retain the collection equipment distributed throughout the world. Furthermore, this approach requires a significant budget.

In this paper, we proposed a novel approach to constructing the worldwide ENF map by analyzing streaming data obtained by online multimedia services, such as “Youtube”, “Earthcam”, and “Ustream” instead of expensive specialized hardware. However, extracting accurate ENF from the streaming data is not a straightforward process because multimedia has its own noise and uncertainty. By applying several signal processing techniques, we can reduce noise and uncertainty, and improve the quality of the restored ENF.

For the evaluation of this process, we compared the performance between the ENF signals restored by our proposed approach and collected by the frequency disturbance recorder (FDR) from FNET/GridEye. The experimental results show that our proposed approach outperforms in stable acquisition and management of the ENF signals compared to the conventional approach.

CCS CONCEPTS

•**Information systems →Data extraction and integration;** *Data mining;* Multimedia streaming;

KEYWORDS

Electrical Network Frequency; Multimedia data; Frequency domain; Power Grid

CIKM'17, November 6–10, 2017, Singapore.

ISBN 978-1-4503-4918-5/17/11…$15.00
DOI: https://doi.org/10.1145/3132847.3132982

ACM Reference format:
Hyunsoo Kim, Youngbae Jeon, and Ji Won Yoon. 2017. Construction of a National Scale ENF Map using Online Multimedia Data. In *Proceedings of CIKM'17, November 6–10, 2017, Singapore.*, , 10 pages.
DOI: https://doi.org/10.1145/3132847.3132982

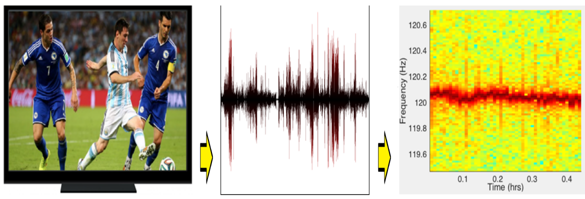
1 INTRODUCTION

The supply frequency of electrical power in distribution networks is called electrical network frequency (ENF). The signal is captured in a particular frequency, either $50Hz$ or $60Hz$, depending on the geographic location. This ENF has temporal fluctuation as the load in the power grid can vary in a time series. Therefore, ENF signals are often used as a spatio-temporal pattern for identifying both location and time.

This electrical pattern would be an important factor in the analysis of power system. ENF is oscillated by the influence of societal events, disaster, and electrical disturbances. With regard to this phenomenon, [5] deals with impacts of societal events on the power system. They explained how the electrical signal oscillates when large-scale events such as Super bowl and Worldcup are progressed. Moreover, they show the possibility of event detection using ENF signals. On the other hands, [16, 19] focused on the electrical accidents on the power system such as line trip or islanding. They provided the method which detects these accidents properly and showed that it is feasibly made using frequency based or angle-based detection techniques. Furthermore, [10, 19] extended range of the events to disasters. In [10, 19], authors handled the feature of signal oscillation not only when line trip accidents occur, but also when and where blackouts outbreak.

An interesting fact is that when digital equipments including computers, digital recorders, video cameras record something, they capture not only intended sound but also 50/60 Hz ENF signals[6, 9] owing to the influences of electrical power through the power plugs. Figure 1 describes the general procedure of ENF extraction from multimedia data through short-time Fourier transform (STFT). Since the STFT is useful to capture the change of the pattern in frequency domain over time, it is possible to extract the ENF signal embedded in the audio/video file.

Following the discovery of ENF patterns in video and audio files, the use of ENF signals emerged as an important and useful tool for many application domains including the analysis of social event impact[5], regional electrical event detection[10], and digital forensics[9]. As ENF signals can be used to identify the location and time, it has recently been exploited for forensic science as well. By simply checking whether the ENF values are natural or not, it is possible to detect the multimedia forgery[8]. Furthermore, by evaluating the correlation between ground-truth data and target value, it is also possible to figure out when and where the signals were extracted[12].



(a) Multimedia data (b) Extracted audio signals (c) ENF patterns in spectrogram

Figure 1: ENF extraction from multimedia data including film and audio files: ENF patterns can be found in the spectrogram obtained from the internal audio signals of multimedia data.

The aforementioned application domains in power systems and digital forensics require a large-scale ENF map that displays and stores all ENF signals in a nation-wide or worldwide range. To construct such a large scale spatio-temporal ENF map, many studies have been conducted on the distribution of the specialized physical devices that capture the ENF signals. The ENF signals are collected from several countries using specialized equipment such as frequency disturbance recorders (FDR)[15]. However, their approach had the following problems. First, physical devices and sensors that were needed to collect the ENF signals should have been distributed in as many places as possible in order to obtain the desired map with higher resolution. Additionally, continued management to monitor and detect abnormal behavior or malfunction of the physical devices was required. Moreover, both installation and management would be expensive and require a considerable budget.

These problems hinder us from accurately collecting and managing ENF signals using physical sensors worldwide. In order to efficiently obtain and process the large-scale ENF signals, we propose a novel approach that composes a worldwide ENF map through collecting and processing multimedia big data on the worldwide web. First, we collected streaming multimedia data from online multimedia services, "Ustream", "Earthcam", etc. Next, we extracted ENF signals from the collected multimedia data.

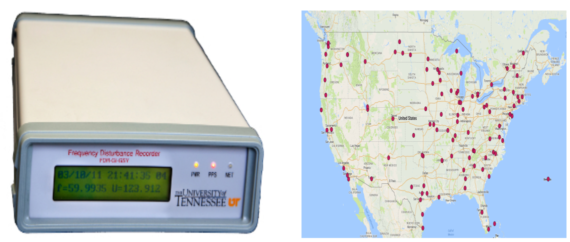
From this perspective, **our key contribution is that a large-scale ENF map can be efficiently and effectively constructed by analyzing online multimedia in time series.** Because our proposed approach collects ENF signals from online multimedia on the worldwide web, it contains the following surprising properties:

- *Purely online/no need of physical devices*: We introduce a novel method which collects worldwide ENF signals through online multimedia without any additional physical equipment.
- *Simple management system*: Because multimedia big data on the worldwide web are processed in a generalized framework, additional management systems that monitor the status of sensors are not needed. For example, we do not need to worry about the failure or malfunction of sensors or devices.
- *Low budget*: Because the proposed system obtains ENF signals from public online multimedia, it is free of charge.

However, our proposed approach requires additional bundles in analyzing multimedia data compared to using physical devices. These additional tasks include three sophisticated signal processing procedures: (1) reducing noise for signal enhancement; (2) signal alignment for extracting temporally longer and stable patterns; and (3) spatial interpolation for unknown signals in the regions that do not have any multimedia data on the worldwide web. Therefore, we focus on three main questions in this paper.

