IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 17 June 2023, accepted 7 August 2023, date of publication 5 September 2023, date of current version 20 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3312181

SURVEY

ENF Based Digital Multimedia Forensics: Survey, Application, Challenges and Future Work

ERICMOORE NGHARAMIKE[1], LI-MINN ANG[1], (Senior Member, IEEE), KAH PHOOI SENG[2,3], (Senior Member, IEEE), AND MINGZHONG WANG[1]

[1]School of Science, Technology and Engineering, University of the Sunshine Coast, Petrie, QLD 4502, Australia
[2]School of AI and Advanced Computing, Xi'an Jiaotong-Liverpool University, Suzhou 215123, China
[3]School of Computer Science, Queensland University of Technology, Brisbane, QLD 4000, Australia

Corresponding author: Li-Minn Ang (lang@usc.edu.au)

ABSTRACT The electric network frequency (ENF) represents the transmission frequency of the electrical grid and fluctuates constantly around 50 Hz or 60 Hz, subject to the region. This constant fluctuation, caused by the continuous mismatch in power demand and supply, makes the ENF a unique signature, which can be utilized for several applications. According to studies, the ENF may be intrinsically implanted in audio recordings captured by digital audio recorders (e.g., microphone-based voice recorder) plugged into the mains supply or are situated close to sources of power and power transmission cable due to electromagnetic field interference originated from power source or acoustic hum and mechanical vibrations emitted by electrically operated devices such as regular household appliances. Recent studies further observed that video recordings made under an illumination source powered by main power can pick up the ENF signal. Following this discovery, several research efforts have been invested towards successful and accurate extraction of the ENF signal, and utilizing the ENF signal retrieved for several applications, including time stamp verification, audio/video authentication, location of recording estimation, power grid identification, estimation of camera read-out time, and video record synchronization. To the best of our knowledge, there has been no comprehensive survey on ENF-based multimedia forensics. Thus, in this paper, we present a comprehensive survey of studies conducted in this field, identifying several application specifics, current challenges, and future research directions.

INDEX TERMS Electric network frequency (ENF), ENF detection, ENF estimation, ENF audio and video forensic, ENF applications.

I. INTRODUCTION

With the majority of the younger generation preferring images and videos as the major mode of communication, an increasing amount of multimedia content is generated and shared via the Internet. Because of the enormous amount of information contained in multimedia content, such as images, audio, and video recordings, they have become targets of malicious attacks to falsify digital content. Advances in digital forensics have grown at an increasing rate in recent years, as digital manipulation techniques are continually expanding and influencing different facets of our socioeconomic and political lives. Currently, several advanced methods rely on generative adversarial networks (GNNs) and autoencoders, and a large number of datasets are publicly available to train them. Deepfakes are now widespread and pose a serious threat to the accuracy of information, affecting many facets of society including the economy, journalism, and politics. As new and more complex types of manipulation emerge, these issues can only worsen. As a result, the government, research community, and various non-profit organizations have intensified their emphasis on mitigating and developing cutting-edge technologies to cope with these phenomena. To encourage research in this area Major IT corporations, such as Facebook and Google, have created massive databases of edited videos available online [1]. There

The associate editor coordinating the review of this manuscript and approving it for publication was Binit Lukose.

have also been efforts by funding agencies to encourage extensive research projects on these topics. For example, in 2006, DARPA launched the Media Forensic Initiative (MediFor) [2] to support well-known research organizations on media integrity worldwide, which produced significant results with regard to methodologies and reference datasets. Subsequently, the Semantic Forensics Initiative (SemaFor) [3] was launched in 2020, which collaborated on text, audio, images, and video to develop semantic-level detectors for fake media to combat state-of-the-art attacks.

Over the past few years, electric network frequency (ENF) has been utilized as a tool in some applications in multimedia forensics. Analysis of the ENF is a forensic tool used to validate multimedia files and spot any attempts at manipulation [4], [5], [6], [7], [8]. The ENF is an electric power grid supply frequency and varies in time about its nominal value of 60 Hz in North America, and 50 Hz in Europe, Australia, and much of the rest of the world because of the inconsistencies caused by power network supply and demand [4], [9]. The nature of these inconsistencies can be observed to be random, unique per time, and usually the same across all locations connected by the same power grid. Consequently, an ENF signal recorded at any point in time while plugging to a certain power mains may serve as a reference ENF signal for the entire region serviced by that power grid for that period of time [4], [10]. The instantaneous values of ENF over time are considered ENF signals. An ENF signal is embedded in audio files created with devices connected to the mains power or located in environments where electromagnetic interference or acoustic mains hum is present [4], [10].

Recent studies have shown that ENF signals can be extracted from video recordings made under the illumination of a main-powered light source [11], [12]. Fluorescent lights and incandestine bulbs used in indoor lighting fluctuate in light intensity at double the supply frequency, making it nearly impossible to notice the flickering that occurs in the illuminated environment. Consequently, videos captured under indoor illumination settings using a camera may contain ENF signals. Research has also recently demonstrated that an ENF can be embedded in an image captured with a rolling shutter camera [13].

The ENF was identified as a relevant resource for multimedia forensic in 2005 when Grigoris [4] first reported that digital audio recordings captured by devices plugged into the power mains contain ENF artefacts. This work was further extended in [5], where experiments were carried out in Romania. Subsequently, several experiments were conducted in other parts of Europe, such as Poland [6], Denmark [10], [14], [15], the United Kingdom [16], [17], [18], the Netherlands [8], and then in North America [9], [12], [19], and in other parts of the world [20], [21], [22]. The experiments were further carried out with devices connected to alternating current (AC) power and battery-powered devices [23]. Research in ENF has been broadly towards the development and improvement of approaches for accurately extracting the ENF signal from media recordings, as well as the application of the extracted ENF signal. A schematic showing the different stages of ENF forensic application is shown in figure 1. In this paper, we present a comprehensive survey of research in this area, identifying current challenges and providing future directions. The remainder of this paper is structured as follows: Section II describes the basic concept and methods employed in creating the ENF reference database. Section III discusses different approaches for estimating ENF. Section IV describes the ENF detection in media recordings. A review of approach approaches for extracting ENF from audio and video recordings and from a single image is provided in Section V. Section VI discusses the factors affecting ENF embedding in audio and video recordings. Section VII discusses forensic and non-forensic applications of ENF. Section VIII presents current challenges and future research directions. Section 9 concludes the paper.

II. BACKGROUND KNOWLEDGE

This section provides a background of the power generation and control mechanism of the electric power grid to aid in understanding of the ENF signal. It also discusses methods for recording the ENF signal at the power distribution level and from the power mains to build a reference ENF database.

A. OVERVIEW OF ENF

In power grids, ENF is the electricity supply frequency. Coal, geothermal, wind, solar, and other energy sources can all be used to generate this electricity. These energies are harvested and transformed into electricity before being transmitted to the power grid [24]. A schematic of power generation and control is shown in figure 2. The generators receive kinetic energy from the turbines and transform it into electrical power. The generator rotates as a result of the provided power (P^s), converting kinetic energy into electrical energy to satisfy the power demand (P^d). The angular velocity of the generator is represented by ω and is inversely proportional to the system/voltage frequency f_s, which represents the ENF [25].

The nominal value of the ENF is 50 Hz in Europe, Australia, and across many other regions of the world, and 60 Hz in the majority of Americas. The ENF is at the nominal value (50 Hz/60 Hz) when the power supply (P^s) is equal to the power demand (P^d). However, because of consumers' varying demand for power per unit time, it constantly varies with its nominal value. These variations, which are within 20 mHz of the nominal value, are due to the power-frequency control systems that maintain equilibrium between electric energy production and consumption throughout the power grid [26]. The nature of the variations may be observed to be random and unique over time, and usually quite the same across all locations connected by the same power grid [18]. Because of this type of power generation, the electric signal of the power grid is modeled as a single sinusoidal waveform with a set frequency, as shown in the following equation:

$$v(t) = A_o Sin(2\pi ft + \emptyset) \tag{1}$$

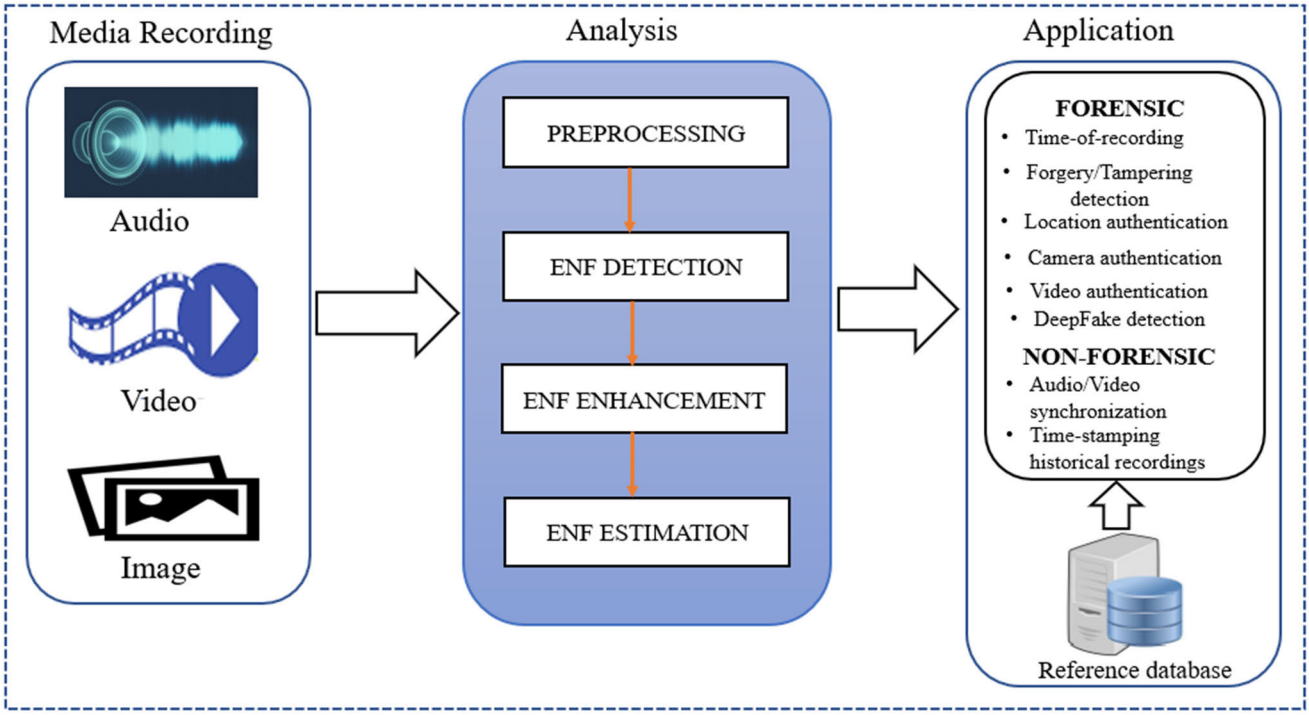


FIGURE 1. Schematic of steps in ENF analysis.

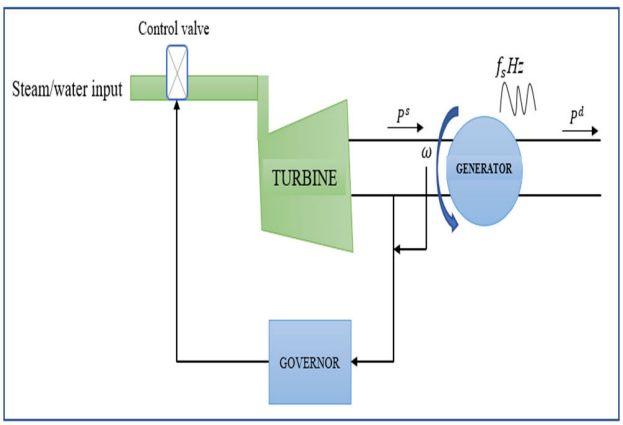


FIGURE 2. Schematics of power generation and control [26].

where A_o and $\emptyset$ represent respectively the amplitude and phase while f denotes the ENF of the grid. Therefore, ENF is defined as the change in instantaneous ENF values over time.

A successful extraction of this unseen signature from a media recording will enable the determination of the time and location where it was made and if any manipulation had occurred.

B. ENF REFERENCE DATA

The ENF signals retrieved from multimedia files may be vital in a variety of practical real-world forensic applications. However, for us to rely on the outcome of its application, we must verify that the extracted signal is actually the ENF signal. Therefore, the ENF criterion was based on the ENF reference data. Usually, power reference signals are obtained using special equipment and are substantially more potent than ENF traces extracted from multimedia. Consequently, they can serve as a reference for the extracted ENF signal from multimedia recordings [4], [27]. Therefore, the ENF signals of multimedia and power grids recorded simultaneously are anticipated to be similar at the same time instants.

Several techniques for acquiring power reference data have been suggested in the literature. The authors in [28] and [29] presented a detailed discussion of the implementation of the wide-area frequency monitoring network (FNET)/GridEye system, a situational awareness tool for an electrical grid that captures high-precision Global Positioning System (GPS) time-stamped data at the distribution level in real time. The FNET has been implemented and utilized to capture the power reference signal from the three North American interconnections [30]: Texas power grid, Eastern power grid, Western power grid.

Figure 3 shows the different components of the FNET/GridEye framework. The frequency disturbance recorder (FDRs) and the associated information management software (IMS) are the main elements of the FNET system. FDRs are system sensors deployed at the distribution level to measure regional GPS-synchronized instantaneous ENF values using phasor approaches [31]. FDRs transmit the measured data via the Internet to the data center, which

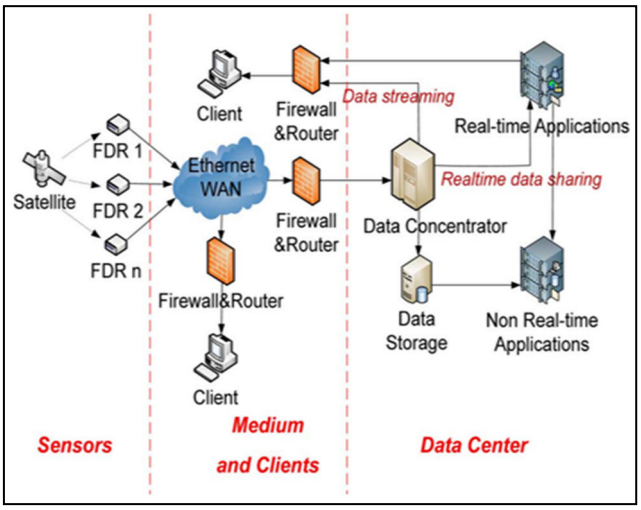


FIGURE 3. Building blocks of the FNET/GridEye system [28].

is managed by the IMS for various applications. More discussion on the FNET/GridEye system can be found in [32], [33], and [34].

The authors in [21] and [35] provided a detailed discussion and analysis of the wide area management system (WAMS) deployed in the Egyptian grid for wide-area synchronized measurements. Contrary to the definition that the ENF signal is similar in all areas of the same grid, the authors opined that the instantaneous ENF value recorded from a single point is not constant throughout the grid when there are system disturbances such as transformer or local network failure and disconnection of unsynchronized local sections of the grid [14]. The authors proposed a robust method that relies on creating an ENF reference database from many FDRs spread across multiple grid points instead of one point. Their choice of the number of sensors to deploy was based on their sensitivity to frequency, as well as the frequency sensor's estimation accuracy. They also suggested a harmony search technique that uses GIS data and wide-area frequency measurements to determine geographically coherent frequency regions for various disturbance situations and allocated the sensors to different regions based on their geographical frequency coherence. Their experimental results revealed that the proposed approach enables the building of a robust ENF reference database and, in turn, improves the accuracy of the matching process.

The authors in [36] considered the methods discussed above expensive, involving a substantial amount of work from design to setup, and lack efficiency to adequately monitor the collection equipment dispersed across the world. They therefore proposed a method to build worldwide ENF map by extracting ENF signal from online streaming multimedia files gathered from sources like "Ustream," "Youtube," and "Earthcam," rather than installing expensive specialized hardware. Their approach employed various signal processing methods to discuss and address several issues, such as accounting for packet loss, aligning various ENF signals from several multimedia broadcasts, and interpolating geographic ENF signals to take consider areas that are not serviced by streaming services. The evaluation of the recovered ENF signal using the proposed method compared to those obtained using the FDR from FNET/GridEye demonstrates that the proposed method performs better in steady acquisition and control of the ENF signals than the conventional approach.

Other systems have been implemented to locally build an ENF reference database. Although the systems discussed above provide enormous advantages in terms of power-frequency monitoring and coverage, researchers can locally obtain ENF references without requiring access to them. A low-cost hardware circuit can be implemented and plugged into an electric wall outlet to capture a power signal or detect ENF fluctuations. In such systems, transformers, such as those utilized in the power supplies of several DC-powered equipment, are usually employed to convert the wall outlet voltage level to a level that can be captured by an analog-to-digital converter (ADC). An example of a generic circuit that can be implemented to record the power reference is depicted in figure 4. The circuit also included an aliasing filter and a fuse to the circuit as a safety measure, subject to the sampling rate of the ADC [37].

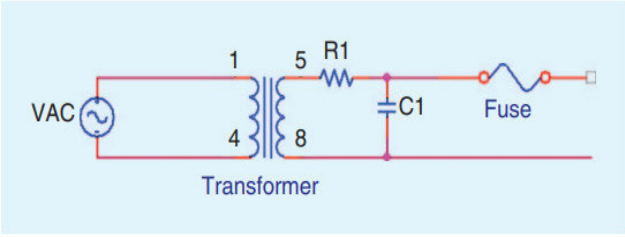


FIGURE 4. A sample generic diagram of sensor hardware [37].

The sensor hardware may be implemented using a variety of different designs.

The authors in [5] implemented an electronic circuit to record the power reference signals. The circuit relies on a transformer capable of producing a 100mV peak-to-peak voltage connected to a personal computer's soundcard. A software system called DCLive Forensics was used to record and analyze the audio signal. The device was used to obtain the ENF reference signal from the Romania electric network and other locations in the European electric network. In the implementation in [38], a step-down circuit was connected a digital audio recorder to obtain the raw power signal in Maryland, part of the US Eastern grid. In this case, an ENF estimation method is employed to further process the obtained signal to retrieve the ENF reference signal.

The authors in [39] implemented a step-down circuitry, together with a Schmitt trigger on a BeagleBone Black board, which applies the Schmitt trigger to spontaneously retrieve the ENF signal from the raw ENF data without requiring an ENF extraction method, as in the implementation in [38]. The

device was used to independently record ENF reference data in Dresden, Germany. Other implementations can be found in [40], [41], and [42]. Depending on the available resources, there are various approaches to obtain the ENF reference. These measurements often yield an ENF signal with a high signal-to-noise ratio (SNR), which can be utilized in ENF investigations as reference data.

III. APPROACHES TO ESTIMATING ENF SIGNAL

This section discusses the procedures employed in the literature to extract ENF traces implanted in media recordings. In these approaches, preprocessing of the media signal may be necessary prior to measuring the varying instantaneous ENF sequences over time to significantly speed up the estimation process and conserve memory for the estimation algorithm. In general, there are three main categories of approaches used for estimating the ENF which: (a) time-domain approach, (b) non-parametric approach and (c) parametric approach.

A. TIME-DOMAIN APPROACH

The time-domain approach consists of the measurement of zero crossings and comparison of the results to the reference ENF values [5]. The time-domain zero-crossing method is relatively simple and does not require segmentation of the recording into successive frames for independent processing [43]. It relies on deriving ENF variations from the zero crosses of a genuine power reference or media recording. [44]. The authors in [45] reported that with the zero-crossing strategy, the ENF-containing signal was pre-filtered around the nominal range to isolate the ENF component from the other contents of the recording. Subsequently, the temporal differences between successive zero values were derived and utilized to realize instantaneous ENF estimation after the zero crossings of the surviving ENF signal were computed.

B. NON-PARAMETRIC APPROACH TO FREQUENCY-DOMAIN ENF ESTIMATION

ENF estimation based on non-parametric methods does not make explicit model assumptions about the data. The majority of these methods rely on the Fourier analysis of the signal. The time recursive adaptive approach (TR-IAA) and spectrogram- or periodogram-based methods that employ the short-time Fourier transform (STFT) comprise the majority of non-parametric frequency-domain methods [46], [47]. The STFT is typically employed for time-varying spectrum signals, such as ENF signals. The signal is first segmented into overlapping frames, and each frame is then subjected to Fourier analysis to ascertain the available frequencies. Therefore, a spectrogram provides a visual representation (2-D representation) of the time-frequency information provided by the STFT. Time-frequency information is usually conveyed using a heat map, with time and frequency as the axes [48]. Because the ENF signal fluctuates very slowly, it is logical to assume that the instantaneous frequency is relatively consistent throughout the frame period for analysis [43].

