ENF Based Location Classification of Sensor Recordings

Adi Hajj-Ahmad, Ravi Garg, and Min Wu

Department of Electrical and Computer Engineering
Institute for Advanced Computer Studies
University of Maryland, College Park, USA
{adiha, ravig, minwu}@umd.edu

***Abstract*—The Electric Network Frequency (ENF) signal can be captured in multimedia signals recorded in areas of electrical activity. This has led to the emergence of many forensic applications based on the use of ENF signals such as validating the time-of-recording of an ENF-containing multimedia signal or estimating its recording location based on concurrent reference signals from power grids. In this paper, we examine a novel application based on the use of the ENF signal that seeks to estimate the power grid in which an ENF-containing multimedia signal was recorded without relying on the availability of concurrent power references. We derive features based on the statistical differences in ENF variations between different grids to serve as signatures for the grid-of-recording of an ENF-containing signal. We use these features in a multiclass machine learning system that is able to identify the grid-of-recording of a signal with a high accuracy.**

I. Introduction

Electric Network Frequency (ENF) signals have emerged in recent years as an important trace of evidence for multimedia forensics. The ENF is the supply frequency in power distribution grids. It has a nominal value of 60Hz in North America and 50Hz in most other parts of the world. The instantaneous ENF usually fluctuates around its nominal value due to load variations and control mechanisms in the grid. The ENF signal is the changing values of the ENF over time. These variations in the ENF can be considered a random process over time, yet the variation trends are almost identical in all locations of the same grid at a given time due to the interconnected nature of the grid.

The ENF signal can be extracted from power signals measured from a power outlet using a step-down transformer and a simple voltage divider circuit. To estimate the instantaneous ENF signal, the power signal is divided into time-frames, and frequency estimation algorithms are applied on each frame to determine its dominant frequency [1]. The importance of the ENF for multimedia forensics emerges because the ENF can also be present in audio or video recordings due to electromagnetic influences in the place of recording. It has been shown that the ENF variations extracted from an audio/video signal match with the ENF variations extracted from a clean power signal recorded at the same time and in the same power grid as the audio/video signal [1]–[3].

WIFS'2013, November 18-21, 2013, Guangzhou, China.
ISBN 978-1-4673-5593-3

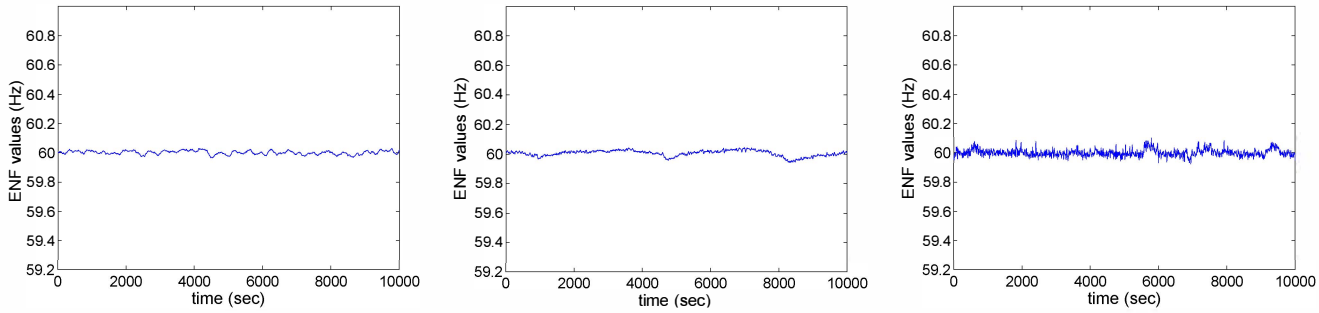
Several forensic applications based on the use of the ENF signal have been proposed. Recent works have shown that the ENF can be used to detect tampering or modifications in a multimedia signal [3], [4]. Also, when making use of power signals as references, the ENF can be used to estimate/validate the time-of-recording of the multimedia signal as well as its location-of-recording in terms of grid or within a grid [1], [2], [4]. Most of these applications require the knowledge of the grid or exhaustively search for the grid in which the multimedia signal was recorded using concurrent power references. In this paper, we investigate a new problem that seeks to estimate the grid in which the ENF-containing signal was recorded, without having concurrent power references.