1. **How and where can we efficiently collect the ENF signals?**
2. **How can we extract stable and longer ENF signals?**
3. **How can we construct the national scale ENF map?**

Given these three main issues, the structure of the paper is as follows. First, we introduce our basic idea of using online multimedia instead of specialized equipment in Section 2. In Section 3, we explain signal processing techniques that were applied to extract clean ENF signals. Next, we conduct experiments to prove the proposed approach is appropriate and available for realistic use in Section 4. In Section 5, we discuss our experimental results and their significance. Finally, we conclude the paper in Section 6.



(a) FDR (b) Locations of FDRs at USA

Figure 2: FDR and the location of the sensors of FNET/GridEye

2 PROBLEMS AND BASIC IDEA

The frequency of the power system could be used as important factors to understand and control the dynamics of the power system. In 2001, the author of [13, 15], therefore, attempted to collect world ENF patterns in a direct way using specialized devices. They collected world-area electric system frequency via a wide-area frequency monitoring network (FNET) and frequency disturbance recorder (FDR).[1] That is, they gathered local electrical patterns from the world with GPS synchronization and transmitting them to their server through the Internet. This research is continued to dealing with an actual implementation and explaining Universal Grid Analyzer (UGA) for promoting a better understanding of the power grid dynamics [17, 18]. Figure 2 shows a FDR and about 80 locations across the continent and around United States.

It is true that the best method to collect clean ENF signals is to install as many physical devices and sensors as possible. However, the approach using physical sensors distributed in wide area may have the following problems. First of them, physical devices and

[1] FDR is a type of phasor measurement unit in smart grid technology.

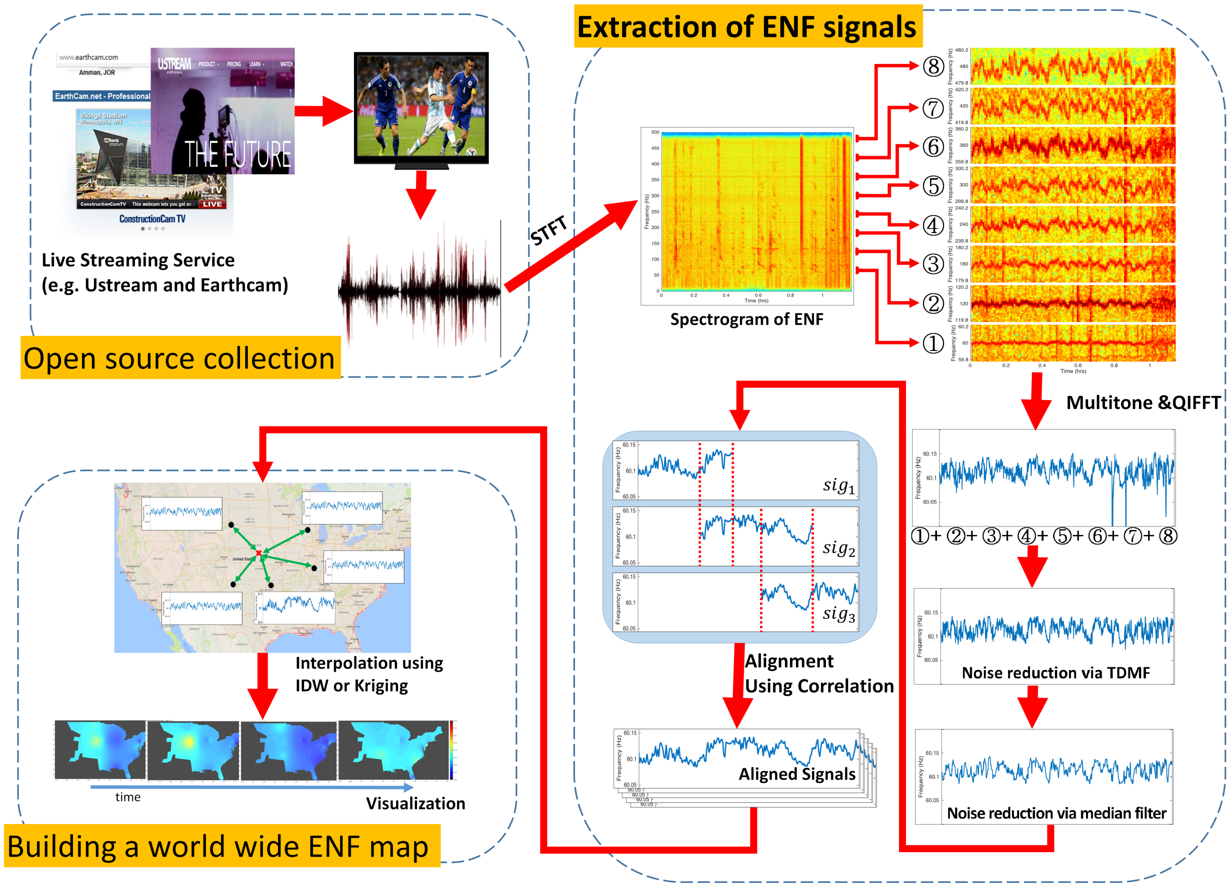


Figure 3: A whole procedure from collecting multimedia data to building a large-scale ENF map

sensors should be developed to collect the ENF signals and they should be distributed as many places as possible in order to obtain more high resolution ENF signals. It also requires continuing management which monitors and detects abnormal behaviors or disorder of the physical devices, including sensor failures. We can easily find that many FDR sensors did not work in FNET/GridEye [2]. For instance, the FDR sensor, planted in eastern Japan, did not work between 05 : 19pm and 10 : 15pm, Feb 16, 2017 (UTC). More than half of the signals were collected as a state of "No Data", which means missing signals or malfunction.

In addition, it is expensive since both installation and management require a lot of budgets. For example, the price of the device, FDR, is about $2,500, which is not cheap. In this paper, we have introduced a new alternative approach to collecting ENF signals from not physical sensors but public multimedia on world wide web. If used data is collected from world wide web, it is free of charge and it takes low cost in managing systems and data.

3 PROPOSED APPROACHES

As discussed in the introduction, we propose an efficient alternative approach to collecting and extracting ENF signals from online multimedia sources. Subsequently, we build a world ENF map using the extracted signals. The whole process from collecting multimedia data to building ENF map is illustrated in Figure 3. As can be seen in the figure flowchart above, our proposed approach is primarily partitioned into three consecutive steps: (1) open source collection, (2) extraction of ENF signals, and (3) building a national scale ENF map. Detailed steps are described through the subsections below.

