IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 25 July 2024, accepted 19 August 2024, date of publication 23 August 2024, date of current version 2 September 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3449099

RESEARCH ARTICLE

An Enhanced STFT Segmentation Framework for ENF-Based Media Forensics

ALI BERK YALINKILIÇ AND SAFFET VATANSEVER, (Member, IEEE)
Department of Mechatronics Engineering, Bursa Technical University, 16310 Bursa, Türkiye
Corresponding author: Saffet Vatansever (saffet.vatansever@btu.edu.tr)

This work was supported in part by Bursa Technical University Scientific Research Units under Project 211N022.

ABSTRACT The electric network frequency (ENF) criterion has gained significant attention over the past two decades as a promising tool in digital media forensics. ENF is the frequency of the alternating current (AC) signal in a mains electricity network, exhibiting continual fluctuations within certain limits around a nominal frequency, contingent upon supplied and demanded power disparities. A sequence of ENF alterations is called an ENF signal, which is inherently embedded in audio and video recordings under certain circumstances. Several efforts have been made to accurately estimate the ENF signal from media. However, no matter how accurately estimated, a media ENF signal may not be reliably used in forensic applications unless sufficiently distinctive. To clarify, ENF may show similar fluctuation patterns at different time intervals. These patterns become more distinct over longer periods of time. Accordingly, working with as large an ENF signal as possible is critical for reliability. To achieve an extended and, thus, more distinctive ENF signal, this study proposes a smart segmentation scheme for Short-Time Fourier Transform (STFT)-based ENF estimation, which derives more data segments from a given media than the conventional STFT technique, leading to increased ENF estimates for any specified STFT parameter setting. The proposed approach can be combined with any ENF accuracy enhancement strategy to obtain relatively more reliable signals. Large-scale tests conducted with different STFT parameters and audio clip lengths showed that the proposed scheme can efficiently improve the performance when used alone or in conjunction with other ENF enhancement strategies.

INDEX TERMS ENF, electric network frequency, media forensics, short-time Fourier transform, STFT segmentation, time-of-recording, timestamp.

I. INTRODUCTION

The electric network frequency (ENF) criterion [1], [2], [3], [4] has proven to be an effective tool in digital media forensics for the last two decades. The ENF is the frequency of AC electricity in a mains power grid. It varies constantly within certain bounds around a nominal frequency (50 or 60 Hz) depending on the supply and demand imbalance in power [5]. In most parts of the world, the nominal value of the ENF is 50 Hz; however, in some regions of North America and Asia, it is 60 Hz. In an interconnected network, the ENF exhibits consistent fluctuations throughout the network [5]. Consequently, the ground truth ENF variations, for any period, can be acquired from any power outlet across the grid. A sequence of successive ENF variations over time is referred to as an ENF signal.

The ENF signal has been revealed to intrinsically integrate into audio recordings captured in environments with mains-sourced electromagnetic fields by a dynamic microphone [1], [2], [3], [4], [6], [7]. Further research has shown that the ENF is also incorporated into audio recorded in settings with acoustic mains hum through an electret microphone [8], [9], [10], [11]. Later research has discovered that the ENF is also inherently embedded in video recordings in settings with illumination from a mains-powered light source [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23]. Indeed, it has been revealed that the ENF trace can be identified even from a single image [24], [25].

Various time and frequency-domain approaches have been adopted in the literature to estimate the ENF from media,

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

including Zero-Crossing (ZC) [2], Short-Time Fourier Transform (STFT) [6], Multiple Signal Classification (MUSIC) [26], and Estimation of Parameters using Rotational Invariant Techniques (ESPIRIT) [26]. The estimated ENF signal can be employed for various media forensic applications for different scenarios or case studies. The time-of-recording detection or verification [7], [13], [27], [28], which typically requires a similarity comparison between the media ENF signal and the ground truth ENF signal, is among the most widely studied. It can potentially be used in legal proceedings, i.e., in court, to support or refute claims. For example, an alleged criminal can be acquitted of a crime if there is a video of them taken somewhere other than the crime scene during the crime and if it can be verified using the ENF criterion that the video recording and the crime coincide [29]. The use of the ENF signal in digital media for forensic purposes is not limited to the time-of-recording detection or verification. By serving as a power signature, it also enables other practical applications, including geo-location estimation [24], [30], [31] (e.g., to identify the country of origin of a recording), multimedia synchronization [32], [33] (e.g., to temporally align videos taken by two cameras to merge their views into a single panoramic view), media authentication [34], [35], [36] (e.g., to determine if a video is original or tampered with), and camera characterization [37], [38] (e.g., to attribute the source camcorder of a video).