We have collected power and audio recordings from a number of different grids around the world. Upon extracting the ENF signals from these recordings, we have noticed that there are differences between them in the nature and manner of the ENF variations. We hypothesize that processing an ENF signal to extract its statistical features may facilitate the identification of the grid in which it was recorded. Following this, we have devised a machine learning system that learns the characteristics of ENF signals from different grids, and used it to classify ENF signals in terms of their grid-of-recording. Such a system that identifies the grid in which a multimedia signal was recorded, without having concurrent power references to compare with, can be very important for multimedia forensics and security. It can pave the way to identify the origins of such videos as those of terrorist attacks, ransom demands, or child pornography and exploitation [5].

The rest of this paper is organized as follows. Section II examines the ENF signals collected from different grids and presents the proposed features to be used for the machine learning implementation. Section III describes the experimental setup and discusses the results obtained. Section IV concludes the paper.

II. Location-Dependent Features in ENF Signals

We have collected power and audio recordings from seven different grids; among them, three grids have a nominal ENF of 60Hz and four grids have a nominal ENF of 50Hz. These recordings are collected over different days and times to



(a) US East (b) US West (c) Quebec, Canada

Fig. 1. Sample ENF signals extracted from power recording from different 60Hz grids.

take into account the possible variations in ENF properties at different day of a week and time of a day, and identify features that are dependent on locations and are robust to these longer-term variations. For each ENF-containing recording, we extracted the ENF signals from consecutive non-overlapping frames, each of which is 5 seconds long. Upon examining the ENF signals obtained from the different grids, we observe several differences that can be used to extract meaningful features for grid classification. In this section, we examine these statistical differences and discuss the features that we choose for grid classification.

A. Comparing ENF Signals in Different Grids

For the 60Hz grids, we have recordings from three North American grids: Eastern North America (or US East for short), Western North America (or US West) and Quebec. Among 50Hz grids, we have recordings from China, India, Ireland and Lebanon. Sample ENF signals extracted from the clean power signals from all these grids are shown in Fig. 1 and Fig. 2.

Examining these figures, we observe several differences between the ENF signals originating from different grids. The first discerning feature is the mean of an ENF signal. We can easily tell if a signal belongs to a 50Hz or 60Hz grid depending on how close its mean is to either 50Hz or 60Hz. Within the signals whose means are similar, there are also some notable differences. For instance, among the 50Hz ENF signals, Lebanon's ENF signal mean is above 50Hz most of the time.