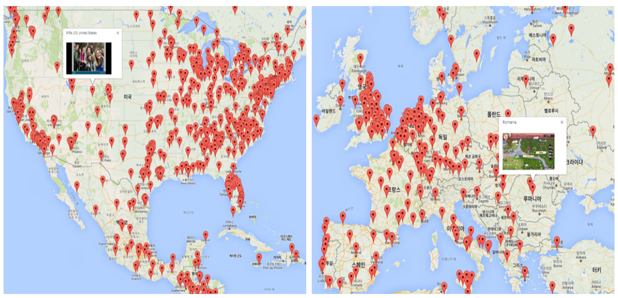
3.1 Collection of multimedia data using open sources

Because we extract ENF signals from multimedia data, we first determined which sources contained good quality ENF signals. Throughout the Internet, there are several candidates from which we can extract ENF signals. For example, social network services, including "Facebook" and "Instagram", contain film and audio. "YouTube" is another significant source of audio and video data. However, we sought open sources that would provide cleaner, more reliable ENF signals in the spectrogram. Moreover, in order to construct a national scale ENF map, sources or services should be used from throughout the world and include geo-information. Using these parameters, we concluded that multimedia data could be obtained from live streaming services which cover large portions of the world and are captured for a long period of time. In this paper, we selected four well-known live streaming services on the Internet:

[2]http://fnetpublic.utk.edu/tabledisplay.html

"Ustream", "Earthcam", "Skyline webcams", and "Explore (providing service through Youtube)". "Facebook" and "Instagram" also have provided good quality multimedia. However, because of the insufficient runtime of the multimedia, we excluded these services.

Ustream is a video streaming service that provides various multimedia to more than 80 million users. Because a portion of these are broadcast over the web, many personal broadcasts are simultaneously provided for viewers. Figure 4 shows the number of people who broadcast through Ustream simultaneously. Figure 4(a) and Figure 4(b) represent the distribution of broadcasters in U.S. and Europe respectively.



(a) U.S. (b) Europe

Figure 4: Distribution of Ustream broadcasters in U.S. and Europe

From the other sources to collect the ENF signal are online webcam networks, "Earthcam", "Skyline webcams", and "Explore", we could reliably extract the stable ENF signals because they provides high-quality sounds. Moreover, the online webcam networks can cover sufficiently wide areas as various videos are recorded with approximated 700 or more cameras throughout the world.

In our research, the multimedia collection is conducted through the following steps: 1) reverse engineering and crawling each web streaming services, 2) identifying live stream URLs by parsing the crawled data, and 3) collecting all live streaming multimedia with the FFmpeg[3] and identified URLs. In this way, we were able to collect the various multimedia from the world effectively. However, there is a point to consider. While the multimedia are being crawled via streaming services, there could be the omission of streaming packets due to the vagary of the streaming condition. For example, streaming downloaders are designed to stack up streaming packets sequentially in time series. However, when a connection failure occurs and a packet is not downloaded, it skips the missing packet, and next packets (following the missing packet) are shifted forward to fill the gap. This phenomenon causes distortion of time series of multimedia, and it is unfortunately impossible to extract proper signal with distorted data. therefore, this mechanism should be considered in the course of collection. To avoid this issue, we recommend designing streaming downloader to not drop missing parts of whole streaming data and keep the vacancy as it is.

3.2 Extraction of more stable and longer ENF signals from multimedia data

3.2.1 How to extract ENF signals. Because multimedia data includes media content along with the ENF signals, the extracted ENF signals can, however, be corrupted by the unwanted content as shown in Figure 1(c). That is, ENF signals might be slightly different from our expectations in reality. Therefore, several additional algorithms should be used to extract ENF signals with high quality by reducing and removing the noise. In our proposed approach, three algorithms for noise reduction are added to improve the quality of ENF signals.

The key algorithm for reducing uncertain noise in multimedia data is to use the multi-tone harmonics method. The multi-tone harmonics method uses a fundamental frequency along with harmonic frequencies for finding the peak position of the ENF. This concept was first introduced by [14], and [4] showed how it could be applied to the extraction of ENF signals. Given a multimedia sound signal $f(t)$ in a time domain, the signal can be transformed to $F(\omega) = \sum_{n=0}^{N-1} f(t)e^{-j\omega t}$ in the frequency domain. While ENF patterns appear around the $50/60Hz$ band of a fundamental frequency, similar ENF patterns are also present in harmonic frequency bands that are multiples of $50/60Hz$. Therefore, to reduce the uncorrelated noise and enhance the ENF patterns in the frequency domain, we created multi-tone spectra by summing all spectrogram at both fundamental and harmonic frequencies. Additionally, we also used the selective harmonics approach to extract more clear signals from multimedia. Contrary to the original method, this method composes harmonics bands selectively. It is useful to extract ENF signals from the multimedia which are polluted by noises. Afterwards, our proposed approach searches for interpolated peaks on the composed spectra and links them using a quadratic interpolated fast Fourier transform (QIFFT) [1, 6] because STFT requires too heavy computation to extract signals quickly from hundreds of multimedia. The sequence of the connected peaks are the ENF patterns in which we are interested, as shown in Figure 5(b).

Figure 5 represents the effectiveness of the multi-tone model and QIFFT. In the left figure, there are eight different spectra obtained at one fundamental frequency, and seven harmonic frequencies. The right side shows the effectiveness of the multi-tone approach. The bottom figure displays ENF signals from the single-tone and the top figure is an ENF signal generated from the composition of the harmonics using a multi-tone method. As can be seen, we can obtain enhanced ENF signals by adding more harmonic spectra.

After applying the multi-tone method and QIFFT, we could still improve the quality of the ENF signal using an additional signal processing filter. The filter that we used for delicately reducing noise near the ENF frequency band is threshold-dependent median filter (TDMF)[6]. This method combines the concept of thresholding approach with the conventional median filter. This performs filtering only for regions exceeding the threshold value. The performance is briefly shown in extraction part of Figure 3. We demonstrated how to obtain the expected ENF signals from online live streaming services.

3.2.2 Building stable and longer ENF signals with signal alignments. It is obvious to obtain longer ENF signals because they can

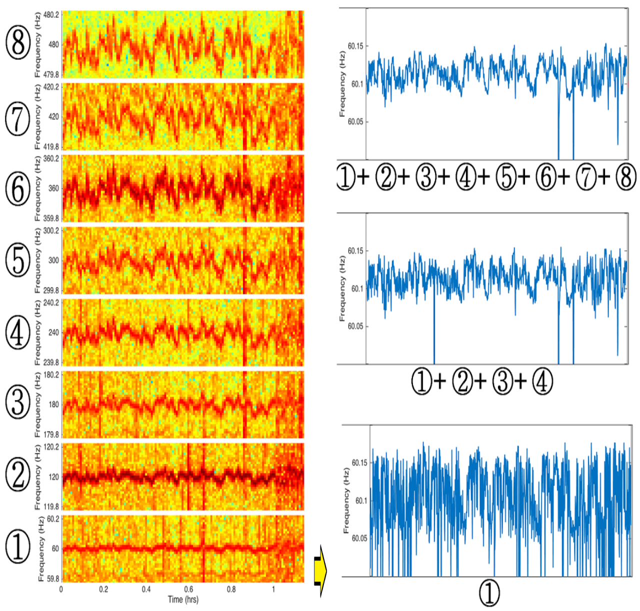


Figure 5: Enhanced ENF signals using the multi-tone method and QIFFT: Given eight different spectra obtained at one fundamental frequency and seven harmonics frequencies, we obtained enhanced ENF signals using the multi-tone method and QIFFT. In this paper, we used accurate ENF signals as shown in the right top figure.