Numerous factors affect the reliability of ENF-based media forensic applications, including the length of the query media, length of the ground truth, and signal-to-noise ratio (SNR) [18], [39], [40]. Several studies have been conducted to obtain accurate ENF estimates under low SNR conditions. Maximum-likelihood estimation (MLE) with spectrum combining [41], [42], a robust filtering algorithm (RFA) for single harmonics [43], adaptive multi-trace carving (AMTC) for robust frequency tracking [44], a multi-tone harmonic robust filtering algorithm (HRFA) for harmonic enhancement [45], a graph-based harmonic selection algorithm (GHSA) [45], and a least absolute deviation (LAD)-based framework [46] are some of the effective strategies introduced to enhance ENF estimation accuracy.

A critical challenge for all ENF enhancement strategies is non-unique ENF patterns, particularly in short recordings. To put it more clearly, ENF tends to show similar patterns, that is, similar fluctuations, over short time intervals because of the comparable discrepancies in the supply and demand of power from time to time. The ENF patterns over longer periods are more distinct owing to the unlikeliness of occurrence and continuation of such discrepancies over a long span. Consequently, the ENF signal estimated from media of longer duration is expected to be more distinctive and, thus, is more reliable than a smaller one because it is less likely to produce false positives in any ENF-related media forensic applications. However, query media may not always be sufficiently long for such reliability. In this circumstance, acquiring as large an ENF signal as possible from any given media is critical.

This work proposes an ingenious segmentation scheme for Short-Time Fourier Transform (STFT)-based ENF estimation, which constitutes additional data segments to achieve increased ENF estimates from a given media compared to the conventional STFT technique. This leads to the acquisition of an extended and, thus, more distinctive ENF signal for any specified STFT parameter setting by enabling additional ENF sample computations from both the beginning and end of the media, which are not considered by the traditional method. To the best of our knowledge, no previous work has attempted an enhancement strategy in the segmentation stage of the STFT to obtain a more effective and reliable ENF signal for ENF-based forensic applications. Because the existing ENF signal enhancement strategies focus on the other stages of the STFT, aiming to improve the estimation accuracy, the proposed scheme is suitable for use in combination with these techniques. To be more specific, the proposed strategy can be integrated into any STFT-based ENF signal enhancement technique. Experimental results reveal that the proposed scheme is considerably effective in improving performance when applied to the conventional STFT technique or used in conjunction with other ENF enhancement strategies.

The rest of this paper is organized as follows. Section II highlights some background information for the traditional STFT-based ENF signal estimation procedure, forming the basis of this work. Section III introduces the proposed STFT segmentation scheme to achieve longer and more distinctive ENF signals. Section IV evaluates the performance of the proposed scheme through large-scale tests on the ENF-WHU dataset [45], [47]. Section V extends the applicability of the proposed technique and explores how it can further increase the effectiveness of existing ENF enhancement strategies when used in conjunction with. Finally, Section VI concludes the paper and provides a summary of the main research findings.

II. CONVENTIONAL STFT-BASED ENF SIGNAL ESTIMATION: HIGHLIGHTS

The Short-Time Fourier Transform (STFT), which is a powerful tool for computing time-dependent changes in the frequency and phase components of a signal, is one of the most widely used methods in ENF-based media forensics to capture ENF alterations in the mains electricity as well as extract these fluctuations from audio or video. The STFT-based ENF estimation technique comprises of two consecutive stages. In the first stage, the given signal (audio or luminance signal (for video)) is divided into partially overlapping segments of equal length [6], as shown in Fig. 1 (Before the segmentation operation, the given signal is expected first to be decimated and bandpass-filtered around the ENF frequency of interest.). Each of the resulting segments is exploited to obtain one ENF estimate in the second stage by detecting the frequency of the highest magnitude around the