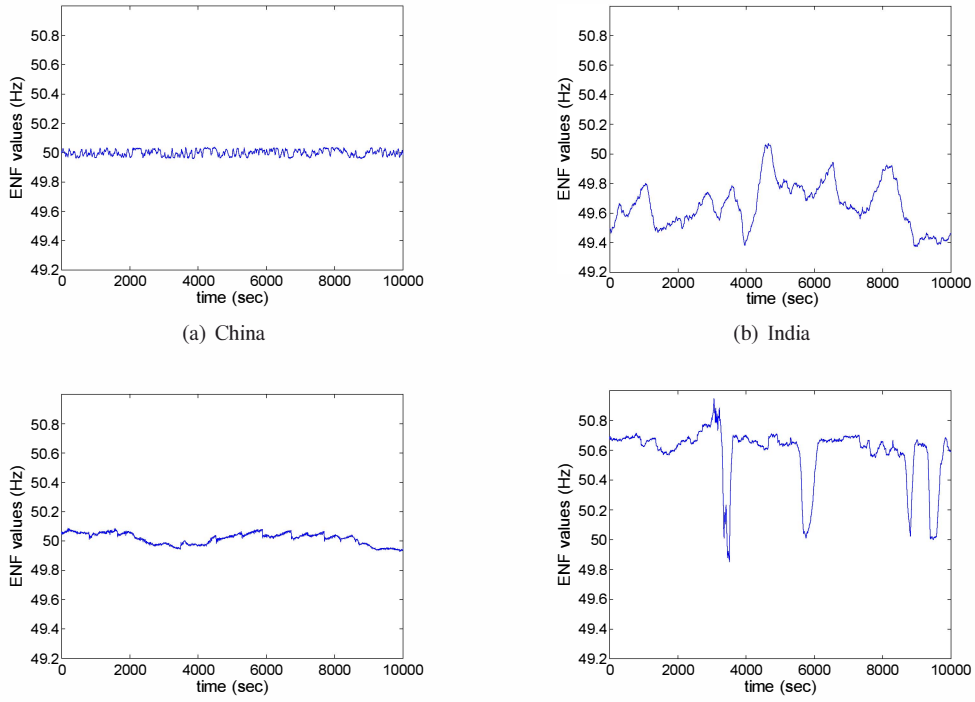
The ENF signals from various grids also differ in terms of their temporal variations. Our data show that among the 60Hz grids, the US East and US West ENF signals appear the most similar to each other, although the variations of US East's ENF are better contained while US West's ENF seems to drift a bit more before returning to the nominal value. Quebec's ENF exhibits more variations than the US ENFs. India has a larger range of variations as compared with most other grids. The ENF for China tends to vary at a different rate than that of Ireland, and Ireland's ENF, though somewhat controlled, shows a tendancy to drift before returning to the nominal value. Lebanon has frequent outliers dropping by almost 1Hz, a characteristic that does not appear in the other ENF signal samples.

To understand these different ENF variations between grids, we recall that the ENF changes due to load changes in the power grid. The control mechanism reacts to such changes by adjusting the power generation to regulate the ENF and try to bring it back toward the nominal value. Different power grids may have different control mechanisms as well as power generation/supply capabilities, thereby affecting the effectiveness and manner in which they are controlling the variations. As ENF variations are similar throughout an interconnected system and are related to the relative imbalance between generation and load and to power-frequency control, larger power grids generally tend to have smaller frequency variations [6]. Our observations on ENF signals reflect these general characteristics of power grids: the US grids have better control mechanisms and power resources than most other grids that we have observed, so their range of ENF variations is very small (around $\pm$0.02Hz). The large grids of US and China exhibit smaller variations than the other smaller grids. Lebanon has a very small grid and limited power resources, which is reflected in its grid's large ENF variations. India's large variations in ENF despite its fairly large grid size can be attributed to its power resources and the control mechanisms governing its grid.

B. Feature Extraction

After examining the empirical differences among grids, we now discuss quantitative features that can be extracted. For our current discussion, we consider having a set of ENF signal segments, $s[n]$'s, of fixed size S from candidate power grids, where these ENF segments correspond to ENF-containing signals whose size is on the order of minutes.

The mean of an ENF segment is a clear feature component to include. Another helpful feature is the variance of the segment, or its dynamic range (the maximum ENF value minus the minimum value). To develop other features, we found it beneficial to apply a transformation to the ENF segment and then examine the statistical properties of the transformation as potential features. More specifically, we consider wavelet signal analysis to study signals at multiple time-frequency



(a) China (b) India

(c) Ireland (d) Lebanon

Fig. 2. Sample ENF signals extracted from power recordings from different 50Hz grids.

resolutions [7]. We apply an L-level dyadic wavelet decomposition, where each level provides an approximation to the original signal and the detailed variations at a specific level of resolution. We compute the variances of the high-pass band of each decomposition level (the details) and also the variance of the lowest time-frequency band (the approximation) as candidate features. Fig. 3 shows an illustration of a 2-level wavelet decomposition. In this case, the signals whose variances we take as feature components are *Approximation* 2 and *Details* 1 and 2.