1) MAXIMUM ENERGY APPROACH

Considering the ENF signal in a frame as a frequency sinusoid near the nominal value embedded in noise, applying the spectrogram method should ideally peak its power spectral density (PSD) at the sinusoidal signal frequency. In this situation, finding the frequency whose spectral component has the maximum power is a simple way to estimate this frequency. However, directly adopting this frequency as the ENF value usually results in a loss of accuracy because the spectrum is computed for a discretized frequency value, and there might be a misalignment of the exact frequency of the maximum energy and the discretized frequency. Consequently, the STFT-based ENF estimation method often employs quadratic interpolation [49] or a weighted approach [12] to perform additional operations to achieve a more precise estimation. Quadratic interpolation is usually performed around the identified spectral peak, while the weighted approach is applied to the detected ENF estimate, in which the frequency bins of the nominal value are weighted according to their spectrum intensities [43]. Examples of spectrogram-based approaches includes the maximum-energy approach and the weighted-energy approach.

2) WEIGHTED ENERGY APPROACH

This approach involves calculating the average frequency of each spectrogram's time bin, considering the weighting of the frequency bins based on their energy levels relative to the nominal ENF value. This results in an average frequency value that is representative of the energy of the signal. The expression of the frequency estimation as described by [46] is given as:

$$F(m) = \frac{\sum_{k=K_1}^{K_2} f(m,k)\, S(m,k)}{\sum_{k=K_1}^{K_2} S(m,k)} \tag{2}$$

where K_1 and K_2 denote the FFT indices of the averaging region's boundary, $f(m,k)$ and $S(m,k)$ represents, respectively, the frequency and energy values for the k^{th} frequency bin of the m^{th} time-frame. Compared to the maximum energy approach, this approach estimates the instantaneous ENF frequencies more accurately because of the robustness of the weighting against outliers.

3) TIME RECURSIVE ITERATIVE ADAPTIVE APPROACH (TR-IAA)

The most recently developed non-parametric frequency estimation approach relies on the TR-IAA [50]. To obtain the spectral estimate for a particular frame, the algorithm formulates a weighted least-squares method to minimize the quadratic function. The TR-IAA is an iterative approach with a convergence period of 10 -15 iterations. The spectral estimate in the first iteration was set to the value from either the spectrogram or the last value from the time frame before

it [51]. This approach requires a longer processing time than the spectrogram-based approach. A quadratic interpolation procedure was applied after the spectral estimate converged to estimate frequency. Compared with the spectrogram-based approach, this approach has been demonstrated to yield somewhat better frequency estimation results, particularly when the frame size is between 20 and 30s [50].

C. PARAMETRIC APPROACH TO FREQUENCY-DOMAIN ENF ESTIMATION

Parametric approaches to frequency-domain ENF estimation are based on making specific model assumptions about the signal and the underlying noise. These approaches produce more precise estimations than non-parametric approaches because of the explicit model assumptions on which they rely. In fact, modeling the signal provides a means of avoiding the idea that the recorded sequence has a value of zero beyond the time period in which it was observed. This makes the resolution dependent on the SNR such that increasing the SNR results in an improved resolution threshold [52], [53]. The MUltiple SIgnal Classification (MUSIC) [54] and Estimation of Signal Parameters via Rotational Invariance Techniques (ESPRIT) [55] are two of the most commonly used parametric approaches that rely on the subspace analysis of a model that includes both the signal and noise components. In principle, these methods can be utilized for the frequency estimation of a signal composed of P complex frequency sinusoids contained in white noise. The value of P for ENF signals is equal to 2 because there is only a single actual sinusoid in an ENF signal [46]. A brief discussion of these methods is provided below.

1) MUSIC

The MUSIC algorithm estimates the essential characteristics of a signal based on the observations provided [54]. This approach assumes that a signal x is composed of p complex sinusoidal components with an unknown angular frequency ω embedded in additive white Gaussian noise. It conducts an eigenspace decomposition of the autocorrelation matrix $R_{xx}^{M\times M}$ based on M observations of signal x, in such a way that it projects the p complex sinusoids onto a signal space that is perpendicular to the noise space. Such a projection may be generated by performing Singular Value Decomposition (SVD) on the $R_{xx}^{M\times M}$ for which the highest eigenvalues correspond to the signal sub-space eigenvectors [54], [56]. The signal spectrum can then be obtained from the noise subspace as follows:

$$P_{mu}\left(e^{j\omega}\right) = \frac{1}{\sum_{k=p+1}^{M} |e^H w_k|^2} \tag{3}$$

where $e = [1 e^{j\omega} e^{j2\omega} \ldots e^{j(M-1)\omega}]$ is a vector that contains M complex sinusoids, e^H denotes the Hermitian operator, w_k is the eigenvector of $R_{xx}^{M\times M}$, and xx is the vector matrix of the signal x.

Compared to the Fourier-based analysis, the MUSIC approach offers a better frequency estimation. The spectrum in equation (3) is a continuously differentiable function in ω that allows for some flexibility because the frequency resolution may be raised to counteract the binning problem, which affects the Discrete Fourier Transform (DFT). It can also intrinsically reduce the level of additive white noise in the measured signal. However, these features come at the expense of reduced stability when there is a drop in the SNR or when the noise stops becoming addictive and white [54]. The MUSIC method may also lead to increased computational complexity when the autocorrelation matrices and the number of frequency points are large. A variant of the MUSIC method called Root MUSIC [57], [58] provides a high-resolution estimation of instantaneous frequency while significantly conserving computational resources.

2) ESPRIT

The ESPRIT approach employs invoked rotational attributes between staggered subspaces to provide frequency estimation. ESPRIT shares similarities with MUSIC, as both approaches rely on subspace analysis. However, ESPRIT differs in that it operates in the signal subspace instead of the noise subspace [55]. In addition, ESPRIT relies on the information in the data matrix to compute the signal subspace, whereas MUSIC explicitly computes the correlation matrix. ESPRIT, like MUSIC, offers a robust estimation of the signal frequency utilizing fewer data points compared to spectrogram-based approaches. Compared with MUSIC, ESPRIT has the advantage of lower computation and storage costs [59].

IV. DETECTING ENF PRESENCE IN DIGITAL MULTIMEDIA RECORDING

Detecting the presence of ENF in digital multimedia recordings is a crucial and fundamental issue that must be resolved to conduct a successful ENF-based forensic investigation process with confidence. The ENF traces can be obtained from media recordings and reflect the way the power network operates during recording. Therefore, it is essential that there is electrical activity present at the location where the recording is made in order to acquire ENF traces. It is commonly accepted that audio recordings created by recorders connected to a power outlet will contain ENF traces owing to the electromagnetic interference produced by the connection between the recorder and the wall outlet [8], [39]. However, when it comes to audio recordings captured with recorders that use battery as a source of power, the situation becomes more complicated because of several factors. For video recordings, studies [11], [12] have shown that ENF can be picked up as time-varying light intensity by multimedia recordings made with cameras, resulting in unseen flickering generated by electric-powered interior lighting. Nevertheless, the ENF signal that is present in audio/video content may be destroyed by strong Doppler effects as a result of normal recording device movement, because the nominal frequency (together with a number of harmonic frequencies) exists in the lower frequency band [60], [61].

We noted that in an audio recording, the intensity of the ENF signal and the square of the distance between the microphone and the nearest ENF source are inversely proportional, whereas in video recording, it is normally rare to observe the flickering of the light intensity [60], [61]. Meanwhile, even when the recording is done near a mains power, the ENF hum's intensity is consistently much weaker compared to the recorded material of interest [62]. Therefore, we can infer that the recorded ENF signal is on average quite weak. It is irrelevant to adopt an ENF-based forensic analysis if the file under investigation is severely distorted by interference, noise, and Doppler effects. It is even more dangerous to progress the analysis assuming that the ENF signal was successfully captured, as this will negatively impact on the related investigation. Therefore, before performing further forensic studies based on ENF, it is necessary to ascertain whether it is captured in the media file under investigation. This step will aid in checking if ENF is present and then inform the progress analysis. Several approaches have been proposed for ENF detection.

The authors of [63] presented a comprehensive examination of the ENF detection problem in audio recording and proposed three practical detectors to address these problems. In their approach, the audio file under examination is down-sampled and bandpass filtered in the preprocessing step, which removes duplicate computational overheads and subdues the interference and out-of-band noise, respectively. Consequently, the detection of the ENF signal within an audio file was explained by identifying a delicate ENF signal that has been corrupted by unknown WSS Gaussian noise with varying colors. In general, this problem is challenging owing to the frail nature of ENF, the random behavior of the amplitudes and instantaneous frequencies, and the initial phase of the ENF signal. Supposing prior information about the ENF signal, the authors developed three Neyman-Pearson (NP) detectors, including a matched filter (MF)-like detector, a general matched filter (GMF), and the asymptotic approximation of the GMF to address this problem. Supposing the unknown parameters of the ENF are constants, the authors suggested two least-squares (LS)-based time domain-detectors known as LS-LRT and naive-LRT. Additionally, assuming that the ENF is steady over a short time interval, the authors suggested a time-frequency (TF) domain detector to conduct short-time Fourier transform-based TF analysis of the audio file utilizing the ENF's prior knowledge obtained from public reference ENF databases. The performance of the proposed detectors was comprehensively evaluated in relation to the test statistic distribution, threshold selection, and computational complexity. The authors reported that the decision thresholds for LS-LRT, naive-LRT, and TF detectors could be efficiently established without knowing the noise parameter. The results of experiments using synthetic and actual audio recordings revealed that the LS-LRT and TF detectors produced highly competitive detection outcomes for normal and lengthy recordings, whereas the naïve LRT detector surpassed the others for extremely small recordings. Their results further demonstrated that estimating ENF using the nominal value is an effective remedy for ENF detection in extremely short recordings.

Liao et al. [64] suggested a multi-tone harmonic combining (MHC) approach that exploits the harmonic components of the ENF to improve the single-tone TF domain ENF detector presented in [64]. Their approach first eliminated contaminated harmonics that may interfere with ENF detection using an enhanced sub-band SNR estimator. Then, the test statistics (TS) of the TF detector were applied to further screen the selected harmonic candidates. Subsequently, the MHC mechanism was applied to combine the harmonic components and generate the TS function used to make the final ENF detection decision. The suggested multi harmonic TS may also be used to measure the quantity and quality of the ENF harmonics that can be utilized. Both the results of the simulation and real-world experiment revealed that the MHC detector achieved improved detection precision compared to the single-tone DF detector.

The authors in [65] introduced a detector for ENF presence that can identify the presence of an ENF signal in a digital video recording. This is accomplished by conducting multiple ENF estimations on stable regions of the video, known as super pixels. The detector computes the average of steady super-pixels instead of all the steady pixels contained in a frame. Each video was segmented into super-pixel regions, and the steady points within each region were identified throughout all frames. The steady pixels within all regions of every frame of the video were combined to generate an intensity signal, from which an ENF vector was computed for each successive video frame in a particular shot. The next step involved analyzing the similarity between the computed ENF vectors to detect whether the ENF was present or absent in the test video. The detector was applied to a video dataset of 160 videos of different lengths recorded under different conditions using cameras that adopt both metal oxide semiconductor (CMOS) and charged-couple device (CCD) sensors, and the findings of the experiments showed that the proposed detector is capable of detecting ENF signals in videos even as short as 2-minutes video clips. The detector can be useful in cases where a set of data on a disc or social media is subjected to ENF-based forensic examination to detect and distinguish ENF-containing videos from ENF-free videos before progressing analysis. In doing so, unnecessary exposure of ENF-free videos to a full ENF-based analysis can be avoided, thereby saving both time and computational effort.

Recently, the authors in [13] and [66] suggested that it is possible to detect and retrieve ENF from a single image captured by rolling shutter cameras under electric light. They showed that the sequential read-out time mechanism of a rolling shutter camera provides the resultant image to pick up instances of electric light signal entering at slightly various points in time, resulting in a brief section of ENF patterns.

Section V presents a detailed discussion of the ENF within a single image.

V. A REVIEW OF RESARCH ON ENF EXTRACTION FROM AUDIO AND VIDEO RECORDING

This section provides an overview of the methods that have been suggested for extracting the ENF signal embedded into audio and video recordings and from a single image.

A. ENF EXTRACTION FROM AUDIO RECORDING

The use of ENF in audio forensic analyses has been widely studied. The crucial initial stage of this process involves successfully extracting the ENF pattern from the audio recordings under investigation. Researchers have proposed several parametric and nonparametric techniques to estimate and extract the ENF signals. Early approaches to ENF extraction from audio recordings were based on the STFT. The questioned signal is first partitioned into overlapping frames, after which Fourier analysis is performed on each frame to ascertain the available frequencies. Next, the instantaneous frequency (IF) of every frame was combined to create an estimate of the ENF [7], [16], [29], [45]. In addition to the widely used STFT-based approach, substantial research efforts have focused on more accurate ENF estimates.

In [50], the authors introduced a technique called the time-recursive iterative adaptive approach (TR-IAA), which is a non-parametric, high-resolution algorithm that is also adaptive. To obtain the spectral estimate for a particular frame, the algorithm formulates a weighted least-squares method to minimize the quadratic function. The TR-IAA is an iterative approach with a convergence period of 10 -15 iterations. The spectral estimate in the first iteration was set by utilizing the value obtained from either the spectrogram or the last value from the time frame before it [67]. This approach requires a longer processing time in contrast to the FFT-based approach. According to the authors, the FFT method achieved a slightly more precise estimation of the network frequency when dealing with the high SNR produced by the first dataset. Nonetheless, in the second dataset, which contained other strong interfering signals, the proposed IAA obtained a higher ENF estimation precision. To enhance the precision of their ENF extraction, the authors formulated a frequency-tracking algorithm rooted in discrete dynamic programming [68], which searches for the path of minimum cost. The approach creates a list of potential frequency peak locations for each frame, and then builds the path with the lowest cost. A cost function is chosen considering the slowly varying nature of the ENF and penalizes large frequency variations between frames. The estimated ENF was determined using the lowest path.

In situations where there is a low SNR, it may be difficult to accurately determine the possible locations of the frequency peaks utilized by the authors in [50]. To circumvent the issue of uncertain peak locations, the authors in [69] and [70] proposed a novel weak frequency component detection and tracking algorithm for conditions with very low SNR and in nearly real-time, termed Adaptive Multi-Trace Carving (AMTC). The process involves using iterative dynamic programming together with adaptive trace compensation on a preprocessed output from a system, such as a spectrogram, to detect frequency traces. The proposed algorithm uses fairly high energy traces that persist for a particular amount of time to prove the existence of the desired frequency components of interest after several forward and backward passes. Experimental results utilizing both synthetic and actual forensic data power signatures reveal that under low SNR conditions, the proposed technique performs better than other representative earlier methods [50] and can be applied in near real-time environments.

Other adaptations to non-parametric approaches have been suggested in previous studies to improve the estimation of the final ENF signal. In [71], the authors presented the idea of computing precise spectral lines at a specific frequency using a DFT algorithm rather than across the entire frequency band. The spectral line computation was then repeated for each target frequency bin using a binary search technique until the concealed ENF signal was extracted. An experimental study of the algorithm using real audio recording and simulated audio signals with various SNRs and error-evaluation criteria proved its effective performance with regard to accuracy and precision. However, SNR significantly influences the performance of the suggested algorithm. A large window length is required under low-SNR conditions, which may weaken its ability to track frequency fluctuations in the time domain.

A frequency-extraction problem was constructed by the author in [72] as a frequency-modulation problem. The author postulated that instead of measuring the ENF directly, it should be considered as a sinusoidal signal at the nominal power network frequency that is subject to frequency modulation by a frail frequency, making it possible to create and analyze an intermediate-frequency signal with a frequency of 0 Hz. According to the authors, this provides substantial data reduction through comprehensive down-sampling to enable the utilization of FM demodulation algorithms to extract the ENF.

In [73], the authors suggested a method that utilizes temporal windowing and a filter-bank carpon spectral estimator. Their method proposed building non-parametric frequency estimation techniques on top of the refined periodograms. Their studies also showed that selecting proper windows can offer an enhanced spectral resolution and improve frequency estimation precision. The superiority of the method over [50] is in its ability to achieve high accuracy even in the recording of 1s frame length.

An approach that combines the Blackman-Tukey spectral estimation method with a modified lag window design was introduced by the authors of [74]. Such a lag window design ensures an accurate ENF calculation under various SNR situations, while striking a balance between minimizing smearing and leakage. They formulated leakage reduction as an energy-maximization problem within the main lobe of the spectral window. The ENF estimation accuracy was

measured using the maximum correlation coefficient (MCC) and minimum standard deviation of the errors. A test of the proposed approach on real-world datasets revealed that it outperformed many existing techniques.

The authors in [75], presented a method that leverages robust principal component analysis to exploit the low-rank nature of the ENF signal, enabling the elimination of speech content interference as well as background noise. This approach enhances the extraction of the ENF estimates, particularly in situations where the SNR is low. ENF estimates were then extracted by enforcing a weighted linear prediction. The authors reported that a small number of signal observations are required for the proposed method to successfully capture ENF variations along the time axis while maintaining adequate frequency precision.

The robustness of ENF signal estimation may be improved by exploiting the existence of ENF traces at multiple harmonics when extracting the ENF signal. In [76], the authors extended a model based on single-tone harmonics to a multi-tone ENF model by applying the maximum-likelihood estimator (MLE). The model employed the Cramer-Rao bound (CRB) to estimate the variance of the ENF Estimator and demonstrated that the proposed model can achieve a theoretical improvement factor of $O(M^3)$ in the estimation accuracy, where M represents the number of harmonics. The experimental findings indicated that the proposed model outperformed estimators based on a single-tone.

In [38], the authors proposed a spectrum combining method that adaptively combines the various ENF components that are present at different harmonics in a signal using the local SNR at each harmonic. Compared with estimates produced using only one component, the technique produced estimates that were more reliable and accurate.

In [77], the authors presented a technique based on the linear canonical transform (LCT) to estimate ENF audio files recorded in a complex noise environment, such as multipath interference. The authors argued that the method proposed in [76] may not be adequate in a noisy network such as China's electrical network, because it did not account for unpredictability in delays and amplitudes. They modeled the electrical signal from such a complex and noisy network as a signal with variable amplitude and narrow bandwidth and applied the LCT, adjusting the coefficient suitably to transform the signal into a frequency spectrum characterized by a sharp pulse.

In [78], the authors presented a strong filter technique called the robust filtering algorithm (RFA) to enhance the ENF signal estimation in audio recording. The RFA accepts as input a preprocessed (down-sampled and bandpass filtered) recording and produces a denoised ENF signal, which is then extracted and further analyzed. RFA achieves its denoising operation by encoding the time-domain expression of the preprocessed audio signal into the instantaneous frequency of an analytical sinusoidal frequency modulated (SFM) signal. A kernel function is then formulated to create a sinusoidal time-frequency distribution (STFD) of the encoded signal, where the STFD peaks correspond to the ENF signal that has been purified from noise. RFA converges within two or three iterations in a single-tone harmonic formulation. The authors noted that this algorithm could be integrated with any frequency estimation method to improve the performance of the estimator by increasing the SNR.

In [79], the authors proposed a robust ENF extraction framework that included an improvement over the single-tone algorithm proposed in [78]. First, the authors introduced a harmonic robust filtering algorithm (HRFA) to extend single-tone RFA to a multi-tone scenario. The HRFA enhances each harmonic element individually without any unwanted interference from other elements, thereby reducing the impact of undesired noise and audio material. Considering that certain harmonic elements may still be distorted despite the use of HRFA, the authors introduced a graph-based harmonic selection algorithm (GHSA) that selects a group of harmonic elements based on their strong correlations with each other. The problem of selecting the most suitable harmonic element was framed as a maximum-weight clique problem, and the authors employed the Bron-Kerbosch algorithm to effectively solve it. According to the authors, the proposed framework surpasses the current single- and multi-tone techniques when used in combination with modern MLE to estimate the ENF. The superiority of the proposed framework was demonstrated through an experimental evaluation using 130 real-world audio recordings from the ENF-WHU dataset.

B. ENF EXTRACTION FROM VIDEO RECORDING

Recent research has focused on the difficult problem of recovering ENF signals from visual content. The fluctuation in light intensity over time included inside the video frames may be leveraged to obtain the ENF signal. Oscillations in the ENF in the grid network influence the brightness of the light emitted by any light source linked to the power grid. Because the light source flickers during both positive and negative AC current cycles, the frequency of the light is twice that of the main power frequency. Thus, the light signal may be considered as an absolute representation of the cosine wave function [65]. For instance, for any video recording made under indoor illumination powered by 50 Hz power mains, because the polarity of the current changes at double the frequency of the main power, the light flickers at 100Hz.

In addition, higher harmonics of decaying energy frequently exist at integer multiples of 100Hz when there is a mild deviation from a perfect sinusoidal mains power signal. In addition, the higher harmonics have a larger bandwidth than the primary component because the actual ENF signal, which is a narrowband signal rather than a completely stable sine wave, is the signal of interest. Consequently, the n^{th}-harmonic component bandwidth is n times the ENF component bandwidth at 100 Hz. Therefore, according to the Nyquist theorem, it is necessary to have a sampling frequency of at least 200 Hz to reliably detect and extract the frequency of illumination from the recorded data [65].

TABLE 1. A summary of research in ENF estimation in audio recordings.