provide more accurate and stable analysis [5, 8–10, 16, 19]. However, it is not easy to directly extract a long ENF signal from a multimedia service. Especially when extracting a signal from a personal broadcasting service such as Ustream, it is impossible to extract a long signal in one region only with a single person's data because the extraction position of the signal varies depending on whether the broadcasting service is on or not. In other words, if the broadcaster offers content within the needed time-frame, we can download the appropriate multimedia data for the ENF signal; however, if not, then we either cannot obtain the full signal or can obtain only incomplete signals. Therefore, in order to collect stable and longer ENF signals, we introduced a signal alignment technique that aligns multiple ENF signals from different multimedia broadcasts. That is, we collect many multimedia data streams in adjacent places and subsequent times with an overlap of the time range. With these gathered signals, we can construct a semi-complete ENF signal. We selected normalized cross correlation (NCC) metrics for aligning the multiple partial ENF signals. The similarity (ρ) can be found using the following equation, $\rho = \frac{\sum_L^{n=1}(x[n]-\bar{x})(y[n]-\bar{y})}{(L-1)\sigma_x\sigma_y}$. In this equation, x and y represent two sequential signals. For example, let us assume that we need a long ENF signal recorded in Pennsylvania, U.S. during 15:00 ~ 17:00 on March 12th, 2017. For this situation, we download a 30 minute long signal every 20 minutes. That is, the time overlap of each sequential multimedia is 10 minutes. We download the multimedia data from randomly selected locations near Pennsylvania. All the multimedia data will be completely different from each other, although they are acquired at close locations. We now express it with symbols. First, let v denote an index of multimedia data downloaded, and starting and ending time points of the download are expressed as (starting time point, ending time point). Then, we have p_v = (starting time point, ending time point). Afterward, the multimedia downloaded from 15:00 to 17:00, are expressed as $\{p_1 = (15:00, 15:30), p_2 = (15:20, 15:50), p_3 = (15:40, 16:10), p_4 = (16:00, 16:30), p_5 = (16:20, 16:50), p_6 = (16:40, 17:10)\}$. We calculate NCC with the overlapping parts of each signal as priority. Obviously, in the time domain analysis, it is impractical to detect alignment points for each multimedia source because they are downloaded from different sources and locations. However, in the view of ENF, the alignment points would be easily detected because we gathered it all within adjacent locations. This is described by Figure 6.

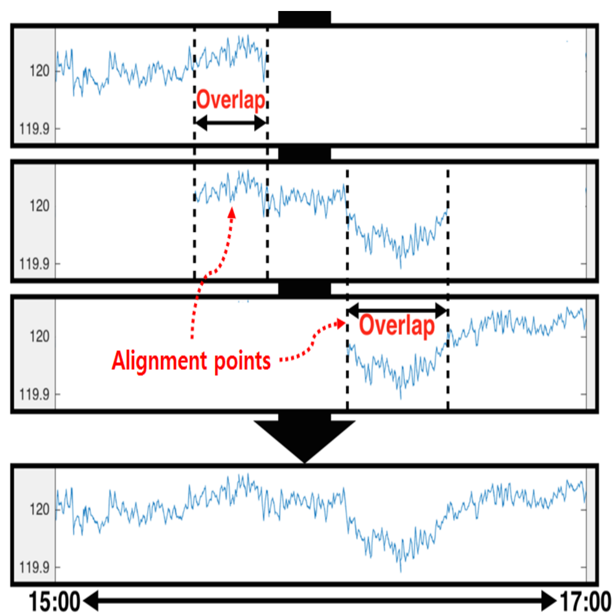


Figure 6: Signal alignment: partial and short ENF signals are merged to a longer and stable ENF using the signal alignment method.

3.3 Construction of a national scale ENF map

The last issue is the method of constructing an ENF map. However, because we can obtain ENF signals from the locations where live stream broadcasts exist, we need to consider how to obtain ENF signals at locations where we could not obtain multimedia. As mentioned earlier, the possibility of ENF extraction is highly dependent on the environment of online services and broadcasters. In other words, the possibility of collecting ENF signals from somewhere does not depend on our intention, but based on whether broadcasters exist in that place. In order to resolve this problem, we used a well-known spatial interpolation approach called inverse distance weighted (IDW) interpolation [2]. A weighted average of ENF signals at known positions is used to infer the ENF signal values of unknown positions and the IDW is defined by $Z^*(u) = \sum_{i=1}^{n} \lambda_i Z(u_i)$

where u_i is the location of the sample position within the neighborhood for $i = 1, 2, ..., n$. $Z^*(u)$ expresses the inverse distance estimate at the estimation location and n implies the number of sample positions. The notation λ_i for $i = 1, 2, ..., n$ denotes the individual weight assigned to a sample position, and conditioning data at sample positions are denoted as $Z(u_i)$. The term for weights are determined as $\lambda_i = \frac{\left(\frac{1}{d_i{}^p}\right)}{\sum_{i=1}^{n}\left(\frac{1}{d_i{}^p}\right)}$ where d_i are the Euclidean distances between estimation location and sample positions and p is a value of the power or distance exponent. More details to obtain the optimal p are given in the experiment section.

4 EXPERIMENTAL RESULTS

Figure 7 plots the constructed ENF maps of the United States and Europe in time series. The ENF signals of the areas where multimedia sources do not exist are inferred using inverse distance weighted (IDW) interpolation given known and sampled data. In this section, we describe the details of the data and the effectiveness of our proposed approach.

4.1 Used dataset and system environments

For performance evaluation, we collected a large amount of data for about nine months since the 11th May, 2016. The sampling rate of the online media was 44000Hz but we down-sampled it to 1000Hz for light computation. For the collection of the data, we used a server with 125.9GiB RAM, 24 Intel(R) Xeon(R) CPU E5-2630, 2.30GHz, and Ubuntu 14.04 (64-bit) OS. Additionally, we used Matlab and Python to collect and analyze the multimedia data and ENF signals.

Moreover, we collected the metadata of the multimedia as well as the multimedia data, since ENF signals and the metadata are necessary for analyzing local events and restoring multiple ENF signals. We collected data from Europe, Asia, Africa, North America, South America, and Oceania about signals, time, and coordinates information through Earthcam, Skyline webcams, Explore, and Ustream. Table 1 demonstrates more details about which countries we could collect multimedia from. The percentage of the signals that can be obtained from multimedia is shown via below subsection.