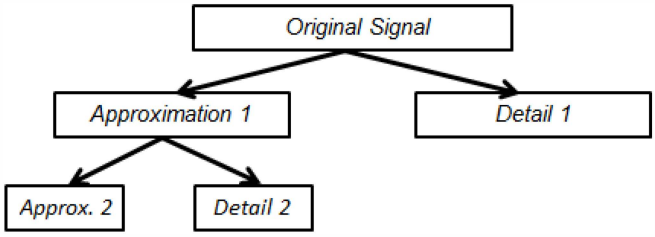


Fig. 3. Illustration of 2-level wavelet decomposition.

Complementing the wavelet features, we extract a set of features obtained from a statistical modeling of the ENF signal. Our recent work has proposed an autoregressive (AR) model of order 2 for ENF signals [8], [9]. An ENF signal $s[n]$ would be modeled as:

$$s[n] = a_1 s[n-1] + a_2 s[n-2] + v[n] \quad (1)$$

The original study was made on ENF signals from the United States, but the idea can be extended to examine ENF signals from other grids. We consider three feature values from this AR modeling: the two AR parameters resulting from the modeling, a_1 and a_2, and the variance of the model's innovation signal $v[n]$. The variance of the innovation signal is an indicator as to how well the signal can be fitted into the AR(2) model, and the AR parameters entail the relations between samples of $s[n]$.

TABLE I
PROPOSED FEATURE COMPONENTS

Index	Features
1	Mean of ENF segment
2	log(range) of ENF segment
3	log(variance) of approximation after wavelet analysis ($L = 4$)
4-7	log(variance) of four levels of detail signals computed through wavelet analysis ($L = 4$)
8-9	AR(2) modeling parameters
10	log(variance) of the innovation signal after AR(2) modeling

III. Experiments and Results

A. Experimental Setup

a) Dataset preparation: The feature components that we use for location classification are summarized in Table I. We apply a log operator on the range and variance feature values to focus on their orders of magnitude. Feature values are extracted from ENF signal segments of equal size S. We use $S = 96$ for our implementation that corresponds to an ENF-containing

TABLE II
AVAILABLE TRAINING AND TESTING EXAMPLES

Number of examples	Training - power	Testing - power	Training - audio	Testing - audio
US East	242	242	64	65
US West	97	98	34	34
Quebec	47	48	0	0
China	197	198	0	0
India	47	47	0	0
Ireland	124	125	82	82
Lebanon	429	429	103	103

TABLE III
DESCRIPTION OF THE TRAINED SVM SYSTEMS

System	Num. of classes	Description
I	7 classes	Trained only on ENF extracted from power signals.
II	4 classes	Trained only on ENF extracted from audio signals.
III	7 classes	Trained on all available training data, assuming that the audio and power ENF signals obtained from the same grid as belonging to the same class.

TABLE IV
ACCURACIES FOR DIFFERENT SYSTEMS AVERAGED OVER 20 ROUNDS

System	Testing on power data	Testing on audio data
I	98.43%	74.76%
II	75.10%	94.33%
III	97.95%	93.71%

signal of 8 minutes long, and the frame size for computing one instantaneous ENF point is 5 seconds. The computed feature values are normalized to the range of $[-100, 100]$ by linear scaling, whereby the k^{th} feature value in a training example is normalized according to the feature values in position k of all training examples. The normalization parameters are stored and later applied to the testing examples to normalize them.

b) Choice of supervised learning mechanism: We use a Support Vector Machine (SVM) to build the location classifier. Our implementation makes use of the LIBSVM library [10], which implements a multiclass SVM that uses an all vs. all approach: For a total of M classes, the system trains $\binom{M}{2}$ binary classifiers; each binary classifier is trained on one of the $\binom{M}{2}$ possible pairs of classes, learning to differentiate between the respective two classes. When testing the trained system, the testing example is passed through all binary classifiers, and votes are assigned to each possible class based on which class emerges as the winner from each binary classification task. The final winning class is the class with the largest number of votes. For a testing example, the LIBSVM implementation also provides M probability (confidence) values, where the j^{th} probability value gives the estimated probability that the testing example belongs to the j^{th} class.