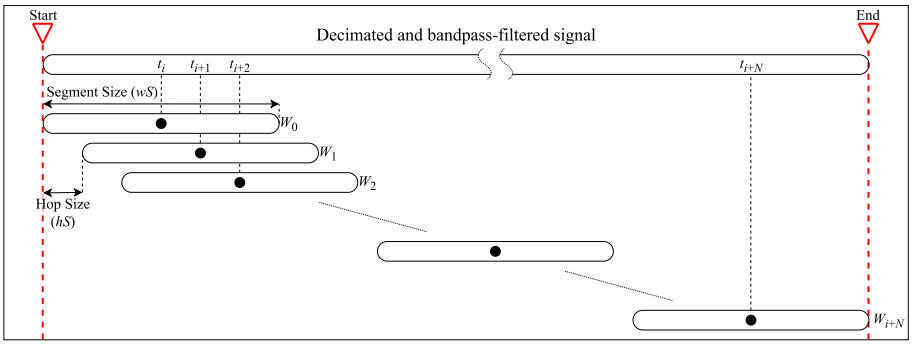

FIGURE 1. Conventional STFT segmentation scheme, in which a fixed segment size is used [6].

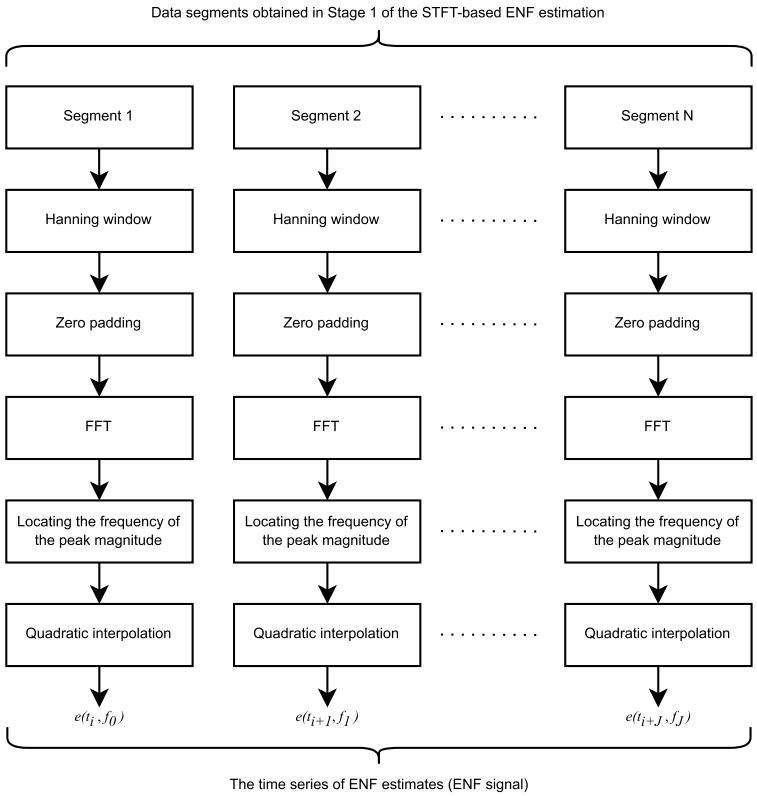

FIGURE 2. ENF signal estimation using the segmented data portions [6].

ENF frequency of interest, as illustrated in Fig. 2 (for details on the intermediate steps, refer to [6]).

The time corresponding to the midway point of an STFT segment defines the time instant of the ENF sample to be computed from this segment. The choice of the STFT segment size is a trade-off between noise and distinction. Although a larger segment size reduces the potential for noisy ENF estimates, it may hinder some ENF fluctuations. The hop size (hS), which is fixed for each consecutive segment, determines the ENF resolution (in samples/second), and picking a large one may also obstruct some ENF alterations to catch.