Ref	Type of Approach	Focus
[7], [16], [29], [47]	Non-parametric	Performing Fourier analysis on each frame of partitioned frames that overlap to ascertain the available frequencies, and combining the IF of every frame to create an estimate of the ENF
[50]	Non-parametric, frequency tracking algorithm	Applying an iterative adaptive approach together with dynamic programming to frequency tracking. Offered high resolution but at high cost of computational complexity.
[70]	Non-parametric	Applying adaptive trace compensation and iterative dynamic programming to identify various weak frequency components under low SNR situations.
[71]	Non-parametric	Achieving high-resolution frequency estimation by considering particular spectral lines rather than the whole frequency band.
[72]	Non-parametric	Modeling the ENF signal as a modulation and demodulation process to considerably reduce the impact of noise and interference.
[73]	Non-parametric, filter-bank Capon spectral estimator	Achieving accurate ENF estimation using fast Capon spectral estimator that relies on Gohberg-Semencul factorization as well as Parzen temporal window.
[74]	Non-parametric, Blackman–Tukey spectral estimator.	Achieving low computational requirement and accurate ENF estimation employing Blackman-Turkey approach that incorporates a lag window that provides an ideal trade-off between smearing and leakage.
[75]	Non-parametric	Ensuring accurate ENF estimation utilizing robust principal component analysis to mitigate noise interference.
[76]	Parametric	Combining the ENF at multiple harmonics to offer a more accurate ENF signal estimation
[38]	Non-parametric and parametric	Combining the ENF at multiple harmonics to offer a more accurate ENF signal estimation
[77]	Non-parametric, linear canonical transform	Mitigating complex noise such as the multipath interference based on linear canonical transform
[78], [79]	Non-parametric, RFA, HRFA	Enhancing the ENF signal relying on harmonic filtering in a low SNR situation.
[46]	Non-parametric and parametric	Attaining high performance while minimizing computational complexity, by utilizing the minimum number of samples per frame feasible

Although the majority of consumer cameras cannot offer such a high frame sampling rate, the illumination frequency can still be estimated from its aliased frequency. Assuming that f_s is the sampling frequency of the camera and f_l is the frequency of the illumination of the light source, the aliased illumination frequency f_a is expressed as [80]:

$$f_a = |f_\ell - j.f_s| < \frac{f_s}{2}, \quad \exists j \in N, \tag{4}$$

where N is the number of samples/observation. Therefore, when a 100Hz illumination signal from a light source is sampled using a camera with a frame-rate of 29.97Hz, the aliased base frequency of the ENF is obtained as 10.09 Hz while the aliased second harmonic is obtained as 9.79 Hz [12]. The aliased effect resulting from camera's insufficient sampling rate is mainly associated with cameras with the global shutter mechanism.

The sampling rates of widely accessible cameras were significantly lower than the nominal ENF value. Common commercially available video cameras come at three sampling rates (fps): 24, 25, and 30 fps. Cameras with 24 fps are utilized for film production, while majority of amateur hand-held cameras have 25 or 30 fps. Cameras with a sampling rate of 25 fps are common in Europe and the majority of Asia and are invented by the PAL analog TV standard, while 30 fps cameras are invented by the NTSC standard and are common in Japan and North America. Most video cameras have a frame rate that varies significantly from 25 to 30 fps. A frame rate of 24.98 is specified by the PAL standard whereas a frame rate of 29.97 frames per second is required by the NTSC standard.

The approaches for estimating the ENF in videos depend on the type of sensor used to capture the video. There are two types of sensors: charged-couple device (CCD) sensors, which utilize a global shuttle mechanism, and the metal oxide semiconductor (CMOS) sensors, which utilize a rolling shutter mechanism. Two types of camera images (CCD and CMOS) can capture the flicker caused by the ENF at either the level of individual frames or rows. This section provides a discussion of the two sensor mechanisms and the research studies that utilize them.

1) VIDEO RECORDING BASED ON GLOBAL SHUTTER MECHANISM

Typically, CCD cameras capture video using global shutters. It captures video by exposing and reading all the pixels of the frame at the same time. The inadequate sampling rate of

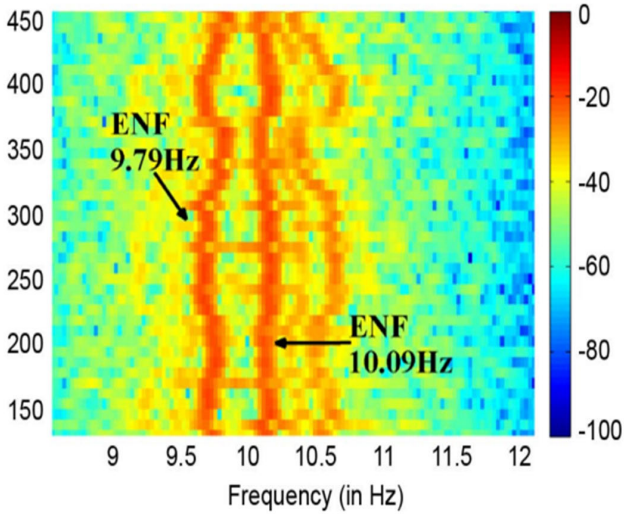


FIGURE 5. ENF fluctuation measured for white wall video showing the aliased ENF signal [12].

this particular camera type causes a problem when attempting to estimate the ENF because of the aliasing effect it induces. However, the aliasing effect has no consequence on the ENF extraction, provided that the frame rate used for sampling is not a divisor of the ENF nominal value or its harmonics. In such a situation, the aliased ENF is obtained at 0 Hz (DC component), which makes it extremely difficult to estimate the ENF.

The authors in [12] conducted the first experiment to demonstrate that the ENF signal can be extracted from a video recording. They utilized a charge-coupled device (CCD) camera to capture two recordings. The first was a video of a white-wall that was lit by fluorescent light. The second was a recording in a room with varying camera positions for surveillance purposes, with occasional movement of people in the foreground. The authors extracted the ENF signal by calculating the average intensity of each frame of the white- wall video to produce an intensity signal. The signal is then run over time through a temporal bandpass filter whose passband matches the desired frequency to extract the ENF. Given that the video content is largely consistent between frames, it was concluded that the ENF signal was responsible for the considerable amount of energy detected in the frequency of interest. The aliased ENF signals extracted from the white-wall video are presented in figure 5. In their second experiment with video recording with movement, directly averaging the pixel values of the entire frame may not be an appropriate preprocessing step prior to carrying out frequency analysis because of the inconsistency in the content of each frame of the video. However, the authors averaged the pixel intensity of relatively steady regions in the video where there were no much inconsistencies and extracted the ENF signal from the averaged pixel intensity spectrogram. The extracted ENF was utilized for time-of-recording and tampering detection applications.

The recordings used in the experiment in [12] were captured in China and India where there is a 100 Hz illumination frequency using a sampling rate of 30 fps. Consequently, there was an aliased component at 10 Hz. ENF extraction is unaffected by this blend of the ENF nominal value and the frame rate of the CCD. However, in countries where the ENF frequency is 120-Hz, such as the US, global shutter videos recorded at a sampling rate of 30fps can pose critical challenges for estimating the ENF.

Researchers have also employed image-segmentation approaches to analyze videos for ENF analysis. The authors in [81] introduced a technique for image segmentation that uses Simple Linear Iterative Clustering (SLIC) [82], [83]. To generate the mean intensity time series, they employed the SLIC algorithm to create regions of similar properties, termed super-pixels, with an average intensity above a certain threshold. According to the authors, the embedded ENF in those regions is not impeded by interference and noise including shadows, textures, and brightness, resulting in more precise estimations irrespective of whether a still or non-still test video is used. The method was applied to a public dataset of static and non-static CCD videos [84], and the mean intensity time-series signal generated was passed through an ESPRIT or STFT method to estimate the ENF.

Another method based on averaging pixels of particular characteristics for extracting ENF from non-still videos was proposed by the author in [85]. The method first employs a background subtraction algorithm named ViBe [86] to mitigate the deviation caused by movement. Before averaging the pixels, a differentiator filter was used at the pixel level to eliminate the time-dependent mean value at each point. Luminance differences beyond a certain threshold were suppressed, requiring the processing of only valid pixels of successive frames. The frame-level signal was passed to a Phase-Locked Loop-based FM demodulator [87] after preprocessing to extract the ENF in an autoregressive manner. This method was used to determine the time when some CCD videos were recorded, and the simulation results showed its effective performance. However, this method may be insufficient for extracting the ENF if the variations caused by movement affect a significant number of pixels in every frame.

2) VIDEO RECORDING BASED ON ROLLING SHUTTER MECHANISM

CMOS cameras capture images based on a rolling shutter mechanism [88]. Unlike global shutters that capture the entire frame of an image at a single instant, rolling shutters sense each frame by scanning across it, row by row, in a vertical or horizontal manner. The frame rows are exposed to light and read out one at a time, after which there is an idle time before moving on to the next frame. As a result, embedding of the ENF is achieved when the intensity of each frame's rows is sequentially captured. Figure 6 depicts this process. Because of the exposure of pixels in different

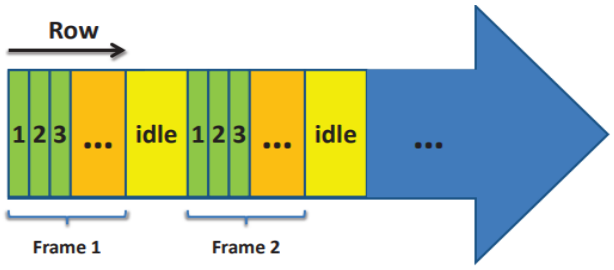


FIGURE 6. Sampling timing of rolling shutters [91].

rows at different times and their simultaneous display during playback, the rolling shutter can generate skew, blur, and other visual distortions, particularly with speedy objects and quick flashes of light [89]. As a result of the associated distortions, the rolling shutter's sequential read-out process has long been considered unfavorable to image/video quality. Conversely, recent research has demonstrated that techniques such as computational photography and computer vision can be employed to take advantage of rolling shutters [88], [90]. The ENF signal can be sensed faster by the rolling shutter than by the global shutter because of its sequential manner of capturing the rows of each frame. Although it is traditionally seen to be unfavorable to image/video quality, the rolling shutter can sample fluctuating light signals at a much more rapid rate. The potential contamination of ENFs by aliasing can be eliminated by sampling at a faster rate [43]. The rolling shutter's ability to sequentially acquire rows of each frame effectively multiplies the sampling rate by the number of rows per frame, thereby addressing the issue of an inadequate sampling rate. However, the rolling shutter method introduces the issue of idle time in successive frames, where sampling does not occur. Therefore, some light samples will not be captured during the idle time period which occurs at the end of every frame.

In [91], the authors initially leveraged a rolling shutter as a solution to the inadequate sampling rate for ENF extraction from video recordings. For their analysis, they formulated an L-branch filter bank model of the rolling shutter mechanism and showed how the dominant ENF harmonic shifted to different frequencies as a result of the idle period in the videos. To extract the ENF, the method ignores the idle time, utilizes the average of each row's pixel values as temporal samples, and concatenates them to form a row signal when the foreground has a uniform color or is eliminated using motion compensation. The row signal is then passed through a spectrogram to generate ENF traces. The ENF was then extracted by determining the primary instantaneous frequency in a narrow band at the desired frequency. The multi-rate signal analysis employed in this method to evaluate the concatenated signal reveals that neglecting the frame's idle time during direct concatenation could lead to a mild alteration in the computed ENF patterns [43].

The authors in [92], [93], and [94] also designed ENF extraction methods based on calculating the average intensity of each frame when the foreground is uniform (in the case of a white-wall video) or eliminated using motion compensation (in the case of videos with motion) to generate an intensity signal that is passed through Fourier analysis to extract the ENF. The extracted ENF was further used for video synchronization applications.

To avert such deformation in [91], the authors in [95] and [96] proposed a periodic zero-padding approach to deal with the idle time problem. The authors argued that instead of ignoring the idle period, equally uniform sampling along time should be conducted by returning zeros to the end of each row signal until they reach the idle time duration before combining them. This zero-padding strategy is capable of producing ENF traces that are free of distortion; however, it does need to know the duration of the idle period in advance. The idle period can be determined using the camera read-out time, which is model-specific [97]. The results of their experiment demonstrated that this technique enhanced the SNR of the estimated ENF signal.

In [98], the authors presented a phase-based method to address the idle time problem when the read-out time was not predefined. Their method aimed to apply row-by-row samples separately on a frame-by-frame basis to prevent any discontinuity owing to the idle period. DFT was applied to each row sample, followed by quadratic interpolation with its adjacent values. They used the read-out time estimated by averaging all sinusoidal cycles acquired from all frames. The precision of the read-out time estimation was verified using a vertical phase method.

In [99], the authors presented a model that replaced the primary ENF frequency with new ENF components determined by the duration of the idle period for each frame. Their model further demonstrated that the strength of the acquired ENF signal and the duration of idle time are inversely proportional. The authors further suggested a new idle period estimation method for camera forensics, which can also be used for videos when the nominal ENF is a multiple of frame rate.

In [100], the authors postulated a MUSIC spectrum-combining method that exploits the rolling shutter mechanism without requiring any prior information of the idle period. Given a test video, the method computes row-by-row, the average pixel intensity, and concatenates them in a singular time-series using a sampling rate of the frame rate multiplied by the number of rows. The time-series signal is then passed through the preprocessing stages, where the local average is subtracted to enable a more precise result, particularly when a video contains movements. The signal was further down-sampled to 1kHz after applying an anti-aliasing filter. Using the resulting signal, MUSIC and Fourier domain analyses were conducted in parallel, and the respective spectra were combined to extract the ENF. The proposed technique was utilized for time-of-recording verification, and the experimental results showed its effectiveness even in a more difficult scenario of 1-minute recording of both static and non-static videos.

TABLE 2. Summary of research in ENF estimation in video recordings.

Ref	Type of Video	Camera/Recorder used	Camera sensor	Duration	ENF extraction steps	Application
[11], [12]	Constant and non-constant scene	Not mentioned	CCD	12mins	Calculating each frame's average intensity of constant scene video and passing it over a band pass filter with a small range that matches the frequency of interest. Carrying out the spectrogram of average pixel intensity of relatively steady regions in video with motion	Time-of-recording, tampering detection, audio-visual authentication and synchronization
[65]	Constant and non-constant scene	Canon PowerShot SX230 HS/Canon PowerShot SX210IS	CMOS/CCD	2 ~ 15mins	Computed the average of steady super-pixels regions to generate the intensity signal. Applied Fourier analysis and quadratic interpolation.	ENF detection.
[81]	Constant and non-constant scene	GOPRO Hero 4 Black/NK AC3061-4KN	CMOS/CCD	12mins	Generated super-pixels having average intensity higher than a certain threshold using SLIC algorithm to form the mean intensity time-series. Applied Fourier and ESPRIT analysis.	N/A
[85]	Non-constant Scene	GOPRO Hero 4 Black/NK AC3061-4KN	CMOS/CCD	2 ~ 10mins	Applied ViBe and differentiator filter at pixel level, averaging pixels with luminance differences beyond a set threshold to form frame-level signal. Applied a Phase-Locked Loop-based FM demodulator	Time-of-recording
[91]	Constant and non-constant scene	Canon PowerShot SX230 HS	CMOS	10mins	Direct concatenation of available row's signals at the beginning of every frame while ignoring the signal during the idle time	N/A
[92], [93]	Constant and non-constant scene	Canon PowerShot SX230 HS and Canon PowerShot SX2300	CMOS	10mins	Computing each frame's row signal and concatenating them to create the source signal and then applying Fourier analysis. Identifying and averaging motion-free regions in each frame to create smoothened version of the source signal and applying Fourier analysis.	Audio-video synchronization, Historical recording synchronization
94	Constant and non-constant scene	Sony Experia M phone	CMOS	10mins	Calculating each frame's average intensity of constant scene video and passing it over a band pass filter with a small range that matches the frequency of interest. Carrying out the spectrogram of average pixel intensity of relatively steady regions in video with motion	Audio-video synchronization
[95], [96]	Constant scene	iPhone 6s back camera	CMOS	10mins	Using the T_{ro} to determine and zero-pad the idle time before concatenating the row signal.	N/A
[98]	Constant scene	Not mentioned	CMOS	2mins	Estimating the phases of the row signals and computing the phase differences between successive frames	N/A
[99]	Constant and non-constant scene	GoPro Hero 4, Nikon D3100, Nikon P100, Canon SX230HS, Canon SX220HS	CMOS	Not mentioned	Analytically modelling the strongest ENF component using possible idle period and then concatenating the row illumination samples using steady pixels average to form the time-series.	Idle time estimation, Time-of-recording.
[100]	Constant and non-constant scene	GOPRO Hero 4 Black/NK AC3061-4KN	CMOS/CCD	11 ~ 12mins	computes row-by-row, the average pixel intensity and concatenate them in a singular time-series using a sampling rate of frame rate multiplied by the number of rows. Applied MUSIC analysis	Time-of-recording
[101]	Constant scene	Huawei P20 Lite, HTC One mini2, Motorola Moto G, iPhone 7, and Samsung Galaxy A71 cameras.	CMOS	5 ~ 30MINS	Computed each frame's row signal and concatenating them to create the source signal. applied Fourier analysis.	Time-of-recording

C. ENF EXTRACTION FROM A SINGLE IMAGE

Recently, the authors of [13] and [66] demonstrated that it is possible to obtain ENF traces from a solitary image captured with a rolling shutter camera. Their work tried to answer questions about the possibility of detecting ENF footprints in an image and whether the location of image capture operate at

a frequency of 50 Hz or 60 Hz. According to their opinion, the sequential read-out time mechanism of a rolling shutter when capturing an image permits the acquisition of samples of the incoming electric light signal at various time points, resulting in a brief section of the ENF traces in the resulting image. As in the case of videos captured using a rolling shutter, every image row is sequentially captured, which includes a temporally fluctuating element attributed to the ENF. All image rows are acquired in the frame period, which according to the common frame rate, is normally 1/30 or 1/25s. This time is too brief to extract a significant temporal signal for the majority of ENF-based analysis tasks [43].

However, the authors illustrated that when smooth image regions are corrupted by real-valued sinusoidal signals, these regions tend to become complex or exhibit greater entropy. Because of the sinusoidal signal's positive and negative additive values, the lone bin in the histogram, prior to corruption, begins to divide into several bins when applied to an image column with uniform intensity. Because of such corruption, the entropy of the constant signal increases from zero to a positive value. The same statistical behavior occurs in the instance of an image column with a linear increase in intensity values; however, the increase is less pronounced than that in the instance of constant intensity. The histogram of the linearly increasing intensity case represents the window prior to corruption. After corruption, the bins on either side of the window divide, potentially increasing entropy. Based on this backdrop, the authors suggested an approach to minimize the entropy for images with parametric surfaces removed. The results of their experiment revealed the effectiveness of the proposed method, where ENF traces were prominent. However, there are several unresolved research problems in the study of ENF in images. For instance, the method suggested in [13] and [66] is only effective for images that contain synthetically embedded ENF. Its effectiveness in real-world photos can be further improved by adopting a more sophisticated physical embedding model [43].

D. FACTORS AFFECTING ENF CAPTURING AND ESTIMATION IN AUDIO AND VIDEO RECORDINGS

ENF traces were obtained from media recordings and reflected the electrical activity of the power network during the recordings. Therefore, it is essential to have electrical activity in the recording location to acquire ENF traces. Recordings made with recorders plugged into a wall outlet are widely believed to contain ENF traces owing to the electromagnetic interference caused by the recorder's connection to the wall outlet [5], [7], [8], [10], [39]. However, for audio recordings obtained with recorders powered by a battery, the situation becomes complicated because of several factors. Therefore, identifying the factors that either facilitate or impede ENF trace acquisition in media recordings can provide valuable insights into circumstances in which ENF analysis is relevant. Furthermore, this can aid in the development of ENF-based applications.

TABLE 3. Factors influencing ENF acquisition in audio recordings obtained with recorders powered by a battery [60].

Factors		Effect
Environmental	Electromagnetic (EM) fields	Foster ENF acquisition when the microphones used for recording is dynamic but not when the microphone is electret
	Acoustic mains hum	Foster ENF acquisition. Can be produced from fridges, power adaptors, fans, and lights.
	Electric cables in vicinity	Insufficient for ENF acquisition
Device-related	Type of microphone	Different types react differently to similar source, such as EM fields.
	Recorders' frequency band	Possible inability of recorders to capture low frequencies, such as around 50/60Hz
	Internal compression of recorder	ENF acquisition may be restricted by strong compression, such as Adaptive Multi-rate.

In general, there are two types of factors that can affect how well ENF traces are acquired: environmental and recording device-related factors. In addition, how different factors interact with each other may also lead to different results. For example, the electromagnetic field in the recording location fosters ENF acquisition when the microphone used for recording is dynamic. However, when the microphone is an electret, the electromagnetic field does not have the same effect. [43], [60]. Sample factors explored in the literature [10], [23], [39] and their effects on ENF acquisition in audio recording are presented in Table 3. The most prevalent source of ENF traces acquisition in audio recordings is the acoustic mains hum, that are generated by mains-powered appliances present in the recording location [39], [102]. A study conducted by the authors in [39] indicated that the ENF traces were highly robust, with traces detectable even in a recording carried out 10m distant from a source of noise located in a separate room. The author in [60] carried out experiments to examine the factors that impact ENF trace acquisition, such as wave interference, recorder movement, and recorder type.