Taking 50% of the data as testing data, and following our choice of S, the numbers of training and testing examples that we have for each grid are listed in Table II. Due to logistical and resource constraints in collecting recordings from various grids around the world, the available data is imbalanced: We have significantly more recordings from some grids than others, and we do not have audio recordings from all the grids considered. This imbalance in training data can create a bias problem or overfitting when testing the system. If a system is trained on a dataset where the majority of the training examples belong to one class, it tends to be more biased to assign the testing examples to this majority class [11]. To tackle this issue, we use a variant of SVM called the weighted SVM, which is supported by LIBSVM.

SVM implementations usually include a fixed cost value C which controls the penalty on making a mistake while classifying an example. The weighted SVM addresses the issue of imbalanced data through assigning different cost values for examples from different classes. The larger class has a smaller cost value than the smaller class, which means that the penalty for making a mistake on an example from the smaller class would be larger [11]. Here, with M classes, the cost for class j that has N_j training examples would be $w_j \cdot C$ with:

$$w_j = \frac{N_{min}}{N_j}, \text{ where } 1 < j < M \text{ and } N_{min} = \min_j N_j. \quad (2)$$

Our implementation chooses the Radial Basis Function (RBF) kernel for SVMs. Using the LIBSVM library, two important parameters need to be chosen for the RBF kernel: the cost parameter C and a parameter γ. For each SVM classifier that we train, we carry out a ten-fold cross validation examination to select these two parameters.

c) Systems to be trained: As detailed in Table II, the data that we have are of two main types: ENF segments extracted either from power recordings or audio recordings. Generally, ENF segments extracted from power recordings are cleaner signals with high signal-to-noise ratio (SNR), while the ENF segments extracted from audio recordings can be noisy. This would affect the quality of the feature values extracted. In order to understand the extent to which a machine learning system would be able to successfully classify ENF segments of varying noise conditions into grids, we train three different SVM systems and then examine the testing results obtained from each. The differences between the three systems are listed in Table III.

If the decision given by the multi-class SVM stage for a testing example has a confidence lower than a threshold (e.g. 0.6), we advance this example to a second stage. In this new stage, we subject the example to a binary SVM classifier trained on the two classes that received the highest confidence in the first stage. This would imply storing $\binom{M}{2}$ binary classifiers trained on all possible binary combinations of the classes, in order to help reach more confident final decisions for the testing data.

B. Results and Discussions

a) Results overview: In this subsection, we present the results obtained for testing the data on the three systems listed in

TABLE V
CONFUSION MATRIX FOR POWER ENF TESTING DATA ON SYSTEM I (ACCURACIES AVERAGED OVER 20 ROUNDS)

Testing Classes	Num. of examples	US East	US West	Quebec	China	India	Ireland	Lebanon
US East power	242	98.86%	1.14%	0	0	0	0	0
US West power	98	2.11%	97.88%	0	0	0	0	0
Quebec power	48	0	0	100%	0	0	0	0
China power	198	0	0	0	99.21%	0.08%	0.71%	0
India power	47	0	0	0	5.30%%	93.42%	1.17%	0.11%
Ireland power	125	0	0	0	0.04%	0	99.84%	0.12%
Lebanon power	429	0	0	0	0	0.20%	0	99.80%

Table III. We choose 0.6 as the threshold for the confidence value to advance to a further binary classification stage. For each system, we show the power and audio testing results separately to understand how well the system is suited for the two types of signals.

To generate our results, we first divide all the available data into six groups, where each group contains approximately one-sixth of the examples from each grid. Then, we train each of our three SVM systems for 20 rounds, considering in each round a different combination of three groups out of the six as training data and the remaining three groups as testing data. The results shown here are averaged over the results of these 20 training/testing rounds, where the number of training and testing examples are approximately equal to each other for a given grid.