Table 1: The number of world multimedia that can be collected from four kinds of streaming services

Europe	Asia	Africa	North America	South America	Oceania
71	22	3	165	3	4

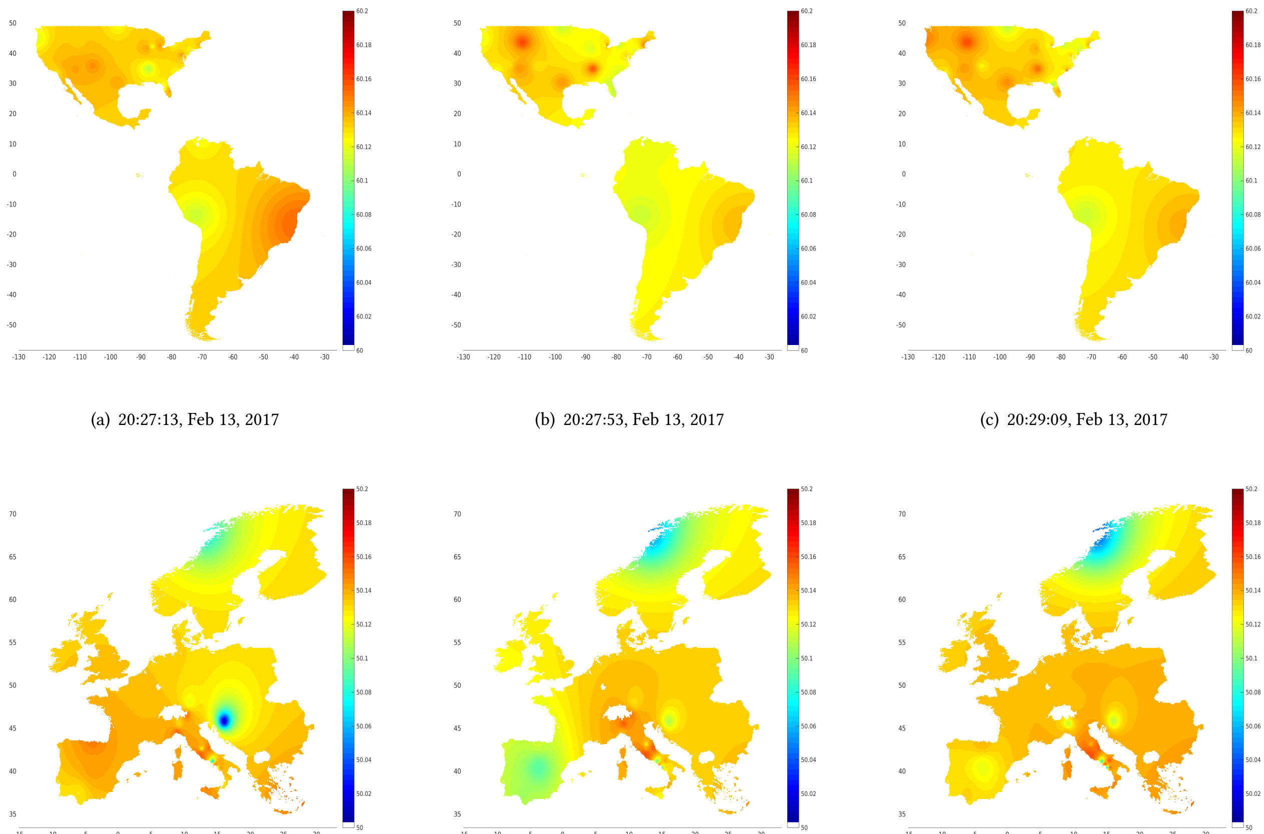
4.2 Validation of ENF signals extracted from online multimedia service

In this subsection, we validate the availability of ENF extraction from online multimedia by checking ENF presence with two steps: (1) data filtering for excluding spoiled multimedia and (2) classification for determining how much percentage of multimedia has ENF signals. The incoming multimedia are highly likely to be spoiled due to unwanted packet loss When massive data from online services are downloaded in communication channels with the limited network capacity. Thus, in order to collect the required amount of multimedia, it is inevitable to extend the number of collecting devices or the network capacity. To describe it more, we conducted a simple experiment and observed that about 150 multimedia can be concurrently downloaded on our single computer device. If the amount of collected data is larger, it would be more effective to collect stable signals that increasing the number of collecting devices. However, if the number of collecting devices and the network capacity are fixed, we need an alternative approach which excludes the spoiled multimedia. In this study, we selectively collected the multimedia data which are not spoiled from such unstable data acquisition process. The second procedure to evaluate the presence of ENF signals is to use a classification algorithm with a threshold to distinguish the ENF signals from noise. It is known that clean ENF signal has normally ±0.02Hz from the mean value and 0.03 is suggested as a threshold value for noise removal [7]. Therefore, we decided to use this criterion for classification. In this work, we have classified multimedia data into two groups depending on the presence or absence of the ENF signal. Based on the classification, our multimedia data were analyzed as shown in Table 2.

Table 2: Validating the presence of ENF signals using the threshold criteria

Source	Signal state (%)	
	Presence	Absence
Earth Cam	85.3	14.7
Skyline webcams	95.2	4.8
Explore	70.6	29.4
Ustream (gaming & wildlife)	42.1	67.9
Overall	72.1	27.9

From Earthcam online multimedia service, we could check that we can gather approximately 92 signals maximally in a single moment throughout the world. However, we selected 68 in total because other data was spoiled due to the limitation of network traffic. Then, about collected signals, our classification module decided that 85.3% of them had ENF signals, and 14.7% had not signals. Though skyline webcams provide hundreds of cameras throughout the world, we collected video from part of them. This is because many of them do not contain audio data which we can extract ENF signals. Therefore, we first filtered out unsuitable data and focused only about 70 multimedia which contain audio data. From the Skyline webcams, we downloaded 62 out of 70 multimedia. Our module decided that 95.2% data has signals, and 4.8% data does not have any signal. After that, we checked out the data collected through the Explore service. We collected a total of 34 signals from Explorer, and our decision module confirmed that only 70.6% of them contained signals. On the other hand, we were able to obtain about 76 videos broadcasting throughout the world from Ustream. For the collection of the Ustream data, we chose “gaming” and “wildlife” categories that are supposed to contain the strong ENF signals. In the Ustream videos, during the target period, there were 42.1% data decided to have ENF signals and 67.9% data decided to do not have any signal. The false positive and false negative error rates for the result of whole classification were approximate 24.85%



(a) 20:27:13, Feb 13, 2017

(b) 20:27:53, Feb 13, 2017

(c) 20:29:09, Feb 13, 2017

(d) 20:25:25, Feb 13, 2017

(e) 20:26:21, Feb 13, 2017

(f) 20:27:29, Feb 13, 2017

Figure 7: Constructed ENF maps of America (upper) and Europe (down) using the inverse distance interpolation method in our proposed approach: each image represents ENF signals with varying time. The time shown is based on UTC.

and 16.41%, respectively. Of course the quality of signals could be different depending on circumstances.