1) WAVE INTERFERENCE

When several pieces of equipment connected to electrical power in a given location, acoustic mains hum can originate from a variety of sources. The different copies of the signal,

along with their reflections, would produce different levels of interference when they converge at the location where the recorder captures the signal, resembling the behavior of sound waves. The interference caused by sound waves that carry ENF traces was investigated in a controlled experiment by the authors of [60]. To conduct their experiment, they positioned two speakers 1 m apart and emitted a synthetically generated tone signal at 240 Hz. They then recorded the signals using two recorders placed at different locations within a room. The results of their study revealed that in a given environment where ENF traces are likely to be present, such as a room, the particular position of a recorder within that environment can influence whether the ENF is captured, and the reliability of the ENF traces that can be captured as a result.

2) RECORDER MOVEMENT

The authors in [60] illustrated that the movement of an audio recorder during recording can impact the quality of the ENF traces that can be captured. These observations could potentially be caused by air pressure fluctuations. Most microphones are susceptible to air pressure variations during mobility, which could lead to substantial noise that might potentially impact the ENF traces that are captured. Another factor is the Doppler effect, perhaps in combination with other sources such as airflow and vibrations. For the experiment, the authors recorded a 10-minute audio at an office in Maryland using an Olympus recorder.

During the first 5-minutes, the recorder was kept stationary, and afterwards, it was casually moved around by a person in the room by hand during the remaining 5-minutes of the recording. A plot of the ENF extracted from the audio recording with the reference ENF recorded simultaneously is shown in Figure 7.

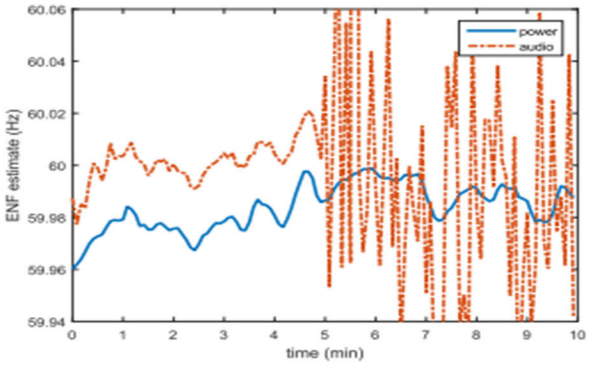


FIGURE 7. ENF signal obtained from the audio recording and a reference power recording captured concurrently [60].

In this experiment, we observed that in the first five minutes, there was a good correlation between the ENF signal in the audio and power signals, but this correlation diminished afterwards, when there was movement of the audio recorder. This indicates that the movement of the audio recorder during recording can affect the accuracy of the ENF traces obtained.

In another experiment, a controlled recording lasting 9-minute was made. This recording had three parts: a motion phase lasting 3-6 minutes where someone held the recorder and walked at a consistent pace towards and away from the source, as well as two stationary parts when the recorder was placed on a table, lasting 0-3 minutes and 6-9 minutes. A speaker, which served as the source was used to emit a 55Hz synthesized single-tone sound. In the section involving motion, the individual that held the camera walked away from the speaker nine times and towards it eight times. The plot of the frequency of the collected sound in the second experiment is depicted in Figure 8 using a 2-second sliding window and 90% overlap.

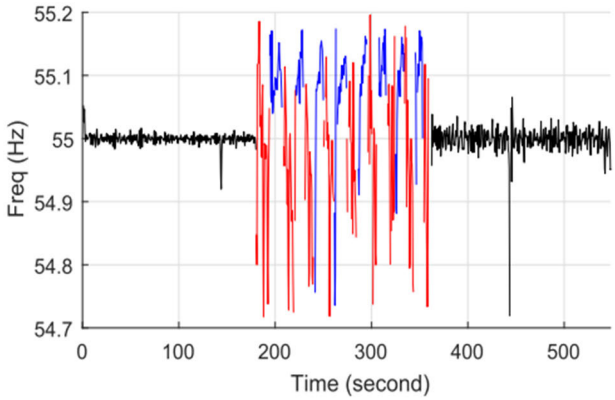


FIGURE 8. Estimated dominant frequencies over time showing the Doppler effect [60].

The graph depicts 17 distinct sub-phases within the motion phase lasting 3-6 min, where the frequencies alternate with respect to the source frequency. These frequencies match the movements towards (shown in blue) and away from (shown in red) the speaker. In general, the quality of the ENF traces that can be obtained may be adversely affected by moving the recorder during the recording process. In practice, proper attention should be given to this issue, for instance, processing the raw ENF trace with suitable compensations or denoising algorithms, as this would impose further challenges on the subsequent ENF-based forensic application.

3) RECORDER (RECEIVER) TYPE

The authors of Hajj-Ahmad et al. [60] demonstrated that the recorder or receiver of a recording device utilized to generate an audio recording can impact how ENF traces are recorded in the resultant recording. Their experiment employed diverse recorders in similar recording settings, which showed the capture of ENF traces of different strengths and multiple harmonics. Additional investigations in this area will improve our understanding of the usefulness of studies and aid in the development of ENF-based applications that can be scaled up.

In video recording, ambient conditions, camera settings and motion, data properties, and the inverse square law of light can affect the capture or estimation of ENF signals [61], [82], [103]. In [61], the authors studied how the impact of a main-powered illumination source as well as the compression of the video recording, affects the capturing and/or how good the captured ENF signal can be. They performed experiments using video recordings made under different lighting sources, such as white light-emitting diodes (LED), incandescence, and white compact fluorescent (CFL). The videos were divided into different sizes and compressed at various bit rates such as 5000Kbps, 1000Kbps, 500Kbps, and 100 Kbps. Furthermore, a compressed Facebook version of every video was generated by consecutively uploading and downloading from the website. Following their experiments for recording time, the author reported that the best ENF estimation was achieved for video captured under LED lighting, while the ENF signal quality declined dramatically under CFL lighting. They also noted that compressed videos at low bit rates (e.g., 100Kbps) and Facebook uploads noticeably diminished ENF detection performance on video captured under LED, while they resulted in a complete failure in ENF detection and time of recording verification for 2 min and 5 min videos captured under CFL.

The challenges of object motions in the recording scene, camera motion, and camera brightness changes in ENF capturing and estimation have been studied and addressed by the authors in [82]. The authors presented motion and brightness compensation schemes that enabled the effective estimation of the ENF.

In [103], the authors provided an analysis of the frame rate harmonics problem caused by the inverse square law of light. This problem often emerges when attempting to estimate the ENF signal from rolling shutter videos, especially when the intended ENF frequency is a multiple of the video frame rate, causing an overlap between the harmonics of the ENF and frame rate. As such, the authors presented a method based on refining the luminance waveform to significantly reduce the impact of frame rate harmonics and consequently improve the measured ENF signal.

VI. ENF APPLICATIONS

ENF analysis is an application-driven research topic. Researchers have employed this tool in for several multimedia forensic applications, such as verifying recording time, detecting forgery/tampering, authenticating location, and identifying camera/devices used for recording. There are also applications beyond forensic analysis, such as multimedia synchronization. This section provides a comprehensive review of studies conducted in various application areas.

A. TIME OF RECORDING VERIFCATION

Initial studies on ENF-based forensic analysis [4], [5], [6], [7], [9], [10], [104] concentrated primarily on the development of audio timestamp verification systems. The ENF pattern of an audio signal should closely resemble the ENF pattern of a power reference recording performed simultaneously. Therefore, given a test audio signal with a power reference captured at the stated audio recording time, the extracted ENF sequence of the audio signal can be compared to the ENF sequence derived from the power reference signal. The stated duration of the audio signal's recording time can be considered authentic if the derived audio ENF sequence matches that of the reference ENF sequence [43].

When verifying that the timestamp of an audio file and the questioned ENF contained a specified recorded time, a speedy conclusion may be reached by visual comparison of the questioned ENF with the simultaneously recorded reference ENF. The least efficient visual comparison provides the quickest effective method for assessing ENF matching [6], [8]. However, visually searching for a match between the questioned recording and the reference ENF becomes impractical when the questioned recording has no declared date on which it was recorded. This is because finding a corresponding segment of the reference ENF from a huge reference database is necessary for matching the questioned ENF. In addition, a visual comparison did not provide a quantitative measurement of the level of similarity between the two sequences. These drawbacks limit the adoption of visual comparison in real applications to estimate recording time. Usually, the time-frequency pattern of the reference ENF acquired is significantly greater than the length of the recorded audio to enable a search within the timeframe in which the audio data were probably obtained. Hence, to determine the level of similarity between the derived ENF signal and all possible reference segments, either the minimum mean squared error (MMSE) or maximum correlation coefficient can be employed [8]. For a time-frequency pattern of the extracted and reference ENF signals represented as $\boldsymbol{r(m)}$ and $\boldsymbol{a(m)}$ with length $\boldsymbol{H}$ and $\boldsymbol{M}$ respectively, $\boldsymbol{H} > \boldsymbol{M}$, the MSE, $\boldsymbol{e}$, is expressed as

$$\boldsymbol{e}\,(\boldsymbol{i}) = \frac{1}{\boldsymbol{m}} \sum_{\boldsymbol{m}=0}^{\boldsymbol{M}-1} \left(\boldsymbol{r_i}\,(\boldsymbol{m}) - \boldsymbol{a}\,(\boldsymbol{m})\right)^2 \tag{5}$$

and the correlation coefficient, $\boldsymbol{\rho}$, is expressed as

$$\boldsymbol{\rho}\,(\boldsymbol{i}) = \frac{\sum_{\boldsymbol{m}=0}^{\boldsymbol{M}-1} \left[\left(\boldsymbol{r_i}\,(\boldsymbol{m}) - \bar{r}i\right)(\boldsymbol{a}\,(\boldsymbol{m}) - \bar{\boldsymbol{a}})\right]}{\sqrt{\sum_{\boldsymbol{m}=0}^{\boldsymbol{M}-1} (\boldsymbol{r_i}\,(\boldsymbol{m}) - \bar{\boldsymbol{r}}_i)^2} \sqrt{\sum_{\boldsymbol{m}=0}^{\boldsymbol{M}-1} (\boldsymbol{a}\,(\boldsymbol{m}) - \bar{\boldsymbol{a}})}} \tag{6}$$

where $\boldsymbol{i} \in \{0, 1, 2, \ldots, \boldsymbol{H} - \boldsymbol{M}\}$ and $\boldsymbol{r_i}\,(\boldsymbol{m}) = \boldsymbol{r}(\boldsymbol{m} + \boldsymbol{i})$.

The timestamp estimate can be calculated using the value that produces the MMSE or MCC [22]. Several suggestions have been made to improve the matching criteria in published studies.

In [105], the authors suggested a method that employs a piecewise linear autoregression (AR) process to model the ENF signal, with the aim of improving ENF-based recording time estimation matching. The model enabled the authors to disintegrate the ENF signal into predictive and innovation processes and analyze its performance using a hypothesis detection framework. The innovative signals generated by

the AR modeling of the ENF signal were then extracted and utilized for matching instead of the original ENF signals. Their experiments, which were conducted using a hypothesis detection framework, demonstrated that the method provides a more confident result in recording time estimation and validation.

To improve the matching accuracy when the extracted ENF signal is affected by in-band noise as well as the resolution problem, the authors in [22] introduced a threshold-based dynamic matching method called error correction matching method (ECM) that can automatically correct the noise-impacted frequency estimation before performing the matching. In their approach, a threshold is set based on the windowed length of the STFT, which determines the frequency resolution of the windowed signal. Based on this threshold, an auto-correction procedure is performed during the matching process. The superiority of the proposed method was demonstrated through experimental results obtained using both synthetic and actual signals. However, its performance has the drawback of requiring additional computational time. Consequently, it can only be used for verifying audio timestamps when a user specifies a short search range for the reference ENF [106].

The authors in [106] presented a similarity measurement method for matching two ENF signals called bitwise similarity matching (BSM). The BSM simply assesses the extent of the correlation between two ENF sequences by binarizing the local difference scale. A pair of bits whose absolute error is below a predefined threshold value is considered to be matched. After the process of converting into binary form, the dissimilarity between two sequences is simplified into a solitary sequence of bits, with successive 1s indicating a segment that is locally matched. The results of their experiment revealed that the proposed BSM technique is effective and efficient, and has a considerably reduced computation time.

In [107], the authors provided a comparative study of ECM and BSM techniques in comparison to the classical MMSE matching criterion. The study opined that the use of a predetermined threshold provides the advantage of tolerating inevitable errors during the matching process, unlike in the MMSE. On the other hand, tolerable matching errors may result in inaccurate matches in some scenarios, whereas the MMSE baseline may produce an accurate match. The threshold-based approaches under such scenarios are inferior to the baseline method. Their analysis revealed that in practical situations, the ECM and BSM methods may not necessarily perform better than the baseline MMSE method owing to the complexity of the scenarios. However, the results of the seven scenarios investigated by the authors show that the ECM produced the most precise matching, whereas BSM compromised matching precision in favor of faster computation time.

The authors in [108] constructed a robust approach for time-stamping relatively contaminated ENF contents. The approach employs a threshold value to detect all useful samples of the estimated ENF signal, specifically those that are comparatively non noisy. The threshold values were specified for the maximum allowable departure from the nominal value and the speed of the ENF variation based on the reference signal. For both the extracted ENF signal and the reference data, a binary mask is created to indicate the samples that are useful. Next, the masks were used to execute a modified NCC analysis between the extracted ENF signal and reference data. The results of the experiment indicated that the proposed method considerably enhanced performance.

The authors in [99] presented a technique for verifying the recording time of ENF signals estimated from videos recorded using rolling shutter cameras. As previously stated in Section V, rolling shutter cameras usually have an idle period between successive frames where no exposure is performed, leading to the loss of some samples of the ENF in the resulting video. This method involves interpolating the absent period of illumination within every frame by making different assumptions about period of inactivity to compensate for the lost data points. An ENF signal was then obtained from each interpolated time series and compared to the ENF reference signal using correlation coefficients to determine or authenticate the recording time.

The process of verifying the authenticity of a time-stamp through ENF features is considered successful when the ENF obtained from a suspicious media file is accurately and exclusively matched to a segment of the ENF reference that corresponds to a particular time-frame that may include the actual time-stamp of the recording being tested. In terms of uniqueness, this implies that one best match should be produced in the match process based on metrics such as MSE or correlation coefficient (CC). In terms of accuracy, it implies that the best match should show that there is sufficient similarity between the reference segment and the derived ENF from the test to make a judgement, such as having a sufficiently small MMSE or CC value close to one. Therefore, given a test recording as well as a match scope, the question of the reliability of the time-stamp authentication result logically arises. The authors in [109] attempted to answer this question by examining the factors that determine the dependability of ENF matching such as the SNR, duration of the reference signal, duration of the test recording, and ENF estimation temporal resolution. They introduced a time-frequency domain (TFD) synthesis scheme and employed it for analysis utilizing both synthetic and actual data. After analyzing the data, they observed that the most critical external component was the SNR, whereas the duration of the test recording was the most significant of the other three intrinsic factors, followed by the duration of the reference ENF. They further observed that the ENF matching process was insensitive to ENF temporal resolution.

An actual scenario where ENF analysis was used to verify the recording time of a media recording in a law court was presented in [6]. The authenticity of a recording of the conversation between two business men provided as evidence

was questioned by a public prosecutor in Cracow, Poland in 2003. It was a 55-minutes long recording where the recording time and date stated by the evidentiary recorder differed by 196 days and 14 hours, respectively, from the time and date claimed by the witnesses to the conversation. The Institute of Forensic Research, tasked with examining audio recordings, identified the existence of an ENF signal through their analysis.

The ENF signal taken from the recording used as evidence was matched with ENF signals obtained from the power network operator for reference. The result of the analysis showed that the true time of the evidentiary recording was the time claimed by witnesses. The reason for this discrepancy is that the user of the recording device incorrectly set the date and time before starting the recording. Figure 9 shows a comparison between the ENF signal extracted from the evidentiary recorder and the reference ENF signal obtained from the power network operator.

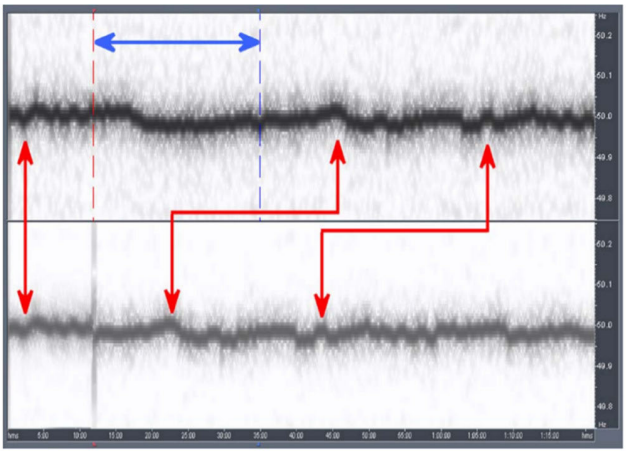


FIGURE 9. Comparison of signals before (upper) and after (lower) modification. Red arrows represent typical sections in both signals; blue arrows show which component of the original signal was deleted from the lower signal [6].

The ENF signal initially caught the attention of the forensic community owing to its application in time-of-recording authentication. However, various obstacles remain to be overcome before it can be widely used in practical situations. First, the aforementioned study assumed no signal tampering. If the signal has been altered, the strategies described in this sub-section might not be effective. Second, determining whether a media's ENF sequence matches all potential ENF sequences may require a high computation cost, based on the information initially available about the recording. A solution to this problem is proposed in [110]. Third, this application presupposes that the time and place where the media recording in question took place already have its reference ENF measured. If the reference data are insufficient or non-existent, a forensic investigator will have to employ alternative methods to conduct the investigation. One such method determines the grid-of-origin (which will be discussed later in this section) of a media recording without requiring simultaneous reference of ENF data [43].

B. LOCATION AUTHENTICATION

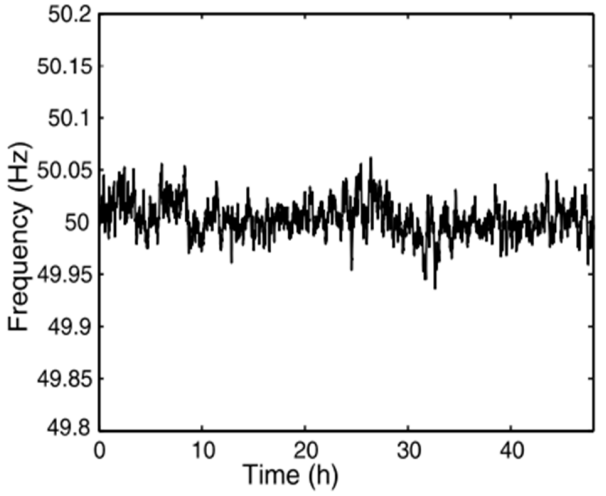
This sub-section discusses the use of ENF signals for location-information inference. The details of a media recording can be obtained from ENF patterns derived from such recordings (e.g., audio and video recordings) and utilized to discern its location. Various techniques have been suggested in the literature that can employ the ENF signal to infer the grid where a media recording is captured, referred to as inter-grid localization, and to pinpoint the environment within a grid where a media recording is obtained, referred to as intra-grid localization.

1) INTER-GRID LOCALIZATION

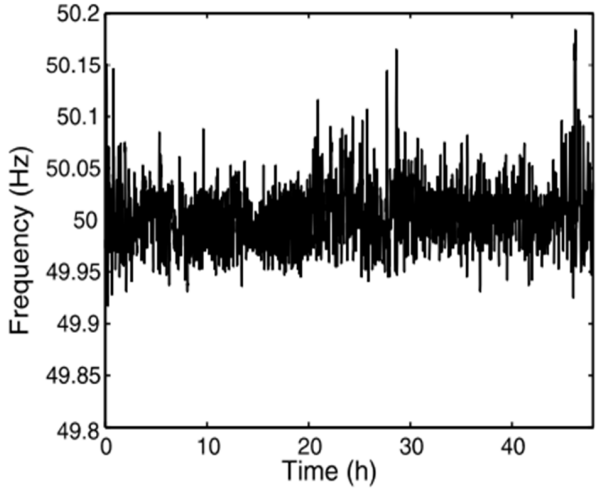
Several methods have been implemented that attempt to determine the source grid of a multimedia recording containing the ENF signal when the power reference signal recorded simultaneously is unavailable. Such methods can find useful applications in security and forensic analysis to determine the origin of ENF-containing multimedia files, especially those involved in domestic violence, child exploitation and pornography, ransom demands, and terrorism attacks [111]. Such methods can also be leveraged to significantly minimize the computational complexity when performing recording time or intra-grid localization estimations. In a situation where a forensic examiner is provided with media recording with no information about its time and location, he/she can initially apply these methods to determine the source grid where the media file was recorded. Subsequently, the reference data of the selected grid can be utilized to perform additional forensic processes, including recording time authentication and precise location estimates [112].