Table IV lists the accuracies achieved for testing on each system for both the clean ENF from power signals and the noisy ENF from audio signals. Testing the power data on a system trained on power data (i.e. Systems I and III) results in high testing accuracy of about 98% while testing this power data on the system trained only on audio data (i.e. System II) results in a lower accuracy of only about 75%. Similarly, testing the audio data on a system trained on audio data (i.e. Systems II and III) results in high accuracy of about 94% while testing this audio data on a system not trained on audio data (i.e. System I) results in a lower accuracy of only about 75%. This shows that it is important to incorporate into training similar signal conditions that the system would be anticipated to see in testing.

b) Close examination of results: To understand these results better, we examine the confusion matrices of the three systems considered. These matrices are shown in Tables V, VI and VII, respectively. In each of these tables, the labels of the rows denote the actual grids and the conditions of the signals tested, while the labels of the columns denote the grids predicted by the system. The tables show how well the testing examples from each grid and signal condition were classified when applied to our trained systems. We include the approximate number of testing examples available for each class to highlight the difference in the number of examples available for each class in order to help present the proper context for the corresponding testing percentages. As Systems I and II proved to be ineffective at classifying data they are not trained on, we do not present the testing of audio data on System I or power data on System II due to space limit.

TABLE VI
CONFUSION MATRIX FOR AUDIO ENF TESTING DATA ON SYSTEM II (ACCURACIES AVERAGED OVER 20 ROUNDS)

Testing Classes	Num. of examples	US East	US West	Ireland	Lebanon
US East audio	65	93.25%	6.75%	0	0
US West audio	34	11.27%	88.73%	0	0
Ireland audio	82	0	0	98.96%	1.04%
Lebanon audio	103	0	0	3.63%	96.37%

The entries on the diagonals of the tables show the percentage of correct classification. We can see that with the incorporation of mean ENF value as a feature, the 50Hz signals are never mistaken for 60Hz signals, and vice versa.

Considering the power signals, we can see that the correct classification percentages in Tables V and VII fall in the high range of 93-100%. Quebec signals are notable for their 100% identification rate, due to the clear distinction in the range and nature of their variations as compared with the more controlled US signals. US East and US West signals become mistaken for each other, which is understandable given the close similarity between them in control mechanisms and power resources as seen in Fig. 1. Notably, signals from India can be mistaken for signals from China or Lebanon. A possible reason behind this is that India's ENF signal is inconsistent in its properties, with the varying mean and different variations. In addition, we have fewer data from India as compared with the data from other 50Hz grids.

Considering the audio signals, we can see that the correct classification percentage range in Tables VI and VII drops to 88-93% for the 60Hz US signals and 95-99% for the 50Hz signals in Ireland and Lebanon. The different performance between the two groups is because the US signals are more similar to each other than Ireland vs. Lebanon. The drop in classification accuracies as compared with the power signals can be explained by the nature of the audio ENF signals. The ENF from audio signals tends to be noisier than from the power signals, and the amount of noise and distortions can be different even within signals of the same class, due to different recording conditions. This can create confusions for the machine learning system when defining the class

TABLE VII
CONFUSION MATRIX FOR POWER AND AUDIO ENF TESTING DATA ON SYSTEM III (ACCURACIES AVERAGED OVER 20 ROUNDS)

Testing Classes	Num. of examples	US East	US West	Quebec	China	India	Ireland	Lebanon
US East power	242	99.19%	0.81%	0	0	0	0	0
US West power	97	4.83%	95.17%	0	0	0	0	0
Quebec power	47	0	0	100%	0	0	0	0
China power	197	0	0	0	99.06%	0.05%	0.81%	0.08%
India power	47	0	0	0	4.35%	93.52%	0	1.27%
Ireland power	125	0	0	0	0.12%	0.04%	98.87%	0.96%
Lebanon power	429	0	0	0	0	0.14%	0.01%	99.85%
US East audio	65	88.99%	9.84%	1.16%	0	0	0	0
US West audio	34	9.23%	90.76%	0	0	0	0	0
Ireland audio	82	0	0	0	0.12%	0	99.82%	0.06%
Lebanon audio	103	0	0	0	0.16%	0.05%	4.50%	95.30%

boundaries and could lead to more mistakes in identification. Another reason is the fewer amount of audio data available to us.