The number of ENF samples that can be estimated from a given media depends on the hop size and segment size selected. A larger segment size results in fewer segments to construct along the signal for a particular hop size, leading to fewer ENF estimations. Consequently, more ENF samples that could potentially be estimated from the media are lost. These losses are the ENF time series before the first segment and after the last segment. More specifically, they are the ENF variations during the periods before and after the midpoints of the first and last segments (Recall that an STFT segment's halfway point designates the time instant of the ENF sample to be estimated from it.). To remedy these losses,

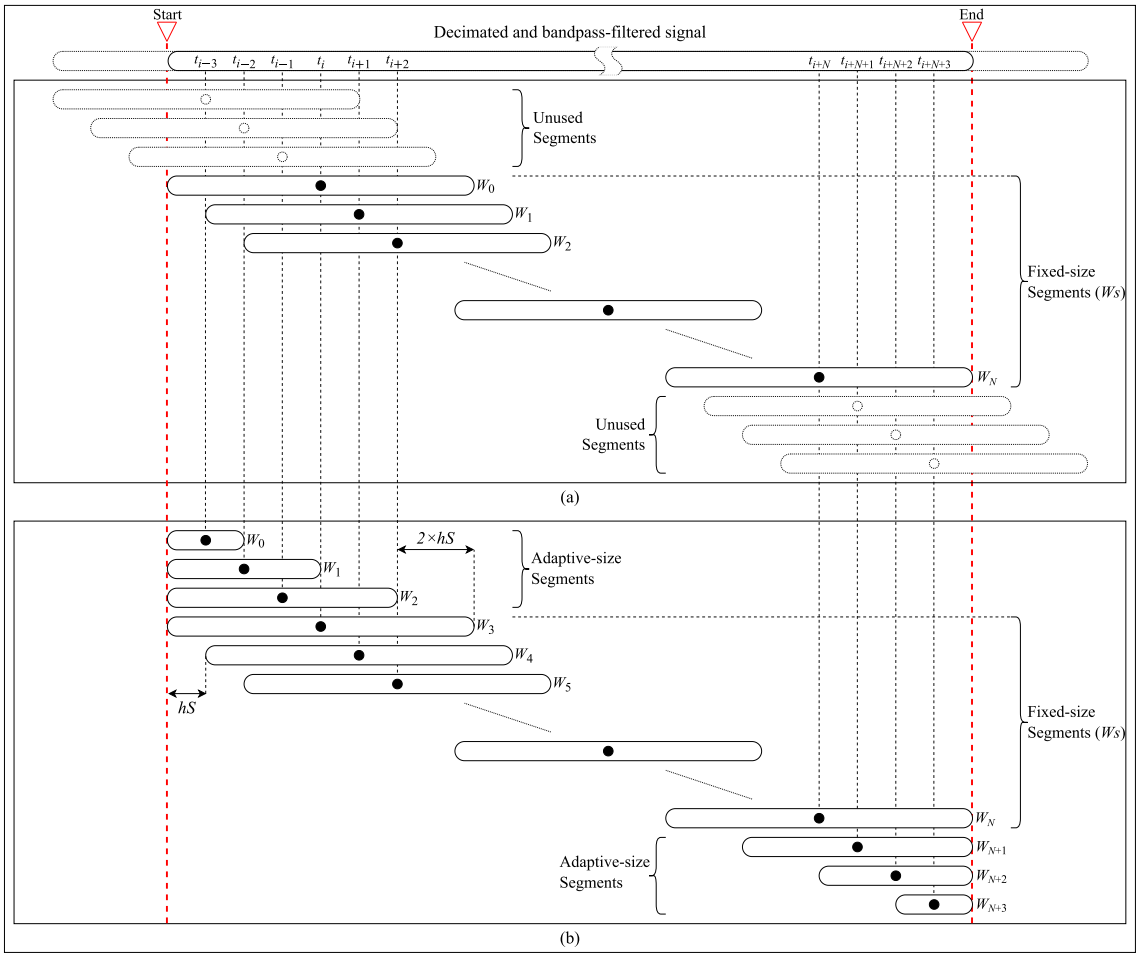

FIGURE 3. An illustration of how the proposed STFT segmentation scheme derives from the traditional technique: (a) traditional technique with non-existing data segments before t_i and after t_{i+N}. (b) proposed approach with additional anterior and posterior data segments (before t_i and after t_{i+N}).