ENF signals obtained from various power grids were observed to exhibit different types of ENF variations, which are usually associated with the control mechanism employed to manage the frequency of the standard value, as well as the power grid size. The average 1-minute frequency over a 48-hours period from three distinct locations and three distinct grids is shown in figure 10 [25]. Spain, connected to the large continental European grid may be observed in this figure to exhibit a narrow range of frequency values and rapid variations. Great Britain, which is the smallest of the three grids, showed relatively large values of frequency variations. In general, frequency variations tended to decrease as the size of the power grid increased.

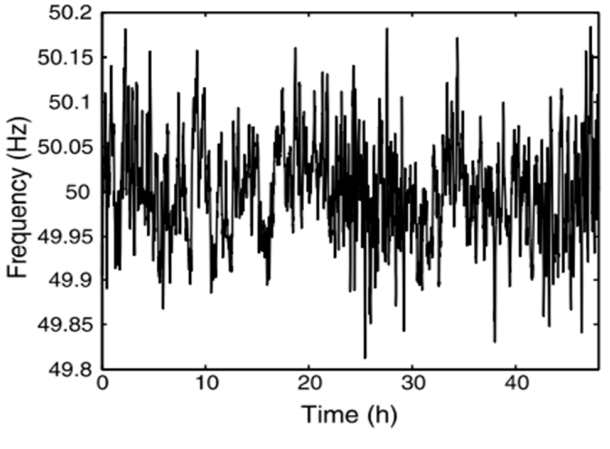
For any test media recording containing the ENF acquired from various grids, the ENF signals in the test recordings can be analyzed to retrieve unique and relevant features. These features can then be exploited to train a machine learning system to classify ENF signals based on their source grid. Researchers have developed several such systems to pinpoint the source grid of any ENF-containing media recording.



(a)



(b)



(c)

FIGURE 10. Measurement of frequency variations in (a) Spain. (b) China. (c) Great Britain [24].

The author in [112] constructed a machine learning system that could determine the recording location where an ENF-containing media file was captured when its simultaneous power reference was unavailable. From 11 distinct grids across the world, the authors obtained power and audio recordings and extracted ENF signal segments of 8-minutes length from each recording. From the segments, statistical, linear-predictive, and wavelet-based features were extracted and utilized to train a multi-class Support Vector Machine (SVM). The authors separated the ''clean'' and ''noisy'' ENF signals obtained the power and audio recordings, respectively, and then combined them in varying proportion during training in various settings assess the impact of the training data type on the resulting testing outcomes. The proposed approach identified ENF signals in the power and audio recordings from 11 target power grids with 88.4% and 84.3% overall precision, respectively. The authors further investigated the use of multi-conditional methods that can accommodate situations in which different noise conditions exist for the training and testing data. When the dataset used for training was restricted to clean ENF signals from retrieved power recordings, this method improved the detection accuracy of noisy ENF signals of extracted audio recordings by 20%.

The SVM method proposed in [112] was improved by the authors in [113] by making the classification method a multi-stage method. The multi-stage classification system utilized the given training datasets to classify the input signal in three stages. First, it distinguished whether the signal was either power or audio. In the second stage, the nominal frequency of the signal was determined, and in the third stage, a region-of-recording analysis was performed. Their proposed method yielded an improvement of 17.33% in the overall accuracy.

The authors in [114] investigated the use of five machine-learning algorithms in region-of-recording identification using power and audio recordings captured from ten various power grids. The machine-learning algorithms examined were SVM, K-nearest neighbors, linear perceptron, random forests, and neural networks. In their analysis, they employed the features used in [112] and extra features related to the extrema, rising edge, and autocorrelation. The features associated with the extrema and rising edge in the ENF signal demonstrated an improved performance, ranging from 3% to 19%. In their setup, the RF classifier achieved an overall highest accuracy of 91%.

The authors in [115], [116], [117], [118], [119], and [120] also presented machine learning systems for inferring source grids using power and audio recordings captured from nine grids [121]. In the implementations in [115], [116], and [117], statistical and wavelet-based features were extracted and utilized to train a multi-class SVM. The overall accuracies achieved by the works in [115] and [116] were 88% and 87%, respectively, while the work in [117] achieved an overall accuracy of 94.7%. In the implementation in [118], extra features such as spectral centroid and roll-off together with statistical and wavelet-based features are used to train Random Forests, SVM, and AdaBoost classifiers, with Random Forests achieving the most effective overall accuracy of 88%. In addition, in the setup used in [119], different combinations

of statistical and wavelet-based features and features associated with extrema and rising edges were fed to a Binary SVM classifier, which yielded a classification accuracy of 92%. In the implementation in [120], the authors trained a Linear SVM and a Bagged Tree using statistical and wavelet-based features. The experimental results revealed that the Bagged Tree outperformed the Linear SVM, with an average accuracy of 82%.

In [122], the authors used a dataset of power signals recorded at the distribution level from six distinct grids in the US and ENF signals from audio and video online datasets to infer the source grid. The extracted statistical and wavelet-based features were used to train the SVM and XGBoost algorithms, respectively. The accuracy of the proposed algorithm was evaluated for each dataset using 15-minutes, 30-minutes, and an hour time unit. The experimental results revealed that the XGBoost classifier achieved a more accurate estimation result when the number of analysis data points increased.

To ensure cost-effective and scalable location inference, the authors in [123] introduced a method that uses a LISTEN attack to construct an ENF signal MAP from online streaming multimedia data, extract the ENF signal from candidate devices, and estimate their grid-of-origin. The LISTEN attack approach first crawls and scraps audio and video data from multimedia services (e.g., "EarthCam" [124], "Skyline" [125], "Explore" [126]) with recording location information and is assumed to have been captured using AC main-powered devices. Then, the multimedia data are analyzed using a series of signal processing methods to detect, retrieve, and create a map of the signal. The second step involves setting a target victim and extracting the ENF signal from the recorded voice or device of the victim. In their experiments, they were able to determine the source grid of audio streams obtained from Skype and Torfan across seven different power grids. They achieved an accuracy rate of 85-90% in classifying audio lasting between 10-40 minutes. As we'll see, their suggested solution also addressed intra-grid localization.

2) INTRA-GRID LOCALIZATION

At an inter-grid level, research has shown that it is possible to distinguish between recordings obtained across various grids, because the ENF signal variations at a specific moment are often distinct across grids that are operated independently. At the intra-grid level, it is assumed that at different points within the same grid, the ENF variations recorded simultaneously are very similar. However, studies have revealed that there are detectable disparities among these variations, which may likely result from alterations in the power consumption of a certain city as well as the duration required to transmit a reaction to load changes to other sections of the grid [35]. The varying ENF values may also be caused by system disturbances, including short circuits, line switching, and generator failures [24]. A small load change in a particular region may have a localized impact on the ENF in that region. However, a significant change, such as generator disconnection, may affect the entire grid. This change is propagated over the grid of the Eastern US at a regular rate of approximately 500 miles per second [33].

The authors of [46] and [127] hypothesized that load changes may create position-specific fingerprints in the ENF pattern, which may then be used to identify the exact site within a grid where the recording was made. They expected that the ENF signals for locations near one another would be significantly similar to those farther apart, because of the limited transmission rate of the frequency disturbances throughout the grid. In their experiment, the authors in [127] made three concurrent power signal recordings in Cottage Park, Maryland, Princeton, New Jersey, and Atlanta, Georgia, which are part of the US eastern grid. Figure 11 shows a plot of the ENF signals obtained from the recordings.

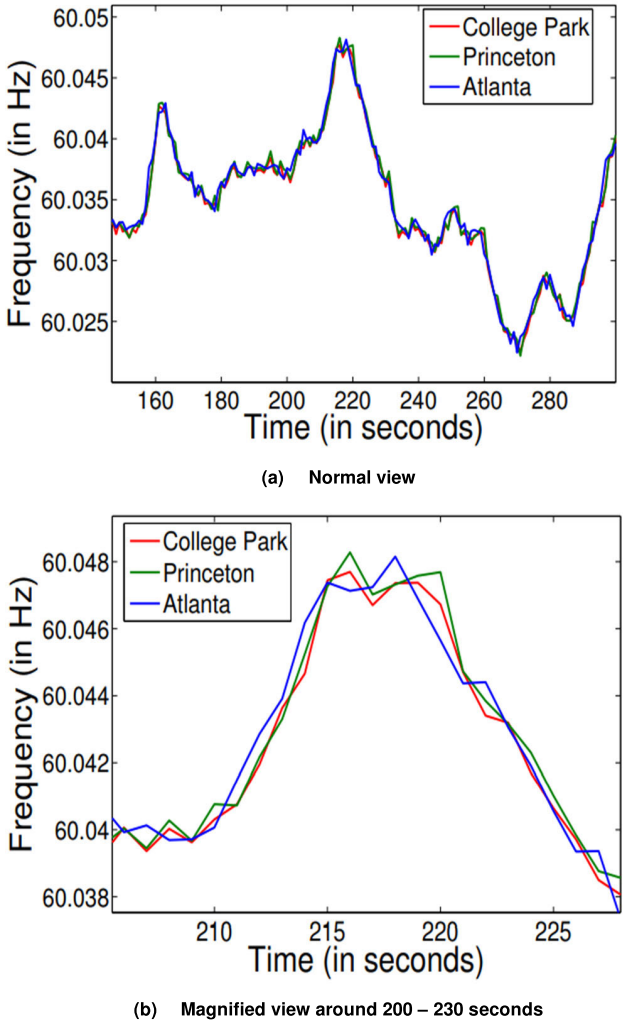


FIGURE 11. ENF signals samples obtained from power signal recordings made concurrently in three locations in the US eastern grid [127].

They observed a high correlation between the three ENF signals at the microscopic level; however, some variations

between the three signals were detected in the magnified plot presented in figure 11(b). Next, the extracted ENF signal underwent high-pass filtering to isolate variations, and the similarity of location across recordings was analyzed by measuring the CC between the filtered segments. The CC between distinct 600s ENF segments from various sites at the same time is presented in Figure 12. As shown in figure 15, that as closer two cities are, the higher the CC between their ENF signals, and the more distant they are, the lower the CC of their ENF signals. Based on this observation, the authors [127] designed a half-plane intersection approach to estimate the position of recordings containing ENF signals. They discovered that the precision of localization may be enhanced by increasing the number of ancho-nodes positions. Their results revealed that the proposed approach could yield an estimation precision of 90% under specified situations.

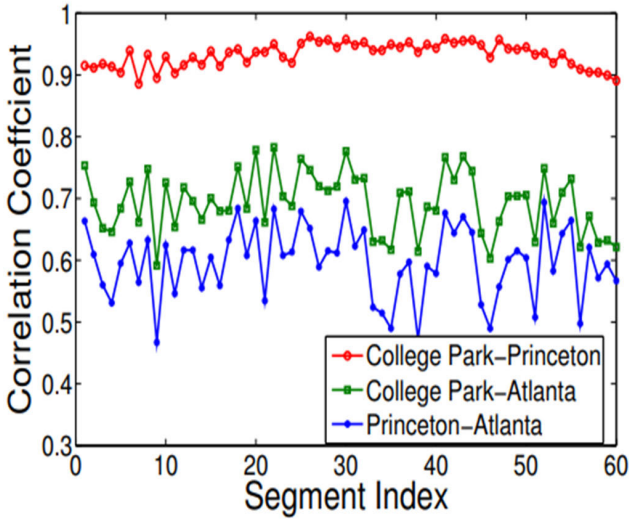


FIGURE 12. Three-location data from the US east coast and the pairwise correlation coefficient between their high pass-filtered ENF signals for query segment of 600 seconds long [127].

The LISTEN framework proposed in [123] for identifying the grid of recording also includes the capability to infer a particular site where a target recording was made using the map of the ENF map constructed from the ENF trace obtained from various live-streaming websites. After determining the source grid, they calculated the Euclidean distance between an interpolated time-series sequence of signals in the selected source grid and the ENF signal present in the target recording. This technique then pinpoints the recording site to a particular area within the grid.

The authors in [128], [129], and [130] utilized the FNET/GridEye system to pinpoint locations within a grid using the features of noise and variations in ENF signals derived from clean power signals without relying on the simultaneous ENF power references. These types of noise characteristics and ENF signal variations are caused by electromechanical propagation, nonlinear loads, and recurring local disturbances [128]. In [128] and [129], wavelet-based features extracted from ENF signals were utilized to train a neural network. The results in [128] found that the proposed system achieved more accuracy at a larger geographical scale, as well as when there was less time between the recordings. In [130], statistical and wavelet-based features were fed into a Random Forest classifier for training. According to the experimental results, the precision of identification is heavily influenced by two key factors: the distance between the measurement sites and the quality of the frequency signals.

In [131], the authors introduced a trilateration-based method that employed the Half-Plane Intersection approach presented in [127] and a correlation quantization method for intra-grid localization without requiring training or reference recording acquired from the query location. As the distance between the recording locations increases, the correlation between the two high-passed signals retrieved from these areas diminishes. This finding was demonstrated through the analysis of recordings taken in different geographical areas within the same grid. Based on the correlation-distance relationship, the authors suggested a correlation quantization scheme. In the proposed scheme, the correlation values were quantized in distance ranges and used by a trilateration protocol to evaluate the possible recording area for a given city. The performance of the proposed method was evaluated using an ENF file acquired from five different areas. Four of these areas were used as anchor cities, whereas the ENF file of the fifth city was used as the query to evaluate its area. The results of their study indicated that the localization precision was enhanced with the correlation quantization technique compared with the half-plane intersection technique. In another experiment, the authors used a combination of the two methods, which yielded the best localization effectiveness with regard to the localization probability and area. Their results also show that by connecting more sites as anchor nodes, the distance-correlation relationship may be further improved, and the localization accuracy is enhanced, resulting in a scalable finer localization approach.

C. FORGERY/TAMPERING DETECTION

ENF-based authentication relies on the basic idea that when a recording containing ENF traces is manipulated, the modifications will also alter the extracted ENF signals. Therefore, ENF analysis of a manipulated signal exposes discontinuities in the retrieved ENF signal, raising suspicions of manipulation. Such manipulations include deleting portions of the recording, adding unrelated segments, and combining clips of several recordings [132], [133].

Earlier studies have shown that with the availability of reference ENF, checking for manipulation can easily be done by a comparison of the extracted ENF and reference ENF signals obtained from the reported time and location of the recording, to either confirm or cast doubt on the integrity of the recording.

The authors in [7] presented a scenario in which the comparison of an ENF signal derived from a video file with its

TABLE 4. Summary of research in ENF localization.

Ref	Localization Type	Dataset Used	Features extracted and used	Classifier	Overall Accuracy
[112]	Inter-grid localization	Power and audio signal recording in from eleven and eight candidate grids respectively including US East, US West, Texas, Quebec, Cruise Ship, China, India, Ireland, Lebanon, Tenerife, and Turkey.	Statistical, wavelet and linear-predictive-related features.	SVM	88.4% (power signal) and 84.3% (audio signal)
[113]	Inter-grid localization	SPCUP2016 dataset	Statistical and window features	KNN, Multi-stage SVM	17.33% improvement of [122]
[114]	Inter-grid localization	SPCUP2016 dataset	Statistical, wavelet, and linear predictive features, features connected to rising edge, extrema, and autocorrelation.	SVM, KNN, Linear Perceptron, Random Forests, and Neural Networks	91%
[115], [116], [117]	Inter-grid localization	SPCUP2016 dataset	Statistical and wavelet features	Multi-class SVM	88%, 87%, 94.7%
[118]	Inter-grid localization	SPCUP2016 dataset	Statistical, wavelet, spectral centroid, and roll-off features	Random Forest, SVM, and AdaBoost	88%
[119]	Inter-grid localization	SPCUP2016 dataset	Statistical, wavelet-based features, and associated with rising edge and extrema	Binary SVM	92%
[120]	Inter-grid localization	SPCUP2016 dataset	Statistical, wavelet-based features.	Linear SVM and Bagged Tree	82%
[122]	Inter-grid localization	Power system distribution level ENF recorded data and online audio and video dataset.	Statistical and wavelet-based features.	SVM and XGBoost	88.4%
[123]	Inter-grid/intra-grid localization	Online streaming multimedia data (audio and video) from “Skyline”, “EarthCam”, and “Explore”.	Signal map	Bayesian Framework	76%
[46], [127], [131]	Intra-grid localization	Power and audio data recording from College Park (MD), Princeton (NJ), Cambridge (MA), Atlanta (GA), Champaign (NC), and Raleigh (IL) in the US.	Pairwise correlation coefficients		96.9%, 92.5%
[128]	Intra-grid localization	FNET/GridEye historical frequency dataset of selected locations in the US Eastern Interconnection (EI)	Statistical and wavelet-based features	Feed-forward Artificial Neural Network	90% and 80% on a 6- and 12-month time interval respectively.
[129]	Intra-grid localization	FNET/GridEye historical frequency dataset	Background noise-based features	Neural Network	94.9% of 1-month time interval.
[130]	Intra-grid	FNET/GridEye historical frequency dataset of 12 cities within the Western Electricity Coordinating Council (WECC) interconnection in the US.	Statistical and wavelet-based features	Random Forest	83%

reference signal exposes the inclusion of an extraneous video segment, as shown in figure 13. However, various alternative techniques have been discussed in the literature to identify manipulations in recordings containing ENF, even in cases where no reference ENF information is available. In general, when manipulations occur in a recording containing the ENF signal, discontinuities in the ENF phase are likely to exist in the parts where the manipulation occurred.

The authors in [20] conducted experiments to study the impact of phase changes on ENF signals using original and edited speech signals. Editing of the edited signal was performed at the original sampling frequency by swapping two small frames. The researchers noted that the ENF signal implanted in a recording can experience a modulation effect owing to phase variations when the recording is bandpassed within the nominal ENF band. Figure 14 shows an example,

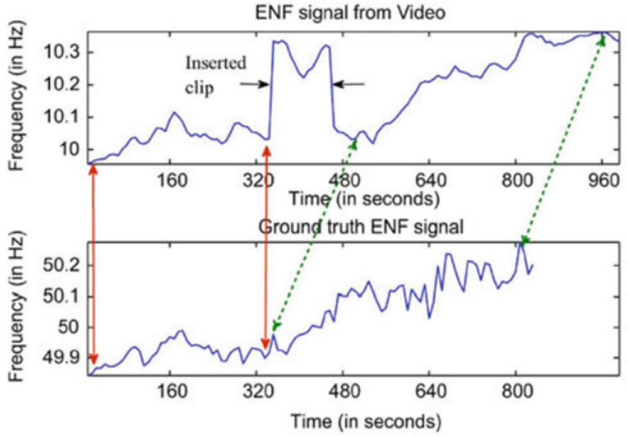


FIGURE 13. ENF matching result indicating detection of video alteration based on ENF traces [7].

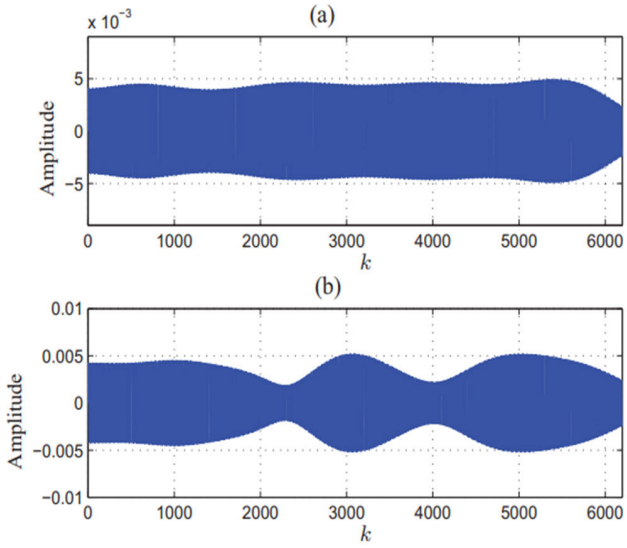


FIGURE 14. Original and edited audio signal from Brazil bandpassed around the nominal value of 60Hz, where (a) is the filtered original signal and (b) is the filtered edited signal [20].

where noticeable discrepancies exist between the relative amplitude of the edited signal and that of its original version. We note that a drop in amplitude occurred at spots where the authors made changes to the audio signals.