Comparing Systems I and II with System III, we can see from Table IV that the correct classification percentages for System III are slightly smaller than their counterparts in Systems I and II. System III defines classes as a mixture of audio and power ENF signals, which means that signals belonging to one class have a larger range of differences from one another due to their varying noise levels.

Overall, by incorporating training examples of multiple conditions, we have developed a machine learning based system (System III) that achieves a high classification accuracy on ENF signals extracted from both power and audio recordings. Meanwhile, if the test signal's condition in terms of noise and distortion is known a priori or can be estimated well, classification performance may have a small amount of further improvement by employing a system that is well trained on the corresponding conditions (Systems I and II).

IV. CONCLUSIONS AND FUTURE WORK

In this paper, we have developed a machine learning based system that can identify the grid of origin of an ENF signal without having concurrent power references. ENF signals from different power grids display different statistical characteristics, which can be exploited to identify their power grid. These differences are attributed to the size of power grids and the techniques and available power/energy resources by which the grids are controlled and operated. We have presented and compared three machine learning systems that are trained on identifying the origin of ENF signals embedded in clean power signals and/or noisy audio signals from different grids. For the system that gives the all-around high performance, System III, we have achieved an average accuracy of 98% on identifying ENF signals extracted from power signals from seven candidate power grids, and an average accuracy of 94% on identifying ENF signals extracted from audio signals from four candidate power grids.

This work presents a new capability of using ENF signals for multimedia forensics, by identifying the grid of origin of an audio/video recording via extracting and classifying its ENF signal. In our ongoing and future work, we are collecting more data from these and other power grids, and investigating how to further enhance the performance. We plan to explore additional features and examine the benefits of incorporating them.

ACKNOWLEDGMENTS

This work is supported in part by NSF grants #1309623, University of Maryland ADVANCE grant and Kulkarni Fellowship. The authors would like to thank Imad Atshan, Yunfang Feng, Jad Hajj-Ahmad, Shan He, Wenjun Lu, Michael Luo, Ashwin Swaminathan and Avinash Varna for their assistance in collecting power and audio recordings from different power grids.

REFERENCES

[1] A. Hajj-Ahmad, R. Garg, and M. Wu, "Instantaneous frequency estimation and localization for ENF signals," in *Proc. of APSIPA Annual Summit & Conf.*, 2012.

[2] C. Grigoras, "Applications of ENF analysis in forensic authentication of digital and video recordings," *Journal of Audio Engr. Society*, 2009.

[3] D. Rodriguez, J. Apolinario, and L. Biscainho, "Audio authenticity: Detecting ENF discontinuity with high precision phase analysis," *IEEE Trans. on Info. Forensics & Security*, 2010.

[4] R. Garg, A. L. Varna, and M. Wu, ""Seeing" ENF: natural time stamp for digital video via optical sensing and signal processing," in *Proc. of the 19th ACM International Conf. on Multimedia*, 2011.

[5] "UNICEF: Child protection from violence, exploitation and abuse child trafficking," URL: http://www.unicef.org/protection/57929_58005.html.

[6] M. Bollen and I. Gu, *Signal Processing of Power Quality Disturbances*. Wiley-IEEE Press, 2006.

[7] P. P. Vaidyanathan, *Multirate Systems & Filter Banks*. Prentice Hall, 1993.

[8] R. Garg, A. L. Varna, A. Hajj-Ahmad, and M. Wu, ""Seeing" ENF: Power signature based timestamp for digital multimedia via optical sensing and signal processing," *IEEE Trans. on Info. Forensics & Security*, 2013.

[9] R. Garg, A. L. Varna, and M. Wu, "Modeling and analysis of electric network frequency signal for timestamp verification," in *Proc. of IEEE Workshop on Info. Forensics & Security*, 2012.

[10] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. on Intelligent Systems & Technology*, vol. 2, pp. 27:1–27:27, 2011, software available at http://www.csie.ntu.edu.tw/∼cjlin/libsvm.

[11] Y. Huang and S. Du, "Weighted support vector machine for classification with uneven training class sizes," in *Proc. of the 4th International Conf. on Machine Learning & Cybernetics*, 2005, pp. 18–21.