Section III presents a technique that enables ENF estimations for these periods via an adaptive segmentation scheme.

III. PROPOSED TECHNIQUE

This section proposes a smart segmentation scheme for STFT-based ENF signal estimation to obtain a larger, hence more distinct ENF signal from a given media in comparison with the conventional technique. Fig. 3 (b) shows a block diagram of the proposed method, which is sketched under the block diagram of the conventional scheme (Fig. 3 (a)) to elucidate how the proposed technique develops and differs from the traditional one. As is evident from the figure, the proposed method does not modify or remove any segment obtained by the traditional scheme (i.e., it preserves and uses all segments obtained by the traditional strategy) but constructs additional segments at the beginning and end. Therefore, the proposed approach can be considered complementary to the conventional scheme. Considering that the time instant of the first ENF sample (t_i in Fig. 1 (a)) corresponds to the midpoint of the first segment (refer to Section II), there is no way to estimate any sample associated with the previous time points using the conventional scheme unless the media record starts earlier. The same goes for estimating the ENF samples for the next time points unless the media record ends later. The proposed technique introduces an adaptive segmentation scheme for these periods to remedy this issue and acquire the ENF samples that were missed before and after the first and last segments of the standard technique. More specifically, to acquire the posterior ENF samples, that is, missing samples at the end, it suggests shrinking the segment size for each subsequent sample one after the other by $2 \times hS$, starting from the left end of the last segment, yet leaving the right end as is (i.e., spanning through the end of the record), as shown in Fig. 3 (b). Consequently, each subsequent segment becomes smaller than the previous segment by $2 \times hS$. A similar procedure is used to compute the anterior missing ENF samples. However, this time, the reduction starts from the right end of the first segment and moves leftward, leading to each prior segment becoming smaller than the next by $2 \times hS$, and each segment spans through the beginning of the record. At first glance, one may consider that the proposed technique leads to a different

Algorithm 1 A Redesigned STFT-Based ENF Signal Estimation Procedure for Query Media Using the Proposed Segmentation Scheme

$\mathbf{D} \leftarrow$ *given media*
$\mathbf{C} \leftarrow$ *query clip of* $\mathbf{D}$
$sM \leftarrow$ *initial moment of* $\mathbf{C}$ *(in sec)*
$cL \leftarrow$ *length of* $\mathbf{C}$ *(in sec)*
$wS \leftarrow$ *length of fixed-size segments (in sec)*
$sDw \leftarrow$ *smallest size of adaptive segments (in sec)*
$eDw \leftarrow$ *largest size of adaptive segments (in sec)*
$hS \leftarrow$ *hop size (in sec)*
$fs \leftarrow$ *sampling frequency*
$\mathbf{C} \leftarrow \mathbf{D}[sM \times fs : (sM + cL) \times fs - 1]$
$eDw \leftarrow wS - 2 \times hS$
$n \leftarrow 0$
for $k \leftarrow sDw : 2 \times hS : eDw$ **do**
    $\mathbf{w} \leftarrow \mathbf{C}[0 : (k \times fs) - 1]$
    $\mathbf{E}[n] \leftarrow$ ENFestimation($\mathbf{w}$)
    $n \leftarrow n + 1$
end for
for $k \leftarrow 0 : hS \times fs : (cL - wS) \times fs$ **do**
    $\mathbf{w} \leftarrow \mathbf{C}[k : k + (wS \times fs) - 1]$
    $\mathbf{E}[n] \leftarrow$ ENFestimation($\mathbf{w}$)
    $n \leftarrow n + 1$
end for
for $k \leftarrow eDw : -2 \times hS : sDw$ **do**
    $\mathbf{w} \leftarrow \mathbf{C}[end - (k \times fs) : end]$
    $\mathbf{E}[n] \leftarrow$ ENFestimation($\mathbf{w}$)
    $n \leftarrow n + 1$
end for

ENF resolution along the interval with the new segments compared to the phase of fixed-size segments. However, a shrink in a segment by $2 \times hS$ from one end and maintaining the other end moves the midpoint by as much as one hS. Therefore, the ENF resolution is sustained throughout the media. The new segments can be considered a shrunken form of the non-existing segments of the standard STFT from both ends, imagining that the recording started earlier and finished later, as shown by the dotted thin line in Fig. 1 (a). As is evident from Fig. 1 (a) and (b), the midpoint of the non-existing segments of the standard scheme matches those of the proposed segments.