The authors of [134] presented a method that identifies and quantifies the phase discontinuity of the power signal. The audio recording was downsampled and band-pass filtered around the nominal ENF value. The output was regarded as a solitary tone for which the phase could be evaluated using a high-precision Fourier analysis method called DFT [135]. Plotting the estimated phase serves as a useful visual aid not only to detect the editing spots (which are indicated by abrupt phase shifts) but also to deduce the sort of audio editing that was performed (the addition of unrelated or removal of part of audio fragments), as shown in figure 15. The authors further introduced an automatic method that utilizes features derived

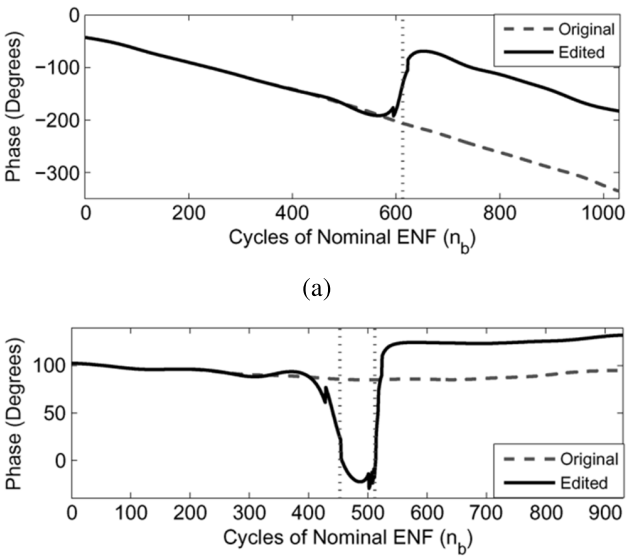


FIGURE 15. Estimation of phase from edited Spanish audio signal dataset using DFT, indicating (a) segment deletion and (b) segment insertion [134].

from the evaluated values to measure the discontinuity of the ENF phase and to certify an original or edited signal. They defined the suggested feature F as:

$$F = 100log\left\{\frac{1}{N_f - 1}\sum_{n_f=2}^{N_f}\left[\emptyset\left(n_f\right) - A_{\emptyset}\right]^2\right\} \tag{7}$$

where N_f represents the total number of frames used to estimate the phase, $\emptyset\left(n_f\right)$ represents the estimated ENF phase of frame n_f, and $A_{\emptyset}$ denotes the average of the calculated phases. The hypothesis groups H_0, H_E is then defined for the process of detection, where H_0 and H_E denote the original and edited audio signals hypothesis, respectively. If the value of F surpasses a certain threshold $\boldsymbol{\gamma}$, then hypothesis H_E holds, indicating the detection of editing in the audio signal. Otherwise, the audio signal is original, that is, hypothesis H_0. Their experiments using Spanish databases (AHUMADA and GAUDI) yielded 6% EER. Their experiments using Brazilian databases (Carioca 1 and Carioca 2) yielded a 7% EER.

In [136], the authors improved their work in [134] to account for a situation in which the ENF signal of the test recording could not be found at the ENF nominal value owing to corruption. This type of situation can also be caused by attacks targeted at preventing forensic analysis [137], [138]. With the assumption that non-linearity in the recording technique allows for higher harmonics of the ENF to be found in the signal, the authors described extracting the ENF from higher harmonics for their automatic tampering detector approach.

The authors in [139] developed a new method for identifying edits in forensic audio analysis, which included improvements in their methodology [134]. The new method applies a

data-driven threshold-based approach to ENF analysis as well as a decision criterion to identify abnormal ENF variations associated with edit activities. After conducting a qualitative investigation into the impact of noise contamination and the duration and location of edits on detection capability, the authors evaluated the effectiveness of the approach with regards to EER detection. Experimental results using the Carioca dataset of edited and unedited signal, which was also used in [134], revealed that the proposed new method accomplished a 4% EER, whereas the edit detector reported in [134] achieved a 7% EER. The proposed method also achieved an EER of 6% on the new Spanish dataset. According to the authors' findings, amplitude clipping and additive broadband noise have a significant impact on the detection effectiveness of the proposed method, indicating that further study is required to enhance detection performance in more difficult settings.

In [140], the authors enhanced their method in [139] by exploiting the typical ENF fluctuation pattern induced by audio editing to improve the detection criteria. Their major improvement on the threshold-based detection strategy of [139] is to check whether an identified anomaly is consistent with a particular pattern of ENF fluctuations triggered by edits in the signal under testing, which reduces the number of false positives produced by the edit detector. Experiments were conducted to directly compare the proposed improved method with the baseline method of [139]. The results revealed that the improved method is more dependable than the baseline method because it tends to provide a lower percentage of EER detection: a decrease from an EER of 4% to an EER of 2% in the Carioca 1 database and a decrease from 6% to 1% EER in the Spanish database.

The authors in [141] constructed a novel method for verifying the authenticity of audio recording, which employs a phase-locked loop (PLL). Their method utilizes a voltage-controlled oscillator (VCO) in the PLL to generate an artificial signal that is comparable to the input signal. By comparing the artificial and input signals, any discrepancies between the two can be detected, and the VCO signal can be adjusted to better match the ENF signal. However, if there are severe phase fluctuations, substantial disparities will occur between the VCO and ENF signals, indicating possible manipulation. The authenticity of the audio was subsequently established through an automated process that involved comparing VCO frequency fluctuations with an estimated threshold. Experiments using the proposed method on the Carioca and AHUMADA databases indicated that the proposed method can identify ENF discontinuities at a high accuracy level, attaining 2% EER and 4.5% EER, respectively, which is approximately the same performance achieved in previous studies [134], [139].

The authors in [132] introduced an audio authentication system that utilized the ENF to conduct both timestamp validation and tampering detection. Given a testing audio recording, the system processes the recording to check for tampering. If no tampering is detected, the system performs a timestamp verification. Otherwise (if tampering is detected), the system performs tampering detection operations to ascertain the nature of tampering (deletion, insertion, or splicing) and the region where tampering occurs. The system employs an absolute error map (AEM) to determine the authenticity of the recording being analyzed. This involved comparing the extracted ENF signal from reliable segments of the recording with the ENF signal from the reference database, which represents a two-dimensional representation of all the local errors that occurred during the marching operation. The AEM was then used to make authentic decisions. The features of the AEM demonstrate how its line pattern corresponds to three tampering operations: insertion, deletion, and splicing. In addition, the authors developed two techniques that can recognize horizontal line patterns in the AEM and generate the necessary validation output without any manual intervention. If the signal under study is determined to be authentic, the system delivers a timestamp when a match is discovered. If not, the two algorithms inspect the AEM to classify the tampering type and identify tampered regions. Synthetic performance analysis and experimental findings demonstrated the effectiveness of the concept and a number of real-world considerations.

In [142] and [143], the authors proposed tamper detection techniques that incorporate the ESPRIT-Hilbert ENF estimator together with an SVM classifier termed SPHINS to detect tampering. The proposed SPHINS techniques employ an ESPRIT-Hilbert ENF estimator, which provides a summary of the ENF disturbances by utilizing sample kurtosis and feeding them into an SVM classifier. Their proposed SPHINS technique achieved an improved performance of 4% EER on the original clean form of the Carioca 1 database compared with the techniques in [134] and [139], especially in scenarios with low SNR and nonlinear digital saturation.

In [144], the authors presented a method for detecting audio manipulation based on autoregressive (AR) modeling. Their approach first exposes the extracted ENF signal to a wavelet filter to generate and highlight more detailed ENF variations, which are then used for AR modeling. The generated AR coefficients were utilized in the training process of the SVM classifier as input features to detect manipulation. The authors reported that their proposed method significantly outperformed the existing work method proposed in [140] under noisy conditions and offered resilience to MP3 compression.

In [145] and [146], a method that uses the maximum offset for cross-correlation (MOCC) to detect audio authenticity was proposed. This method involves comparing the ENF signal obtained from a questioned audio recording with the reference signal. The obtained ENF signal is split into blocks that overlap, and an enhancement scheme is used to improve the quality of the ENF signal. The MOCC between the questioned audio recording's ENF and the reference ENF was computed, and the changes in the MOCCs from different

blocks were compared and used to determine audio authenticity. According to the authors, their suggested method not only detects the boundaries of the edited region but also recognizes the type of forgery that has occurred.

In [147], the authors suggested an ENF-based algorithm for detecting inter-frame video forgery in static video recording. Given a suspected video, the algorithm retrieves its ENF and employs cubic spline interpolation to address data deficits in the retrieved ENF. From the interpolated ENF signal, the algorithm first obtains the CC between every consecutive period, and then utilizes the sharp decreases in the value of the CC to identify the presence and location of the forgery. Their experimental findings demonstrated the effectiveness of the proposed method in identifying inter-frame video forgeries, including frame deletion, frame insertion, and frame duplication. Additional experiments indicated that when it comes to identifying inter-frame forgery in static scene surveillance videos, the proposed algorithm outperformed several other contemporary algorithms [148], [149].

In [150], the authors presented a technique for identifying audio tampering that relies on the consistency of the ENF component (ENFC). From the extracted ENFC, the phase features and instantaneous frequency features were extracted using DFT [135] the and Hilbert transform, respectively. The feature set represents a measure of the phase magnitude as well as the instantaneous frequency changes of the ENFC, and as a pointer to the consistency of the ENFC. After extracting these features, they were utilized as inputs to an SVM classifier to detect tampering. Experiments demonstrated that the proposed method provides a high degree of classification precision. However, adding Gaussian white noise reduces the precision of the proposed technique for signals with an SNR of less than 10 dB.

Recaptured audio recordings contain two ENF signals [151] the ENF signal picked up during the original recording (content ENF signal) and the ENF signal picked up during the recapturing process (recaptured ENF signal). If the recapturing of the recording takes place in a region with the same nominal ENF value as that of the original recording, then it is possible that the ENF patterns from both recording operations will overlap. The two signals may have different intensities, where higher and lower intensities are regarded as the dominant ENF and latent ENF respectively. Conventional ENF estimation algorithms, as demonstrated in the study of [151], can only extract the dominant ENF signal. Because audio recapturing may potentially be employed by an adversary as an anti-forensic strategy for manipulating ENF traces to misguide forensic investigators, approaches to extract various overlapping ENF signals are necessary to complement current strategies against such anti-forensic operations [137].

The authors of [151] and [152] introduced approaches to tackle the challenge of dealing with recaptured audio signals. In [151], the authors presented a decorrelation algorithm to retrieve ENF patterns in a more difficult case, in which the two patterns overlap each other in frequency, employing the assistance of a power reference signal. Assuming the availability of the power grid reference data, the suggested decorrelation algorithm may be utilized to detect audio recapture, that is, to determine whether a particular audio recording is a genuine or recaptured copy. In [152], the author described an approach for discerning whether a recording is genuine or recaptured, which relies on a convolutional neural network (CNN). The CNN uses spectral characteristics derived from the fundamental ENF and its harmonics. The research also examined how the analysis window impacts the effectiveness of the approach and uses intermediate feature maps to obtain an understanding of the learning and decision-making process of the CNN.

D. CAMERA AND DEVICE IDENTIFICATION

Researchers have also developed ENF-based applications that employ ENF pattern implanted in audio or video recordings to identify the device model or camera used to acquire the audio or video chip. For audio recordings, certain signal-derived features that allow for the differentiation of various devices can be employed with a machine learning algorithm to identify the recording device model. In [153], the author implemented a machine learning solution that is capable of learning the distinctive features of ENF signals from a variety of devices and utilizes the learned features of the ENF signals for its categorization. A system as such this might potentially determine the device that captures an ENF signal without the requirement for simultaneous power references. The author used various recorders, including six mobile phones and one tablet, to conduct the audio recording. The harmonic power coefficients were estimated from the extracted ENF signals and were utilized as features to train an SVM classifier. The evaluation findings suggest that the proposed method is effective.

For video recordings, an analysis can be performed to distinguish the inherent characteristics of the acquired camera. Research in this area has concentrated on CMOS cameras equipped with rolling shutters, with an emphasis on estimating the read-out time T_{ro} of the camera. The T_{ro} refers to the duration required by the camera to capture the rows of a solitary frame, and it is a camera-specific parameter that can be leveraged to characterize CMOS cameras with a rolling shutter mechanism. It is not usually listed in the user manual or specification catalog of most cameras and is typically shorter than the frame period [91].

Some studies have exploited T_{ro} on flicker-based video forensics, which tackled concerns relevant to investigations into movie piracy in the entertainment industry [80], [154]. This study focused on pirated videos created by filming a video displayed on an LCD screen using a camcorder. Such pirated recordings usually display a unique artifact known as the flicker signal, which results from the interaction between the LCD screen's backlight and the video camera's recording mechanism. In [80], the authors presented several estimation methods that analyzed flicker signals and estimated their frequencies and the T_{ro} value. These methods are used to

identify the specific camera and LCD screen used to produce pirated video.

Similar to the flicker signal, the ENF signal is a trace that can be inherently implanted in a video, owing to the camera's recording process and signal existence within the recording scene. The LCD screen's backlight serves as the environmental signal for flicker signals, whereas the electric lighting signal in the recording scene triggers the ENF signal. Inspired by the similarity between the two signals, the authors of [97] suggested a nonintrusive approach that utilizes the ENF recorded in a video to identify or describe the camera used to create the video by computing the T_{ro} of the camera. Their approach essentially estimates the T_{ro} of the camera for each frame using the vertical phase analysis modeled as:

$$T_{ro} = \frac{L\omega_b}{2\pi f_e} \tag{8}$$

where L denotes frame's number of rows, ω_b represents the vertical radial frequency which is calculated from the vertical phase line's slope, and f_e represents the ENF component that fluctuates about the nominal frequency. The slope of the vertical phase line may be acquired by estimating the ENF phases $\emptyset[\ell](\ell \in \{1, 2, 3, 4, 5, \ldots, L\})$ for every one of each row. The $\emptyset[\ell]$ is extracted by applying the Fourier Transform to the time series of the l th-row, which can be generated by calculating the average intensity values of the lth row from every video frame. The proposed approach was tested by employing ENF-containing videos captured by five distinct cameras. The experimental results demonstrated that the approach attained great precision in computing the T_{ro} value, with a relative estimation error of less than 1.5%. However, the suggested approach relies on alias ENF, and thus, may be ineffective when the frame rate of the camera that obtains the video is a divisor of the ENF nominal value (0 Hz alias ENF).

To address this limitation, the authors in [99] introduced a method that can handle cases where the video camera's frame rate can cause a 0Hz ENF. Relying on a model that shifts the nominal illumination harmonic to various frequencies based on idle period length. The approach evaluates the two frequency locations with the highest power where the ENF components appear and estimates T_{ro} based on the power ratio between the two components.

E. MULTIMEDIA SYNCHRONIZATION

Synchronizing multimedia data is required to maintain the original relationship between heterogeneous multimedia data, ensuring that they remain synchronized before their final presentation [155]. Traditional techniques that synchronize multimedia content involve either passively or actively measuring timestamps or finding similar related information from two or multiple recordings. In audio synchronization, contextual information such as voice and music can be utilized. On the other hand, overlapping visual sequences, even those obtained from various viewing viewpoints, can be used to synchronize videos [43].

The ENF-based approach to synchronization, involves aligning the implanted ENF signals of audio and visual tracks to synchronize the recordings. As it is not dependent on audio or visual data from multimedia signals, it can complement traditional synchronization procedures and may assist in addressing issues that would otherwise be intractable. Traditional synchronization techniques require dedicated hardware, whereas synchronization solutions relying on collected ENF traces do not require dedicated hardware. However, for the ENF-based synchronization procedure to be effective, the ENF signal contained in the audio and visual track recordings must have sufficient strength to permit the estimation of reliable ENF signals. This approach is mainly applicable in use cases where several recordings need to be synchronized, either because they overlap in time or because they do not. If numerous recordings from one power grid are captured with an overlap in time, ENF can synchronize them without difficulty.

The authors of [92], [93], and [94] proposed approaches for synchronizing videos by aligning intrinsically embedded ENF signals. When a video includes both visual and audio tracks, either modality can be used as the source of synchronization ENF traces. When an audio track is unavailable, as is sometimes the case in video surveillance settings, it becomes crucial to use a visual track for multimedia synchronization [93]. The use of ENF to synchronize audio streams from wireless and inexpensive USB sound adaptors was demonstrated by the authors in [156]. When using multiple sound cards, it is common for streams to be out of synchronization because of the differences in the sample buffers filling rates of the sound cards themselves.

If there was no time overlap between recordings, reference ENF databases were necessary to precisely pinpoint the time every track was recorded. If the source location is known, each recording must match a single or more reference databases to calculate its timestamp [43].

F. VIDEO AUTHENTICATION

The use of the Internet of Video Things (IoVT) is now an integral part of smart city infrastructure, where situation awareness (SAW) is crucial for monitoring and managing cities [157]. As IoVT systems are increasingly being deployed, a large amount of visual data is generated and executed at each instant. However, with the advancement of artificial intelligence (AI) technology, it has become easier to create fake video and audio feeds as well as adulterated images, which can deceive smart city security operators [158]. Therefore, verifying the authenticity of visual and audio feeds is essential for ensuring safety and security. One way to achieve this is to use an ENF signal obtained from the power grid, which is a reliable signature for authentication. In [159], the authors suggested a method for video authentication using an ENF and steady superpixels. This method is known as the EVAS. To improve the segmentation performance, memory efficiency, and computational speed, the authors utilized the SLIC algorithm. To avoid the

effects of ENF estimation variations, video super-pixels were created by grouping pixels with consistent intensities and textures. The EVAS method authenticates a video by comparing the estimated ENF with the ground-truth ENF through reference-matching. However, in surveillance camera video recording, moving objects may cause disruptions in illumination samples, leading to an imprecise estimate of the ENF. The authors overcame this challenge using Selective Superpixel Masking, which compensates for occlusions attributed to moving objects. The proposed method was implemented on an edge-based platform to authenticate and monitor indoor surveillance using the reference data.

G. DEEPFAKE DETECTION

Although advanced end-to-end encryption models are offered by contemporary video conferencing technologies, they do not authenticate the media broadcast through them, leaving the responsibility of verification to end users [160]. This creates a vulnerability, as an attacker can use replay or deepfake attacks to fabricate audio or video streams and manipulate the perception of real-time transcribed events. The progress made in generative models, including deepfake technology, enables attackers to mimic a specific individual and manipulate online interactions. The successful execution of such an attack may cause disinformation, which has the potential to disrupt society and erode the fundamental basis of trust [161].

In [162], the authors introduced DeFakePro, a method for detecting deepfakes in online video conferencing platforms, using a decentralized consensus mechanism. This technique utilizes the ENF present in digital media recordings as a unique environmental signature, which is used in the proof-of-ENF (PoENF) algorithm to establish consensus. The PoENF algorithm compares the variations in the ENF signal to validate media broadcasts on conferencing platforms. DeFakePro can detect deepfake video recordings broadcast by malicious participants through video conferencing settings. The system ensured the validity of all incoming media on both the audio and video channels.

VII. CHALLENGES AND FUTURE WORK

1) In ENF-based forensic studies, several authors [4], [8], [22], [30], [72], [112], [163] reported that the length of a questioned recording should be 10 min or longer to guarantee improved ENF extraction and reliable matching. This poses a great challenge in the applicability of ENF because the length of audio evidence in practical scenarios can be as short as 30 s or less. In general, the ENF analysis performance drops dramatically when the length of the questioned recording is below 10 min because the randomness or uniqueness of the ENF pattern weakens with shorter durations. However, the studies in [164] and [165] demonstrated that ENF may be detected in somewhat short recordings, such as 2 min, indicating the feasibility of conducting credible ENF estimates and subsequent forensic analysis on short recordings. Accordingly, future research should focus on developing and utilizing high-resolution frequency-estimation systems for improved ENF detection and estimation with smaller frame sizes.

2) The approaches proposed in the literature for inferring both inter-grid and intra-grid localization utilize the power ENF data. The ENF estimations derived from a power signal are regarded as its purest and most easily accessible form. However, multimedia ENF data prove more challenging to work with because of its noisy nature compared with power ENF data. The noisy ENF signal in multimedia data potentially makes the localization process challenging because state-of-the-art approaches use the high-resolution frequency variations of the ENF signal to construct useful metrics for localization. We noted that reducing the frame size to 1 s caused an even greater decrease in the SNR of the audio ENF signals. Nevertheless, a frame size of 1 s is considered ideal to achieve acceptable localization accuracy [127]. The noise sensitivity analysis reported by the authors in [131] showed that a questioned signal SNR greater than 20 dB is required for a satisfactory localization performance. In light of these findings, it is challenging to use ENF signals in multimedia recordings for localization, by utilizing current ENF estimation methods as well as localization feature extraction approaches. To extract an ENF signal with high a SNR from media recordings, a breakthrough is needed in ENF estimation approaches.

3) We noted that the ENF-based approaches proposed in the literature for tampering detection can handle insertion, deletion, and splicing attacks, but only when a lone attack occurs at a single point inside an audio recording. In addition to from the limitations of the current methods, there are challenges in handling a combination of several attacks such as insertion, deletion, splicing, replacement, and shuffling, which can sometimes appear at multiple points. More investigations on the effective use of ENF for audio tampering detection are needed, although the solution presented in [132] is more of a decent starting point than a completely dependable solution. Such studies may not utilize the assumption of discontinuity at tampered points as the only criterion but may consider other possible ways to effectively exploit accessible real-life data from both the questioned audio file and the reference.

4) Detecting the ENF is a critical initial step in ENF-based multimedia authentication systems, which has proven challenging using signal processing and detection theory [62]. For decades, the generic and hypothetical answer to the fundamental question of what places/locations can a recording device capture the ENF has been that the recording device should be near ENF sources, which still needs some kind of practical guideline. Therefore, there is a need for other methods that can be used to promote ENF detection from a forensic standpoint. As noted by the authors in [62], an approach that may practically answer this question is to actively measure the strength of the ENF signal at different spots. The sensor of the measuring device, for instance, a modified sound-level meter, should exclusively measure the ENF bandwidth. With such an ENF level measuring meter, the strength of the ENF

signal at different location such as restaurants, supermarkets, cafes, classrooms, quiet office rooms, convenience stores, lecture halls, home rooms, event trains, can be widely measured. The selection of the spots to measure is influenced by the environmental noise level, interior space size, type of likely media material to be captured, and typical objects occupying the area, among other factors. With sufficient experimentation, we can learn the strength of the ENF signal in a particular or similar sort of spot. Law enforcement authorities can exploit such experimental information when media evidence is noted to have been created at a measured spot. Therefore, instead of treating a questioned audio file from the signal processing viewpoint as interference, it was carefully examined to determine the recording spot. This will allow experimental information regarding the strength of the ENF in that spot to be utilized. Even though it is supplementary support, this idea and further research in this direction can usually offer relevant forensic clues in practice.