That is, the results explain the possibility of collecting world ENF signals through an online multimedia service instead of using additional hardware. Comparing with ground-truth ENF signals from FNET[18], we additionally could represent how much these signals are reliable. FNET system was established to collect ENF signals from the world. They installed about 167 devices in the world. However, only 100 devices were working at 08:20 PM, Feb 15, 2017 (UTC). We do not know why the many devices were not working. However, we argue that our approach is more efficient than FNET because we could collect ENF signals more than 167 places in the world.

4.3 Availability of signal alignment technique

We use the signal alignment technique to construct a semi-complete signal by using two or more different partial signals extracted from adjacent places as mentioned in 3.2.2. However, we should check the similarity between signals obtained from different locations to guarantee its availability and practicality. To figure out how the distance influences the similarity of the paired signals, we calculated each pair's correlation coefficient by using an NCC. We classified each signal's pair into a more detailed group with varying ranges such as "0 ∼ 300", "300 ∼ 600", "600 ∼ 900", "900 ∼ 1200", and "1200 ∼." Then, we calculated and averaged the NCC of each signal pair to prove that the measured distance influences the similarity of the signals.

Table 3: Signal similarity (ρ) according to the distance between two sources

Miles	0∼300	300∼600	600∼900	900∼1200	1200∼
ρ	0.81	0.55	0.53	0.45	0.32

Table 3 describes the result of the signal similarity experiment depending on the distance between paired signals. As mentioned previously, we averaged the coefficient values of each group separated by the distance of measured locations, and calculated the confidence interval based on the t-distribution. As we assumed, the

Table 4: Error rates of signal alignment technique depending on overlap lengths ($t + \tau$) and signal similarities (ρ)

$t + \tau$	ρ =0.33	ρ =0.44	ρ =0.56	ρ =0.70	ρ =0.82
20 min	23.69%	17.30%	17.17%	8.14%	1.42%
25 min	10.24%	5.99%	4.29%	4.48%	1.18%
30 min	3.08%	1.8%	1.91%	1.15%	0.00%

signal pairs collected in near locations had a higher similarity value than the pairs collected in far locations. Specifically, the group labeled "0 ~ 300," the nearest signals group, had the highest similarity, over 0.81, and the second group, which is farther than the first group, had the second highest coefficient value of approximately 0.55. The third group, sequentially, had a value slightly lower than the second group, to the nearest 0.53. The fourth group, which included 900 ~ 1200 miles signal pairs, had a lower coefficient value 0.45 than the third group. Last, the farthest group, "1200~," had the lowest similarity at approximately 0.32. Given the relation between the distance and the similarity of the paired ENF signals, We finally aligned the paired signals by searching for the aligning points using the normalized cross correlation (NCC).

Table 4 presents the error rates of signal alignment technique with varying signal similarities (ρ) and overlap lengths ($t + \tau$). For the experiment, we prepared 30 signal pairs ($sig_1^{(i)}, sig_2^{(i)}$) of about 1-hour length each for $i = 1, 2, ..., 30$. Two signals have the same section $t + \tau$ and random noise σ, that is ($sig_1{}^{(i)}+\sigma_1{}^{(i)}$, $sig_2{}^{(i)}+\sigma_2{}^{(i)}$, for i = 1, 2, ..., 30). With this signal pairs, we have experimented to find the requirements for the signal alignment by changing τ and ρ. For this procedure, we extracted signals via QIFFT. More precisely, the parameter of QIFFT were set as 2^{13} window size and 2^{12} overlap length. That is, ($t + \tau$) corresponding to 20 minutes, 25 minutes, and 30 minutes were describing the lengths of 300 points, 375 points, and 450 points respectively. The error rates are expressed in the table above. We obtained the error rates by calculating how much the alignment points deviate from correct positions. Therefore, the bigger the error rate, the less accurate the result. For more accurate result, we repeated this experiment 30 times with each signal pairs. The results are shown in the Table 4. The results show that both the overlap length and the similarity of the signals affect the performance, but the overlap length has larger effect. Therefore, to obtain more stable and reliable signals from this technique, using signal pairs with approximately 30 minutes of overlap length or at least 0.82 ρ is recommended.

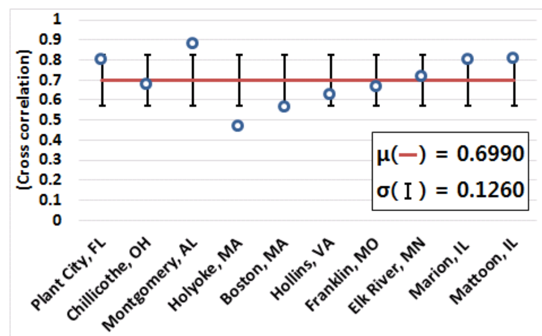
4.4 Validation of the usefulness of the IDW interpolation

In order to verify the availability of the inverse distance weighted (IDW) interpolation, we need to

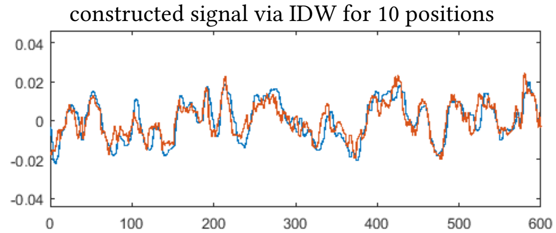
- find the optimal p to minimize the error of IDW using 4 fold cross-validation and
- compare the constructed ENF map with the ground-truth data.

First of all, we first randomly divided the data into 4 subsamples and labeled it into two groups: a training set and a testing set. The training set and the testing set were composed by 3 and 1 subsamples respectively. Afterward, we repeatedly calculated IDW for the target area with the training set and varying p from 1 to 3. Then, by comparing the each result of IDW and the testing set, we obtained the parameter p which leads the smallest error between the training set and the testing set. Parameter p would be changed depending on target area. Therefore, we need to estimate locally optimal p but we left this for further study.

Then, we conducted another experiment to prove the availability of the IDW interpolation by comparing to ground-truth data as follows. First, IDW interpolation is calculated using ENF data from 52 locations of eastern U.S. The data was collected for 40 minutes from 03:40 PM (UTC), 10 May 2016. Then, we checked the similarity using normalized cross correlation (NCC) between interpolated ENF signals and underlying ground-truth ENF signals collected from the FNET/GridEye server. The comparison is plotted by Figure 8.