To obtain the ENF estimate from each new segment, the proposed method uses the second stage of the conventional STFT-based ENF estimation approach (Fig. 2) in the same way as the standard technique. The proposed method primarily involves enhancing the first stage of the conventional STFT-based ENF estimation approach. Therefore, it should be emphasized that the objective of the proposed method is not to improve the ENF signal quality to be estimated from the media but to enlarge and make the estimated signal more distinct and unique by computing additional ENF samples that are not considered by the traditional method. It may

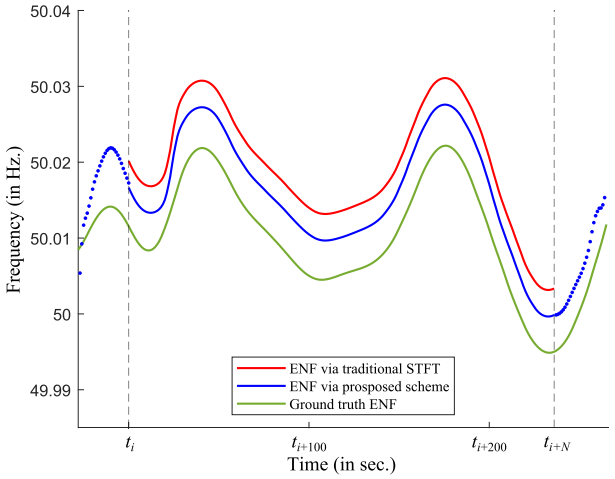

FIGURE 4. A comparison of the estimated ENF signals using the proposed approach and the conventional scheme for a 5-minute audio clip. A hop size (hS) of 1 second, a fixed segment size of 64 seconds, and a minimum adaptive segment size of 16 seconds were used. t_i is 32, and t_{i+N} is 268, which are, respectively, the midpoints of the first and last segments obtained by the conventional method.

be particularly significant for short-duration recordings for distinctiveness.

A pseudocode demonstrating how to estimate the ENF signal from a query clip of a given signal (audio or luminance signal (for video)) using the proposed segmentation method is provided in Algorithm 1. It is particularly important to understand the adaptation of the proposed scheme to the reference signal (to be discussed next in Algorithm 2). Algorithm 1 comprises three consecutive loops for the three phases of the proposed method (Fig. 3 (b)): anterior adaptive segmentation, standard segmentation, and posterior adaptive segmentation. In the first loop, the ENF is estimated for each of the suggested anterior segments of varying lengths (Fig. 3 (b)), where each successive segment is larger than the previous segment by $2 \times hS$, and each starts from the beginning of the query signal. Here, the smallest segment size (sDw) is preset based on user preference, considering that it should be less than the length of the fixed-size segments (wS) by a multiple of $2 \times hS$. Accordingly, this phase's largest segment size (eDw) is $2 \times hS$ smaller than the length of the fixed-size segments (wS). In the second loop, the ENF is estimated for each fixed-size segment acquired using the standard STFT segmentation scheme (Fig. 1), where the hop size is one hS. In the final loop, the ENF is estimated for each of the suggested posterior segments, where each successive segment is smaller than the previous segment by $2 \times hS$, and each spans through the end of the query signal. Here, the smallest and largest segment sizes are set to be the same as those in the first loop. Algorithm 1 follows the same ENF estimation procedure as the standard technique (Fig. 2) for every segment in each loop. Regarding the time complexity, the proposed technique increases the operational time by approximately *the number of new segments* $\times$ *a segment's processing time*.