5) ENF analysis can be exploited in new cases beyond audio and video forensics. For example, fine localization fingerprints in ENF can be employed to create secure connected autonomous Internet of Things (IoT) and cyber physical systems (CPS) [166]. One possible scenario that could occur involves utilizing ENF patterns captured by specialized sensors used in IoT/CPS applications to aid in authenticating location, either by authenticating location in security tokens or by providing ''Proof-Carrying Sensing'' for CPS. For example, in the case of IoT/CPS deployment, each connected device may contain a photodiode covering the ENF frequency range, which is relevant for indoor sensing. There could also be acoustic sensors that are designed to record ENF with a high SNR or sensors that capture ENF from devices plugged into the power mains. Employing the approach for inferring the location proposed in [131], the devices may use the high SNR ENF from each other to authenticate the location of each device after validating the reported location with the estimated location, which utilizes both the ENF signal of the device and the ENF signal of grid node.

VIII. CONCLUSION

This paper has provided a summary of research endeavors pertaining to the ENF signal, a random and unique identifier that may be acquired from multimedia recordings produced in areas where electrical activity exists. Initially, we discussed the techniques and sensor hardware for recording ENF reference data because for us to rely on the outcome of the application of the ENF signal obtained from media recordings, we must verify that the signal we extracted is actually the ENF signal. Subsequently, we discuss ENF detection in media recordings. We note that it is dangerous to perform an analysis assuming that the ENF signal is successfully captured, as this will negatively impact the related investigation. We further explored the process of extracting ENF signals from audio and video recordings. The various factors that influence the capture of ENF signals in both audio and video recordings were also analyzed. Next, we review the potential applications of the ENF fingerprint for media file recording time verification, tamper detection, location authentication, recording sensor/device identification, and multimedia synchronization. We also investigated applications for video authentication and DeepFake detection. Finally, we highlight the current challenges in the ENF analysis and suggest future research directions.

REFERENCES

[1] L. Jiang, W. Wu, C. Qian, and C. C. Loy, ''DeepFakes detection: The deeper forensics dataset and challenge,'' in *Handbook of Digital Face Manipulation and Detection*, vol. 303. Cham, Switzerland: Springer, 2022.

[2] *Media Forensics (MediFor)*. Accessed: Nov. 2022. [Online]. Available: https://www.darpa.mil/program/media-forensics

[3] *Semantic Forensics (SemaFor)*. Accessed: Nov. 2022. [Online]. Available: https://www.darpa.mil/program/semantic-forensics

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

[5] C. Grigoras, ''Applications of ENF criterion in forensic audio, video, computer and telecommunication analysis,'' *Forensic Sci. Int.*, vol. 167, nos. 2–3, pp. 136–145, Apr. 2007.

[6] M. Kajstura, A. Trawinska, and J. Hebenstreit, ''Application of the electrical network frequency (ENF) criterion,'' *Forensic Sci. Int.*, vol. 155, nos. 2–3, pp. 165–171, Dec. 2005.

[7] A. J. Cooper, ''An automated approach to the electric network frequency (ENF) criterion: Theory and practice,'' *Int. J. Speech Lang. Law*, vol. 16, no. 2, pp. 193–218, Apr. 2010.

[8] M. Huijbregtse and Z. Geradts, ''Using the ENF criterion for determining the time of recording of short digital audio recordings,'' in *Proc. 3rd IWCF*, vol. 1, Aug. 2009, pp. 116–124.

[9] R. W. Sanders, ''Digital audio authenticity using the electric network frequency,'' in *Proc. Audio Eng. Soc. Conf., 33rd Int. Conf., Audio Forensics-Theory Pract., Audio Eng. Soc.*, Denvor, CO, USA, Jun. 2008, pp. 5–7.

[10] E. B. Brixen, ''Techniques for the authentication of digital audio recordings,'' in *Audio Engineering Society Convention*. New York, NY, USA: Audio Eng. Soc., May 2007.

[11] R. Garg, A. L. Varna, and M. Wu, '''Seeing' ENF: Natural time stamp for digital video via optical sensing and signal processing,'' in *Proc. 19th ACM Int. Conf. Multimedia*, Scottsdale, AZ, USA, Nov. 2011, pp. 23–32.

[12] 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. Inf. Forensics Security*, vol. 8, no. 9, pp. 1417–1432, Sep. 2013.

[13] C. W. Wong, A. Hajj-Ahmad, and M. Wu, ''Invisible geo-location signature in a single image,'' in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 1987–1991.

[14] E. B. Brixen, ''Further investigation into the ENF criterion for forensic authentication,'' in *Audio Engineering Society Convention*, vol. 123. New York, NY, USA: Audio Eng. Soc., Oct. 2007.

[15] E. B. Brixen, ''ENF; quantification of the magnetic field,'' in *Proc. 33rd Int. Conf., Audio Forensics-Theory Pract.* New York, NY, USA: Audio Eng. Soc., 2008, pp. 1–12.

[16] A. J. Cooper, ''The electric network frequency (ENF) as an aid to authenticating forensic digital audio recordings—An automated approach,'' in *Proc. Audio Eng. Soc. 33rd Conf., Audio Forensics-Theory Pract.*, Jun. 2008, pp. 1–10.

[17] A. J. Cooper, ''Digital audio recordings analysis: The electric network frequency (ENF) criterion,'' *Int. J. Speech Lang. Law*, vol. 16, no. 2, pp. 193–218, 2009.

[18] A. J. Cooper, ''Further considerations for the analysis of ENF data for forensic audio and video applications,'' *Int. J. Speech Lang. Law*, vol. 18, no. 1, pp. 99–120, Sep. 2011.

[19] R. Sanders and P. S. Popolo, ''Extraction of electric network frequency signals from recordings made in a controlled magnetic field,'' in *Audio Engineering Society Convention*, vol. 125. New York, NY, USA: Audio Eng. Soc., Oct. 2008.

[20] D. P. Nicolalde and J. A. Apolinario, "Evaluating digital audio authenticity with spectral distances and ENF phase change," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Apr. 2009, pp. 1417–1420.

[21] M. M. Eissa, M. M. Elmesalawy, Y. Liu, and H. Gabbar, "Wide area synchronized frequency measurement system architecture with secure communication for 500 kV/220 kV Egyptian grid," in *Proc. Int. Conf. Smart Grid (SGE)*, Aug. 2012, pp. 1–10.

[22] G. Hua, J. Goh, and V. L. L. Thing, "A dynamic matching algorithm for audio timestamp identification using the ENF criterion," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 7, pp. 1045–1055, Jul. 2014.

[23] J. Chai, F. Liu, Z. Yuan, R. W. Conners, and Y. Liu, "Source of ENF in battery-powered digital recordings," *Audio Eng. Soc. Conv.*, vol. 135, pp. 1–6, Oct. 2013.

[24] M. Bollen and I. Gu, *Signal Processing of Power Quality Disturbances*. Hoboken, NJ, USA: Wiley, 2006.

[25] B. M. Weedy, B. J. Cory, N. Jenkins, J. B. Ekanayake, and G. Strbac, *Electric Power Systems*. Hoboken, NJ, USA: Wiley, 2012.

[26] T. Baksteen, "The electrical network frequency criterion: Determining the time and location of digital recordings," MSc. thesis, Dept. Appl. Maths, Fac. Elect. Eng., Maths, Comput. Sci., Delft Univ. Tech., Delft, The Netherlands, 2015.

[27] B. Gerazov, Z. Kokolanski, G. Arsov, and V. Dimcev, "Tracking of electrical network frequency for the purpose of forensic audio authentication," in *Proc. 13th Int. Conf. Optim. Electr. Electron. Equip. (OPTIM)*, May 2012, pp. 1164–1169.

[28] Y. Zhang, P. Markham, T. Xia, L. Chen, Y. Ye, Z. Wu, Z. Yuan, L. Wang, J. Bank, J. Burgett, R. W. Conners, and Y. Liu, "Wide-area frequency monitoring network (FNET) architecture and applications," *IEEE Trans. Smart Grid*, vol. 1, no. 2, pp. 159–167, Sep. 2010.

[29] Y. Liu, Z. Yuan, P. N. Markham, R. W. Conners, and Y. Liu, "Application of power system frequency for digital audio authentication," *IEEE Trans. Power Del.*, vol. 27, no. 4, pp. 1820–1828, Oct. 2012.

[30] Y. Liu, C. Yao, C. Sun, and Y. Liu, "The authentication of digital audio recordings using power system frequency," *IEEE Potentials*, vol. 33, no. 2, pp. 39–42, Mar. 2014.

[31] A. G. Phadke, J. S. Thorp, and M. G. Adamiak, "A new measurement technique for tracking voltage phasors, local system frequency, and rate of change of frequency," *IEEE Power Eng. Rev.*, vol. PER-3, no. 5, p. 23, May 1983.

[32] Z. Zhong, C. Xu, B. J. Billian, L. Zhang, S.-J.-S. Tsai, R. W. Conners, V. A. Centeno, A. G. Phadke, and Y. Liu, "Power system frequency monitoring network (FNET) implementation," *IEEE Trans. Power Syst.*, vol. 20, no. 4, pp. 1914–1921, Nov. 2005.

[33] S.-J. Tsai, L. Zhang, A. G. Phadke, Y. Liu, M. R. Ingram, S. C. Bell, I. S. Grant, D. T. Bradshaw, D. Lubkeman, and L. Tang, "Frequency sensitivity and electromechanical propagation simulation study in large power systems," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 54, no. 8, pp. 1819–1828, Aug. 2007.

[34] *FNET Server Web Display*. Accessed: Nov. 2022. [Online]. Available: http://fnetpublic.utk.edu

[35] M. M. Elmesalawy and M. M. Eissa, "New forensic ENF reference database for media recording authentication based on harmony search technique using GIS and wide area frequency measurements," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 4, pp. 633–644, Apr. 2014.

[36] H. Kim, Y. Jeon, and J. W. Yoon, "Construction of a national scale ENF map using online multimedia data," in *Proc. ACM Conf. Inf. Knowl. Manag.*, Nov. 2017, pp. 19–28.

[37] P. Top, M. R. Bell, E. Coyle, and O. Wasynczuk, "Observing the power grid: Working toward a more intelligent, efficient, and reliable smart grid with increasing user visibility," *IEEE Signal Process. Mag.*, vol. 29, no. 5, pp. 24–32, Sep. 2012.

[38] A. Hajj-Ahmad, R. Garg, and M. Wu, "Spectrum combining for ENF signal estimation," *IEEE Signal Process. Lett.*, vol. 20, no. 9, pp. 885–888, Sep. 2013.

[39] N. Fechner and M. Kirchner, "The humming hum: Background noise as a carrier of ENF artifacts in mobile device audio recordings," in *Proc. 8th Int. Conf. IT Secur. Incident Manage. IT Forensics*, May 2014, pp. 3–13.

[40] V. A. Nita, R. A. Dobre, A. Drumea, A. Ciobanu, C. Negrescu, and D. Stanomir, "Electrical network signal's waveform and frequency logging for forensic," in *Proc. 9th Int. Symp. Adv. Topics Electr. Eng. (ATEE)*, May 2015, pp. 156–161.

[41] D. K. O. Jones, D. D. L. E. L. Hamilton, C. Robinson, J. T. Reed-Jones, and K. Morrison, "An electric network frequency analysis technology demonstrator for educational purposes," in *Proc. IEEE Conf., Rec.*, vol. 52438, Sep. 2021, pp. 16–17.

[42] V. A. Nita, A. Ciobanu, C. Negrescu, and D. Stanomir, "Low cost electric network signal and frequency recorder," in *Proc. 10th Int. Symp. Adv. Topics Electr. Eng. (ATEE)*, Mar. 2017, pp. 13–16.

[43] A. Hajj-Ahmad, C. W. Wong, J. Choi, and M. Wu, "Power signature for multimedia forensics," in *Multimedia Forensics*, H. T. Sencar, L. Verdoliva, and N. Memon, Eds. Singapore: Springer, 2022, pp. 235–280.

[44] T. E. Gemayel, "Feasibility of using electrical network frequency fluctuations to perform forensic digital audio authentication," M.S. thesis, Dept. Elect. Eng. Comput. Sci., Univ. Ottawa, Ottawa, ON, Canada, Tech. Rep., 2013.

[45] C. Grigoras, "Applications of ENF analysis in forensic authentication of digital audio and video recordings," *J. Audio Eng. Soc.*, vol. 57, no. 9, pp. 643–661, 2009.

[46] A. Hajj-Ahmad, R. Garg, and M. Wu, "Instantaneous frequency estimation and localization for ENF signals," in *Proc. Asia Pacific Signal Inf. Process. Assoc. Annu. Summit Conf.*, Dec. 2012, pp. 1–10.

[47] S. Haykin, *Advances in Spectrum Analysis and Array Processing*. Upper Saddle River, NJ, USA: Prentice-Hall, 1991.

[48] J. Weinrib, "An exploration of ENF signal detection," Doctoral dissertation, The Cooper Union for the Advancement of Science and Art, Dept. Elect. Eng., Albert Nerken School Eng., New York, NY, USA, 2019.

[49] J. O. Smith and X. Serra, "PARSHL: An analysis/synthesis program for non-harmonic sounds based on a sinusoidal representation," in *Proc. Int. Comput. Music Conf. (ICMC)*, Aug. 1987, pp. 290–297.

[50] O. Ojowu, J. Karlsson, J. Li, and Y. Liu, "ENF extraction from digital recordings using adaptive techniques and frequency tracking," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 4, pp. 1330–1338, Aug. 2012.

[51] G. O. Glentis and A. Jakobsson, "Time-recursive IAA spectral estimation," *IEEE Signal Process. Lett.*, vol. 18, no. 2, pp. 111–114, Feb. 2011.

[52] D. G. Manolakis, V. K. Ingle, and S. M. Kogon, *Statistical and Adaptive Signal Processing*. New York, NY, USA: McGraw-Hill, 2000.

[53] C. Offelli, C. Narduzzi, D. Petri, and F. E. Zanin, "Comparative analysis of non-parametric and parametric frequency estimation methods," in *Proc. IEEE Instrum. Meas. Technol. Conf.*, May 1993, pp. 603–607.

[54] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. AP-34, no. 3, pp. 276–280, Mar. 1986.

[55] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 984–995, Jul. 1989.

[56] D. Spielman, A. Paulraj, and T. Kailath, "Performance analysis of the MUSIC algorithm," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Apr. 1986, pp. 1909–1912.

[57] A. Barabell, "Improving the resolution performance of eigenstructure-based direction-finding algorithms," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Apr. 1983, pp. 336–339.

[58] Y. P. Liao and A. Abouzaid, "Resolution improvement for MUSIC and ROOT MUSIC algorithms," *J. Inf. Hiding Multimedia Signal Process.*, vol. 6, no. 2, pp. 189–197, Mar. 2015.

[59] G. Karantaidis and C. Kotropoulos, "Assessing spectral estimation methods for electric network frequency extraction," in *Proc. 22nd Pan-Hellenic Conf. Informat.*, Nov. 2018, pp. 202–207.

[60] A. Hajj-Ahmad, C.-W. Wong, S. Gambino, Q. Zhu, M. Yu, and M. Wu, "Factors affecting ENF capture in audio," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 2, pp. 277–288, Feb. 2019.

[61] S. Vatansever, A. E. Dirik, and N. Memon, "Factors affecting ENF based time-of-recording estimation for video," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 2497–2501.

[62] 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.

[63] G. Hua, H. Liao, Q. Wang, H. Zhang, and D. Ye, "Detection of electric network frequency in audio recordings—From theory to practical detectors," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 236–248, 2021.

[64] H. Liao, G. Hua, and H. Zhang, "ENF detection in audio recordings via multi-harmonic combining," *IEEE Signal Process. Lett.*, vol. 28, pp. 1808–1812, 2021.

[65] S. Vatansever, A. E. Dirik, and N. Memon, "Detecting the presence of ENF signal in digital videos: A superpixel-based approach," *IEEE Signal Process. Lett.*, vol. 24, no. 10, pp. 1463–1467, Oct. 2017.

[66] J. Choi, C.-W. Wong, A. Hajj-Ahmad, M. Wu, and Y. Ren, "Invisible geolocation signature extraction from a single image," *IEEE Trans. Inf. Forensics Security*, vol. 17, pp. 2598–2613, 2022.

[67] T. Yardibi, J. Li, P. Stoica, M. Xue, and A. B. Baggeroer, "Source localization and sensing: A nonparametric iterative adaptive approach based on weighted least squares," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 46, no. 1, pp. 425–443, Jan. 2010.

[68] U. Jönsson, C. Trygger, and P. Ogren, "Lecture notes on optimal control: Optimization and system theory," unpublished

[69] Q. Zhu, M. Chen, C.-W. Wong, and M. Wu, "Adaptive multi-trace carving based on dynamic programming," in *Proc. 52nd Asilomar Conf. Signals, Syst., Comput.*, Oct. 2018, pp. 1716–1720.

[70] Q. Zhu, M. Chen, C.-W. Wong, and M. Wu, "Adaptive multi-trace carving for robust frequency tracking in forensic applications," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 1174–1189, 2021.

[71] L. Fu, P. N. Markham, R. W. Conners, and Y. Liu, "An improved discrete Fourier transform-based algorithm for electric network frequency extraction," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 7, pp. 1173–1181, Jul. 2013.

[72] L. Dosiek, "Extracting electrical network frequency from digital recordings using frequency demodulation," *IEEE Signal Process. Lett.*, vol. 22, no. 6, pp. 691–695, Jun. 2015.

[73] G. Karantaidis and C. Kotropoulos, "Efficient Capon-based approach exploiting temporal windowing for electric network frequency estimation," in *Proc. IEEE 29th Int. Workshop Mach. Learn. Signal Process. (MLSP)*, Oct. 2019, pp. 1–6.

[74] G. Karantaidis and C. Kotropoulos, "Blackman–Tukey spectral estimation and electric network frequency matching from power mains and speech recordings," *IET Signal Process.*, vol. 15, no. 6, pp. 396–409, Aug. 2021.

[75] X. Lin and X. Kang, "Robust electric network frequency estimation with rank reduction and linear prediction," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 14, no. 4, pp. 1–13, Nov. 2018.

[76] D. Bykhovsky and A. Cohen, "Electrical network frequency (ENF) maximum-likelihood estimation via a multitone harmonic model," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 5, pp. 744–753, May 2013.

[77] W. Zhong, X. You, X. Kong, and B. Wang, "Electric network frequency estimation based on linear canonical transform for audio signal authentication," in *Proc. 25th Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2017, pp. 286–290.

[78] G. Hua and H. Zhang, "ENF signal enhancement in audio recordings," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 1868–1878, 2020.

[79] G. Hua, H. Liao, H. Zhang, D. Ye, and J. Ma, "Robust ENF estimation based on harmonic enhancement and maximum weight clique," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 3874–3887, 2021.

[80] A. Hajj-Ahmad, S. Baudry, B. Chupeau, and G. Doërr, "Flicker forensics for pirate device identification," in *Proc. 3rd ACM Workshop Inf. Hiding Multimedia Secur.*, Jun. 2015, pp. 75–84.

[81] G. Karantaidis and C. Kotropoulos, "An automated approach for electric network frequency estimation in static and non-static digital video recordings," *J. Imag.*, vol. 7, no. 10, p. 202, Oct. 2021.

[82] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.

[83] D. Stutz, A. Hermans, and B. Leibe, "SuperPixels: An evaluation of the state-of-the-art," *Comput. Vis. Image Understand.*, vol. 166, pp. 1–27, Jan. 2018.

[84] S. Fernandez-Menduina and F. Pérez-González, "ENF moving video database," Tech. Rep., Jan. 2020, doi: 10.5281/zenodo.3549378.

[85] S. Fernández-Menduiña and F. Pérez-González, "Temporal localization of non-static digital videos using the electrical network frequency," *IEEE Signal Process. Lett.*, vol. 27, pp. 745–749, 2020.

[86] O. Barnich and M. Van Droogenbroeck, "ViBe: A universal background subtraction algorithm for video sequences," *IEEE Trans. Image Process.*, vol. 20, no. 6, pp. 1709–1724, Jun. 2011.

[87] W. C. Lindsey and C. M. Chie, "A survey of digital phase-locked loops," *Proc. IEEE*, vol. 69, no. 4, pp. 410–431, Apr. 1981.

[88] J. Gu, Y. Hitomi, T. Mitsunaga, and S. Nayar, "Coded rolling shutter photography: Flexible space-time sampling," in *Proc. IEEE Int. Conf. Comput. Photography (ICCP)*, Mar. 2010, pp. 1–8.