(a) Similarity between ground truth signal and constructed signal via IDW for 10 positions



(b) The ground-truth signal with a red line and, constructed signal with a blue line

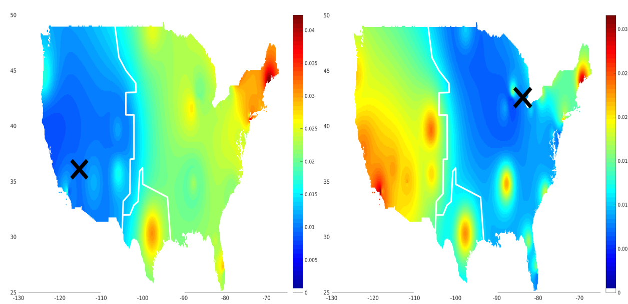
Figure 8: Comparing the similarity between ground truth signal and constructed signal via IDW

4.5 Estimating the time and the location using ENF signals

As we have already referred to, it is known that ENF signals can be effectively used for digital forensics since they can be considered a spatio-temporal pattern for identifying both location and time. A few groups have recently proved this by identifying the targeted power grid using ENF patterns [8, 11, 12]. After collecting unidentified recordings from different grids, the researchers identified the grid, which data were recored in, by using classification algorithms and between-grid distance metrics. However, they could not infer the recording time and location of the multimedia data if the data are collected from a single power grid. We realized that such limited circumstances can be tackled by calculating within-grid distance in our proposed world-wide ENF map. Since we have a full map of world-wide ENF map, we can identify the location and time with the multimedia data which are collected even in the same power grid.

For the evaluation, we attempt to identify the recording location and time of the 33 multimedia data from the online streaming networks in USA. In this experiment, the Leave-One-Out Cross-Validation is used. For each run, we constructed a full ENF map by interpolating 32 ENF signals. The size of the full map is 250-by-660 so it has 16500 cells. Afterwards, in order to calculating the similarity, we calculated Root Mean Square Distance (RMSD) between the ENF signals at 16500 cells and a test ENF signal. Given a test ENF signal $y^* \in \mathbb{R}^T$ and the ith cell of the constructed map $\mathcal{M}^{(i)} \in \mathbb{R}^T$, the location is inferred by $\hat{i} = \arg_i \min T^{-0.5} \sqrt{\sum_{j=1}^{T} \left(y_j^* - \mathcal{M}_j^{(i)} \right)^2}$ for all $i \in \{1, 2, \cdots, 16500\}$. Since there are three power grids in USA, we identify both the grid and its internal location using between-grid distance and within-grid distance.

Figure 9 plots the similarity maps of two different test ENF signals to estimate the hidden location. In this figure, the cross mark represents the hidden location of test ENF signal and the blue area is the location with the lowest distance between the test ENF signals and ENF signals at a cell. As shown in this figure, we can identify the grid and hidden location using our ENF map. In addition, the correct grids are identified and classified with the probability 96.96% (the 32 runs among 33 runs are succeeded and only one run failed). The data was collected for about 80 minutes from 08:25 PM, 13 February 2017 (UTC).



(a) (b)

Figure 9: Location Similarity Map with Root Mean Square Distance (RMSD). The cross mark represents the hidden location of test ENF signal and the blue area is the location with the lowest distance between the test ENF signals and ENF signals at a cell.

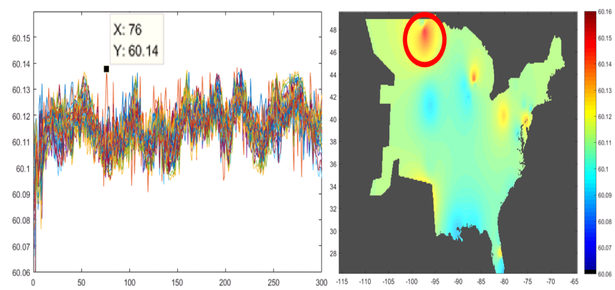
5 DISCUSSION

This paper primarily focuses on two subjects: (1) the collection of the proper online multimedia data for extracting ENF signals; and (2) the reconstruction of a reliable spatio-temporal worldwide ENF map. However, as we have already mentioned in the introduction, ENF signals have currently been used in many application domains, including abnormal event detection and digital forensics analysis.

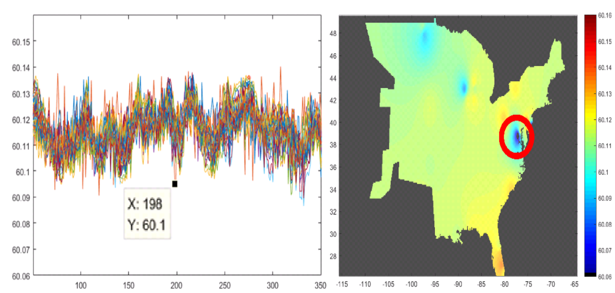
Although applying worldwide ENF maps into these various application domains is not a main goal of this paper, it is obviously interesting issue for the further work. Especially, we found that there are two major applications for the worldwide ENF map: (a) the detection and the prediction of abnormal events and (b) location privacy.

5.1 Detection and prediction of abnormal events using a worldwide ENF map

First of all, we may use the worldwide ENF map in order to detect and predict abnormal signals. For example, from the ENF map constructed by our proposed approach, we plotted about 50 trajectories of ENF signals to about 50 randomly selected locations at two different times in the United States. Figure 10 shows the trajectories of ENF signals at the selected locations. As we can see in the figure, we can find an abnormal status of the power system both in the 1D trajectory and in the 2D ENF map. This shows the usefulness of our ENF map for monitoring the status changes of the power system and detecting abnormal events including blackouts, and sudden disorder of the power system. The data was collected for about 23 minutes from 03:40 PM, 10 May 2016 (UTC).



(a)



(b)

Figure 10: Monitoring of the power system using ENF map: we can find the time and location of the abnormal states in power system using this visualized ENF map. The detection of the abnormal events can automatically be developed with additional outlier detection algorithms.

In addition, we can detect and predict catastrophic disasters like earthquakes and Typhoon by monitoring the real-time worldwide ENF map. That is, ENF signal can be used as a side channel and a unique signature in that the status of the power grid systems can be changed when they are damaged by such unwanted catastrophic disasters. For instance, it would be possible to detect anomalies at a few hours/minutes before the earthquake related to Italy earthquakes (http://www.bbc.com/news/world-europe-38663608). Therefore, the earthquake prediction system can be developed using worldwide ENF map.

5.2 Digital forensics and location privacy

Section 4.5 demonstrates the possibility of inferring the recording location and time of the multimedia data using the reconstructed ENF map. Therefore, we can use this for digital forensics as referred to in the introduction. For instance, when a certain multimedia file is submitted as an evidence to the court, using our worldwide ENF map we can determine (1) when and where the file was created and (2) whether the file is modified.