[89] C.-K. Liang, L.-W. Chang, and H. H. Chen, "Analysis and compensation of rolling shutter effect," *IEEE Trans. Image Process.*, vol. 17, no. 8, pp. 1323–1330, Aug. 2008.

[90] O. Ait-Aider, A. Bartoli, and N. Andreff, "Kinematics from lines in a single rolling shutter image," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Minneapolis, MN, USA, Jun. 2007, pp. 1–6.

[91] H. Su, A. Hajj-Ahmad, R. Garg, and M. Wu, "Exploiting rolling shutter for ENF signal extraction from video," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Piscataway, NJ, USA, Oct. 2014, pp. 5367–5371.

[92] H. Su, A. Hajj-Ahmad, M. Wu, and D. W. Oard, "Exploring the use of ENF for multimedia synchronization," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2014, pp. 4613–4617.

[93] H. Su, A. Hajj-Ahmad, C.-W. Wong, R. Garg, and M. Wu, "ENF signal induced by power grid: A new modality for video synchronization," in *Proc. 2nd ACM Int. Workshop Immersive Media Exper.*, Orlando, FL, USA, Nov. 2014, pp. 13–18.

[94] K. Vidyamol, E. George, and J. P. Jo, "Exploring electric network frequency for joint audio-visual synchronization and multimedia authentication," in *Proc. Int. Conf. Intell. Comput., Instrum. Control Technol. (ICICICT)*, Kerala, India, Jul. 2017, pp. 240–246.

[95] J. Choi and C.-W. Wong, "ENF signal extraction for rolling-shutter videos using periodic zero-padding," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 2667–2671.

[96] J. Choi, C.-W. Wong, H. Su, and M. Wu, "Analysis of ENF signal extraction from videos acquired by rolling shutters," *IEEE Trans. Inf. Forensics Security*, vol. 18, pp. 4229–4242, 2023.

[97] A. Hajj-Ahmad, A. Berkovich, and M. Wu, "Exploiting power signatures for camera forensics," *IEEE Signal Process. Lett.*, vol. 23, no. 5, pp. 713–717, May 2016.

[98] H. Han, Y. Jeon, B.-K. Song, and J. W. Yoon, "A phase-based approach for ENF signal extraction from rolling shutter videos," *IEEE Signal Process. Lett.*, vol. 29, pp. 1724–1728, 2022.

[99] S. Vatansever, A. E. Dirik, and N. Memon, "Analysis of rolling shutter effect on ENF-based video forensics," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 9, pp. 2262–2275, Sep. 2019.

[100] P. Ferrara, I. Sanchez, G. Draper-Gil, H. Junklewitz, and L. Beslay, "A MUSIC spectrum combining approach for ENF-based video timestamping," in *Proc. IEEE Int. Workshop Biometrics Forensics (IWBF)*, Rome, Italy, May 2021, pp. 1–6.

[101] G. Frijters and Z. J. M. H. Geradts, "Use of electric network frequency presence in video material for time estimation," *J. Forensic Sci.*, vol. 67, no. 3, pp. 1021–1032, May 2022.

[102] S. Vatansever and A. E. Dirik, "Forensic analysis of digital audio recordings based on acoustic mains hum," in *Proc. 24th Signal Process. Commun. Appl. Conf. (SIU)*, May 2016, pp. 1285–1288.

[103] S. Vatansever, A. E. Dirik, and N. Memon, "The effect of inverse square law of light on ENF in videos exposed by rolling shutter," *IEEE Trans. Inf. Forensics Security*, vol. 18, pp. 248–260, 2023.

[104] Y. Liu, Z. Yuan, P. N. Markham, R. W. Conners, and Y. Liu, "Wide-area frequency as a criterion for digital audio recording authentication," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, Jul. 2011, pp. 1–7.

[105] R. Garg, A. L. Varna, and M. Wu, "Modeling and analysis of electric network frequency signal for timestamp verification," in *Proc. IEEE Int. Workshop Inf. Forensics Security (WIFS)*, Dec. 2012, pp. 67–72.

[106] L. Zheng, Y. Zhang, C. E. Lee, and V. L. L. Thing, "Time-of-recording estimation for audio recordings," *Digit. Invest.*, vol. 22, pp. S115–S126, Aug. 2017.

[107] G. Hua, "Error analysis of forensic ENF matching," in *Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS)*, Dec. 2018, pp. 1–7.

[108] S. Vatansever, A. E. Dirik, and N. Memon, "ENF based robust media time-stamping," *IEEE Signal Process. Lett.*, vol. 29, pp. 1963–1967, 2022.

[109] G. Hua, Q. Wang, D. Ye, and H. Zhang, "Reliability of power system frequency on times-stamping digital recordings," 2020, *arXiv:2011.00176*.

[110] G. Pop, D. Draghicescu, D. Burileanu, H. Cucu, and C. Burileanu, "Fast method for ENF database build and search," in *Proc. Int. Conf. Speech Technol. Hum.-Comput. Dialogue (SpeD)*, Jul. 2017, pp. 1–6.

[111] A. Hajj-Ahmad, R. Garg, and M. Wu, "ENF based location classification of sensor recordings," in *Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS)*, Nov. 2013, pp. 138–143.

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

[113] P. B. Suresha, S. Nagesh, P. S. Roshan, P. A. Gaonkar, G. N. Meenakshi, and P. K. Ghosh, "A high resolution ENF based multi-stage classifier for location forensics of media recordings," in *Proc. 23rd Nat. Conf. Commun. (NCC)*, Mar. 2017, pp. 1–6.

[114] Ž. Šaric, A. Žunic, T. Zrnic, M. Kneževic, D. Despotovic, and T. Delic, "Improving location of recording classification using electric network frequency (ENF) analysis," in *Proc. IEEE 14th Int. Symp. Intell. Syst. Informat. (SISY)*, Aug. 2016, pp. 51–56.

[115] H. Zhou, H. Duanmu, J. Li, Y. Ma, J. Shi, Z. Tan, and W. Li, "Geographic location estimation from ENF signals with high accuracy," in *Proc. IEEE Signal Processing Cup*, Mar. 2016, pp. 1–8.

[116] R. Ohib, S. Y. Arnob, R. Arefin, M. Amin, and T. Reza, "ENF based grid classification system: Identifying the region of origin of digital recordings," *Criterion*, vol. 3, no. 4, p. 5, 2017.

[117] M. Sarkar, D. Chowdhury, C. Shahnaz, and S. A. Fattah, "Application of electrical network frequency of digital recordings for location-stamp verification," *Appl. Sci.*, vol. 9, no. 15, p. 3135, Aug. 2019.

[118] M. E. Helou, A. W. Turkmani, R. Chanouha, and S. Charbaji, "A novel ENF extraction approach for region-of-recording identification of media recordings," *Forensic Sci. Int.*, vol. 155, nos. 2–3, p. 165, 2005.

[119] D. Despotović, M. Kneževic, Ž. Šaric, T. Zrnic, A. Žunic, and T. Delic, "Exploring power signatures for location forensics of media recordings," in *Proc. IEEE Signal Process. Cup*, Mar. 2016, pp. 20–25.

[120] A. Gaber, A. Zayed, B. Ahmed, E. Elshiekh, H. Aly, I. Sherif, K. Elgammal, O. Ahmadein, O. Elzaafarany, and T. Gama, "Electric network frequency (ENF) recognition," IEEE Signal Process. Cup, Pursuit Process. Team, Final Rep., Mar. 2016.

[121] IEEE SigPort. (2016). *Information on the mAST ENF Power Signature Dataset*. [Online]. Available: http://sigport.org/1108

[122] W. Bang and J. W. Yoon, "Power grid estimation using electric network frequency signals," *Secur. Commun. Netw.*, vol. 2019, pp. 1–11, Sep. 2019.

[123] 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 *Proc. World Wide Web Conf. World Wide Web*, 2018, pp. 339–348.

[124] EarthCam Inc. (2017). *EarthCam—Webcam Network*. [Online]. Available: https://www.earthcam.com/

[125] S. R. L. VisioRay. (2017). *SkylineWebcams | Live Cams Around the World!*. [Online]. Available: https://www.skylinewebcams.com/

[126] The Annenberg Foundation. (2017). *The Largest Live Nature Cam Network on the Planet World!* [Online]. Available: http://explore.org/

[127] R. Garg, A. Hajj-Ahmad, and M. Wu, "Geo-location estimation from electrical network frequency signals," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 2862–2866.

[128] W. Yao, J. Zhao, M. J. Till, S. You, Y. Liu, Y. Cui, and Y. Liu, "Source location identification of distribution-level electric network frequency signals at multiple geographic scales," *IEEE Access*, vol. 5, pp. 11166–11175, 2017.

[129] J. Chai, J. Zhao, W. Yao, J. Guo, and Y. Liu, "Application of wide area power system measurement for digital authentication," in *Proc. IEEE/PES Transmiss. Distrib. Conf. Expo.*, May 2016, pp. 1–5.

[130] Y. Cui, Y. Liu, P. Fuhr, and M. Morales-Rodriguez, "Exploiting spatial signatures of power ENF signal for measurement source authentication," in *Proc. IEEE Int. Symp. Technol. Homeland Secur. (HST)*, Oct. 2018, pp. 1–6.

[131] R. Garg, A. Hajj-Ahmad, and M. Wu, "Feasibility study on intra-grid location estimation using power ENF signals," 2021, *arXiv:2105.00668*.

[132] G. Hua, Y. Zhang, J. Goh, and V. L. L. Thing, "Audio authentication by exploring the absolute-error-map of ENF signals," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 5, pp. 1003–1016, May 2016.

[133] M. Savari, A. W. A. Wahab, and N. B. Anuar, "High-performance combination method of electric network frequency and phase for audio forgery detection in battery-powered devices," *Forensic Sci. Int.*, vol. 266, pp. 427–439, Sep. 2016.

[134] D. P. Nicolalde Rodriguez, J. A. Apolinario, and L. W. P. Biscainho, "Audio authenticity: Detecting ENF discontinuity with high precision phase analysis," *IEEE Trans. Inf. Forensics Security*, vol. 5, no. 3, pp. 534–543, Sep. 2010.

[135] M. Desainte-Catherine and S. Marchand, "High-precision Fourier analysis of sounds using signal derivatives," *J. Audio Eng. Soc.*, vol. 48, nos. 7–8, pp. 654–667, 2000.

[136] D. P. Nicolalde-Rodríguez, J. A. Apolinário, and L. W. P. Biscainho, "Audio authenticity based on the discontinuity of ENF higher harmonics," in *Proc. 21st Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2013, pp. 1–5.

[137] W.-H. Chuang, R. Garg, and M. Wu, "How secure are power network signature based time stamps?" in *Proc. ACM Conf. Comput. Commun. Secur.*, Oct. 2012, pp. 428–438.

[138] W.-H. Chuang, R. Garg, and M. Wu, "Anti-forensics and countermeasures of electrical network frequency analysis," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 12, pp. 2073–2088, Dec. 2013.

[139] P. A. A. Esquef, J. A. Apolinário, and L. W. P. Biscainho, "Edit detection in speech recordings via instantaneous electric network frequency variations," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 12, pp. 2314–2326, Dec. 2014.

[140] P. A. A. Esquef, J. A. Apolinário, and L. W. P. Biscainho, "Improved edit detection in speech via ENF patterns," *IEEE Int. Workshop Inf. Forensics Secur.*, Nov. 2015, pp. 1–6.

[141] M. Fuentes, P. Zinemanas, P. Cancela, and J. A. Apolinário, "Detection of ENF discontinuities using PLL for audio authenticity," in *Proc. IEEE 7th Latin Amer. Symp. Circuits Syst. (LASCAS)*, Feb. 2016, pp. 79–82.

[142] P. M. G. I. Reis, J. P. C. L. da Costa, R. K. Miranda, and G. D. Galdo, "Audio authentication using the kurtosis of ESPRIT based ENF estimates," in *Proc. 10th Int. Conf. Signal Process. Commun. Syst. (ICSPCS)*, Dec. 2016, pp. 1–6.

[143] P. M. G. I. Reis, J. P. C. L. da Costa, R. K. Miranda, and G. D. Galdo, "ESPRIT-Hilbert-based audio tampering detection with SVM classifier for forensic analysis via electrical network frequency," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 4, pp. 853–864, Apr. 2017.

[144] X. Lin and X. Kang, "Supervised audio tampering detection using an autoregressive model," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 2142–2146.

[145] Y. Hu, C. T. Li, Z. Lv, and B. B. Liu, "Audio forgery detection based on max offsets for cross correlation between ENF and reference signal," in *Proc. Int. Workshop Digital Watermarking*. Berlin, Germany: Springer, 2013, pp. 253–266.

[146] Z. Lv, Y. Hu, C.-T. Li, and B.-B. Liu, "Audio forensic authentication based on MOCC between ENF and reference signals," in *Proc. IEEE China Summit Int. Conf. Signal Inf. Process.*, Jul. 2013, pp. 427–431.

[147] Y. Wang, Y. Hu, A. W.-C. Liew, and C.-T. Li, "ENF based video forgery detection algorithm," *Int. J. Digit. Crime Forensics*, vol. 12, no. 1, pp. 131–156, Jan. 2020.

[148] C. C. Huang, Y. Zhang, and V. L. L. Thing, "Inter-frame video forgery detection based on multi-level subtraction approach for realistic video forensic applications," in *Proc. IEEE 2nd Int. Conf. Signal Image Process. (ICSIP)*, Aug. 2017, pp. 20–24.

[149] K. Sitara and B. M. Mehtre, "A comprehensive approach for exposing inter-frame video forgeries," in *Proc. IEEE 13th Int. Colloq. Signal Process. Appl. (CSPA)*, Mar. 2017, pp. 73–78.

[150] Z.-F. Wang, J. Wang, C.-Y. Zeng, Q.-S. Min, Y. Tian, and M.-Z. Zuo, "Digital audio tampering detection based on ENF consistency," in *Proc. Int. Conf. Wavelet Anal. Pattern Recognit. (ICWAPR)*, Jul. 2018, pp. 209–214.

[151] H. Su, R. Garg, A. Hajj-Ahmad, and M. Wu, "ENF analysis on recaptured audio recordings," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 3018–3022.

[152] X. Lin, J. Liu, and X. Kang, "Audio recapture detection with convolutional neural networks," *IEEE Trans. Multimedia*, vol. 18, no. 8, pp. 1480–1487, Aug. 2016.

[153] D. Bykhovsky, "Recording device identification by ENF harmonics power analysis," *Forensic Sci. Int.*, vol. 307, Feb. 2020, Art. no. 110100.

[154] S. Baudry, B. Chupeau, M. De Vito, and G. Doërr, "Modeling the flicker effect in camcorded videos to improve watermark robustness," in *Proc. IEEE Int. Workshop Inf. Forensics Secur.*, Dec. 2014, pp. 42–47.

[155] Z. Huang, K. Nahrstedt, and R. Steinmetz, "Evolution of temporal multimedia synchronization principles," in *MediaSync*. Cham, Switzerland: Springer, 2018, pp. 33–71.

[156] O. Golokolenko and G. Schuller, "Investigation of electric network frequency for synchronization of low cost and wireless sound cards," in *Proc. 25th Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2017, pp. 693–697.

[157] B. Liu, Y. Chen, D. Shen, G. Chen, K. Pham, E. Blasch, and B. Rubin, "An adaptive process-based cloud infrastructure for space situational awareness applications," *Proc. SPIE*, vol. 9085, pp. 149–157, Jun. 2014.

[158] S. Y. Nikouei, R. Xu, D. Nagothu, Y. Chen, A. Aved, and E. Blasch, "Real-time index authentication for event-oriented surveillance video query using blockchain," in *Proc. IEEE Int. Smart Cities Conf. (ISC2)*, Sep. 2018, pp. 1–8.
[159] D. Nagothu, Y. Chen, A. Aved, and E. Blasch, "Authenticating video feeds using electric network frequency estimation at the edge," *ICST Trans. Secur. Saf.*, Jul. 2018, Art. no. 168648.
[160] D. Nagothu, R. Xu, Y. Chen, E. Blasch, and A. Aved, "DeFake: Decentralized ENF-consensus based DeepFake detection in video conferencing," in *Proc. IEEE 23rd Int. Workshop Multimedia Signal Process. (MMSP)*, Oct. 2021, pp. 1–6.
[161] D. Nagothu, R. Xu, Y. Chen, E. Blasch, and A. Aved, "Deterring deepfake attacks with an electrical network frequency fingerprints approach," *Future Internet*, vol. 14, no. 5, p. 125, Apr. 2022.
[162] D. Nagothu, R. Xu, Y. Chen, E. Blasch, and A. Aved, "DeFakePro: Decentralized deepfake attacks detection using ENF authentication," *IT Prof.*, vol. 24, no. 5, pp. 46–52, Sep. 2022.
[163] J. Chai, L. Yuming, Z. Yuan, R. W. Conners, and Y. Liu, "Tampering detection of digital recordings using electric network frequency and phase angle," *Audio Eng. Soc. Conv.*, vol. 135, pp. 1–8, Oct. 2013.
[164] E. Jacobsen and P. Kootsookos, "Fast, accurate frequency estimators," *IEEE Signal Process. Mag.*, vol. 24, no. 3, pp. 123–125, May 2007.
[165] R. Korycki, "Time and spectral analysis methods with machine learning for the authentication of digital audio recordings," *Forensic Sci. Int.*, vol. 230, nos. 1–3, pp. 117–126, Jul. 2013.
[166] M. Wu, F. Pereira, J. Liu, H. Ramos, M. Alvim, and L. Oliveira, "Proof carrying sensing—Towards real-world authentication in cyber-physical systems," in *Proc. ACM Conf. Embedded Networked Sensor Syst.*, Nov. 2017, pp. 1–20.

ERICMOORE NGHARAMIKE received the bachelor's degree in computer science from the Federal University of Technology, Owerri, Nigeria, and the master's degree in network computing from Coventry University, U.K. He is currently pursuing the Ph.D. degree with the University of the Sunshine Coast, Australia. Before commencing the Ph.D. degree, he was a Lecturer and a Researcher with the Department of Computer Science, Federal University Oye-Ekiti (FUOYE), Nigeria. His research interests include the Internet of Things (IoT), wireless multimedia sensor systems, data analytics, machine learning, and multimedia signal processing.

LI-MINN ANG (Senior Member, IEEE) received the B.Eng. (Hons.) and Ph.D. degrees from Edith Cowan University (ECU), Australia. He was an Associate Professor of networked and computer systems with the School of Information and Communication Technology (ICT), Griffith University. He has worked at Australian and U.K. universities, including Monash University, the University of Nottingham, ECU, California State University (CSU), and Griffith University. He is currently a Professor of electrical and computer engineering with the School of Science, Technology, and Engineering, University of the Sunshine Coast (USC), Australia. His research interests include computer, electrical, and systems engineering, including the Internet of Things, intelligent systems and data analytics, machine learning, visual information processing, embedded systems, wireless multimedia sensor systems, reconfigurable computing (FPGA), and the development of innovative technologies for real-world systems, including smart cities, engineering, agriculture, environment, health, and defense. He is a fellow of the Higher Education Academy in the U.K.

KAH PHOOI SENG (Senior Member, IEEE) received the B.Eng. and Ph.D. degrees from the University of Tasmania, Australia. She was a Professor and the Head of the Department of Computer Science and Networked System, Sunway University. Before joining Sunway University, she was an Associate Professor with the School of Electrical and Electronic Engineering, Nottingham University. She has worked or attached to Australian-based and U.K.-based universities, including Monash University, Griffith University, the University of Tasmania, the University of Nottingham, Sunway University, Edith Cowan University, and Charles Sturt University. Prior to joining the Queensland University of Technology (QUT), she was an Adjunct Professor with the School of Engineering and Information Technology, UNSW. She is currently a Professor of artificial intelligence with Xi'an Jiaotong-Liverpool University and an Adjunct Professor with the School of Computer Science, QUT. She has participated in more than U.S. $1.8 million research grant projects from the government and industry in Australia and overseas. She has supervised or co-supervised 15 Ph.D. students to completion and more than 25 higher-degree research students. She has a strong record of publications and has published more than 250 papers in journals and internationally refereed conferences. Her research interests include computer science and engineering, including artificial intelligence (AI), data science and machine learning, big data, multimodal information processing, intelligent systems, the Internet of Things (IoT), embedded systems, mobile software development, affective computing, computer vision, and the development of innovative technologies for real-world applications. She also serves on the editorial board or committees for several journals and international conferences. She is an Associate Editor of IEEE Access.

MINGZHONG WANG joined the University of the Sunshine Coast, in 2015. Before that, he was a Lecturer with the Beijing Institute of Technology, China. He developed and taught a wide range of subjects, including data analysis, artificial intelligence, algorithm design, and computational theory. He has won a grant from the Chinese National Science Foundation for studying service-oriented scientific workflow scheduling in a cloud environment. He has been the principal investigator of several industrial-funded projects, with a focus on enterprise information and business management systems. His current research interests include machine learning, knowledge graphs, and deep reinforcement learning. He is also interested in mobile computing and telepresence robots. He received the Best Paper Award from the 2013 (32nd) International Performance Computing and Communications Conference.