Contrast to the digital forensics, our reconstructed ENF map can be used to infer hidden personal information. The hidden location can be tracked by an attacker with malicious intent. For example, even if a user uploads the video or audio files to SNS without GPS location information, the malicious attacker can track his/her location information by using ENF maps.

5.3 Construction of a world scale ENF map using other services

In this paper, we demonstrated the possibility of construction of world ENF maps for various locations (such as Europe, Asia, Africa, North America, South America, and Oceania). We have endeavored to collect ENF signals from the world as many as possible. As a result, we could collect ENF signals from the more locations than what Grideye/FNET covers. However, it is still not enough to cover the whole world, so we are planning to carry out further research to cover a wider area. For more powerful performance, we will conduct further work using various streaming services such as "Facebook live", "Xiandanjia", and "Yingke" which can cover China, Southeast Asia, and any other countries throughout the world.

6 CONCLUSION

In this paper, we proposed a new alternative approach to collecting ENF signals from online multimedia data and reconstructing a worldwide ENF map. Because the data sources from which the ENF signals are extracted are open in the Internet, our proposed approach has significant benefit compared to previous approaches using physical hardware and sensors. For example, our proposed approach requires a relatively lower budget and simple management in that all things are controlled in software and the data are public-domain and free of charge. However, our approach requires more advanced signal processing tasks because the online multimedia data includes unwanted content. The contents are considered as noise. We demonstrate several algorithms to reduce the noise and enhance the ENF signals.

In conclusion, we claim that our approach is much more efficient and stable than previous methods using hardware. Although the quality of the signals from such hardware would be better, our approach is economically and practically improved over the previous method because we have eliminated tasks associated with devices, such as design, installation, and maintenance.

7 ACKNOWLEDGMENTS

This research is supported by Korea University's special research fund (K1711541). We thanks for Eunchong Lee who helped the crawling of live streaming data.

8 A DEMONSTRATION FILM IN YOUTUBE

Our proposed approach's demo film is also uploaded in youtube. The link is https://youtu.be/co6s9txDVu8 for viewing and downloading. This film includes the way and procedure about how to collect the multimedia data, how to extract and enhance the ENF signals and how to reconstruct the worldwide ENF map.

REFERENCES

[1] Mototsugu Abe and Julius O Smith III. 2004. Design criteria for simple sinusoidal parameter estimation based on quadratic interpolation of FFT magnitude peaks. In *Audio Engineering Society Convention 117*. Audio Engineering Society.

[2] Olena Babak and Clayton V Deutsch. 2009. Statistical approach to inverse distance interpolation. *Stochastic Environmental Research and Risk Assessment* 23, 5 (2009), 543–553.

[3] Fabrice Bellard, M Niedermayer, and others. 2012. FFmpeg. *Availabel from: http://ffmpeg. org* (2012).

[4] Dima Bykhovsky and Asaf Cohen. 2013. Electrical network frequency (ENF) maximum-likelihood estimation via a multitone harmonic model. *Information Forensics and Security, IEEE Transactions on* 8, 5 (2013), 744–753.

[5] Lang Chen, Penn Markham, Chien-fei Chen, and Yilu Liu. 2011. Analysis of societal event impacts on the power system frequency using FNET measurements. In *Power and Energy Society General Meeting, 2011 IEEE*. IEEE, 1–8.

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

[7] Alan J Cooper. 2009. An automated approach to the Electric Network Frequency (ENF) criterion-Theory and practice. *International Journal of Speech, Language and the Law* 16, 2 (2009), 193–218.

[8] Ravi Garg, Adi Hajj-Ahmad, and Min Wu. 2013. Geo-location estimation from Electrical Network Frequency signals.. In *ICASSP*. 2862–2866.

[9] Catalin Grigoras. 2007. Applications of ENF criterion in forensic audio, video, computer and telecommunication analysis. *Forensic Science International* 167, 2 (2007), 136–145.

[10] Jiahui Guo, Yanzhu Ye, Ye Zhang, Yin Lei, and Yilu Liu. 2014. Events associated power system oscillations observation based on distribution-level phasor measurements. In *T&D Conference and Exposition, 2014 IEEE PES*. IEEE, 1–5.

[11] Adi Hajj-Ahmad, Radhika Garg, and Min Wu. 2013. ENF based location classification of sensor recordings. In *Information Forensics and Security (WIFS), 2013 IEEE International Workshop on*. IEEE, 138–143.

[12] Adi Hajj-Ahmad, Ravi Garg, and Min Wu. 2015. ENF-based region-of-recording identification for media signals. *IEEE Transactions on Information Forensics and Security* 10, 6 (2015), 1125–1136.

[13] Bin Qiu, Ling Chen, Virgilio Centeno, Xuzhu Don, and Yilu Liu. 2001. Internet based frequency monitoring network (FNET). In *Power Engineering Society Winter Meeting, 2001. IEEE*, Vol. 3. IEEE, 1166–1171.

[14] Joseph Tabrikian, Shlomo Dubnov, and Yulya Dickalov. 2004. Maximum a-posteriori probability pitch tracking in noisy environments using harmonic model. *IEEE Transactions on Speech and Audio Processing* 12, 1 (2004), 76–87.

[15] Lei Wang, Jon Burgett, Jian Zuo, Chun Chun Xu, Bruce J Billian, Richard W Conners, and Yilu Liu. 2007. Frequency disturbance recorder design and developments. In *Power Engineering Society General Meeting, 2007. IEEE*. IEEE, 1–7.

[16] Yanzhu Ye and Yilu Liu. 2012. Monitoring power system disturbances based on distribution-level phasor measurements. In *Innovative Smart Grid Technologies (ISGT), 2012 IEEE PES*. IEEE, 1–8.

[17] Lingwei Zhan, Jianyang Zhao, Jerel Culliss, Yong Liu, Yilu Liu, and Shengyou Gao. 2015. Universal Grid Analyzer design and development. In *Power & Energy Society General Meeting, 2015 IEEE*. IEEE, 1–5.

[18] Yingchen Zhang, Penn Markham, Tao Xia, Lang Chen, Yanzhu Ye, Zhongyu Wu, Zhiyong Yuan, Lei Wang, Jason Bank, Jon Burgett, and others. 2010. Wide-area frequency monitoring network (FNET) architecture and applications. *Smart Grid, IEEE Transactions on* 1, 2 (2010), 159–167.

[19] Yingchen Zhang Lang Chen Penn N. Markham R. Matthew Gardner Yilu Liu Zhiyong Yuan, Tao Xia. 2010. Inter-area oscillation analysis using wide area voltage angle measurements from FNET. *IEEE Power and Energy Society General Meeting* (2010).