Algorithm 2 A Redesigned STFT-Based Ground Truth ENF Signal Estimation Procedure for Reference Data (mains signal) Using the Proposed Segmentation Scheme

1: $\mathbf{R} \leftarrow$ *reference data*
2: $cL \leftarrow$ *length of query clip of given media* (*in sec*)
3: $\mathbf{r} \leftarrow$ *a clip of* $\mathbf{R}$ *in the length of* cL
4: $wS \leftarrow$ *length of fixed-size segments* (*in sec*)
5: $sDw \leftarrow$ *smallest size of adaptive segments* (*in sec*)
6: $eDw \leftarrow$ *largest size of adaptive segments* (*in sec*)
7: $hS \leftarrow$ *hop size* (*in sec*)
8: $fs \leftarrow$ *sampling frequency*
9: $eDw \leftarrow wS - 2 \times hS$
10: **for** $i \leftarrow 0 : hS \times fs : \text{Length}(\mathbf{R}) - (cL \times fs)$ **do**
11: $n \leftarrow 0$
12: $\mathbf{r} \leftarrow \mathbf{R}[i : i + (cL \times fs) - 1]$
13: **if** $i == 0$ **then**
14: **for** $j \leftarrow sDw : 2 \times hS : eDw$ **do**
15: $\mathbf{w} \leftarrow \mathbf{r}(0 : (j \times fs) - 1)$
16: $\mathbf{G}[i,\ n] \leftarrow \text{ENFestimation}(\mathbf{w})$
17: $n \leftarrow n + 1$
18: **end for**
19: **for** $j \leftarrow 0 : hS \times fs : (cL - wS) \times fs$ **do**
20: $\mathbf{w} \leftarrow \mathbf{r}[j : j + (wS \times fs) - 1]$
21: $\mathbf{G}[i,\ n] \leftarrow \text{ENFestimation}(\mathbf{w})$
22: $n \leftarrow n + 1$
23: **end for**
24: **for** $j \leftarrow eDw : -2 \times hS : sDw$ **do**
25: $\mathbf{w} \leftarrow \mathbf{r}[end - (j \times fs) : end]$
26: $\mathbf{G}[i,\ n] \leftarrow \text{ENFestimation}(\mathbf{w})$
27: $n \leftarrow n + 1$
28: **end for**
29: **else**
30: **for** $j \leftarrow sDw : 2 \times hS : eDw$ **do**
31: $\mathbf{w} \leftarrow \mathbf{r}(0 : (j \times fs) - 1)$
32: $\mathbf{G}[i,\ n] \leftarrow \text{ENFestimation}(\mathbf{w})$
33: $n \leftarrow n + 1$
34: **end for**
35: $l \leftarrow (cL - wS) \div hS + 1$ ▷ length of ENF estimates to be copied to the next ENF clip.
36: $\mathbf{G}[i,\ n : n + l - 1] \leftarrow \mathbf{G}[i - 1,\ n + 1 : n + l]$
37: $n \leftarrow n + l$
38: **for** $j \leftarrow wS : -2 \times hS : sDw$ **do**
39: $\mathbf{w} \leftarrow \mathbf{r}[end - (j \times fs) : end]$
40: $\mathbf{G}[i,\ n] \leftarrow \text{ENFestimation}(\mathbf{w})$
41: $n \leftarrow n + 1$
42: **end for**
43: **end if**
44: **end for**

Fig. 4 shows the estimated ENF signals using the conventional STFT and the proposed method for a 5-minute audio clip, and they are compared with the ground truth ENF acquired by the standard technique. Here, the hop size (hS) was set as 1 second. A fixed segment size of 64 was used for both the conventional technique and the second phase of the proposed method; hence, the largest size of adaptive segments was specified as 62 seconds. The smallest adaptive segment size was set to 16 seconds. As is evident in the figure, the proposed strategy effectively extended the ENF signal of the traditional approach by extracting additional ENF samples from the beginning and end, which the standard STFT method does not consider. Specifically, the proposed technique acquired 24 extra ENF samples for the specified parameters at both ends.

Algorithm 2 presents a pseudocode for estimating the ground truth ENF signal from a reference signal (reduced mains voltage data) using the proposed STFT segmentation scheme. In this algorithm, the reference signal is first divided into partially overlapping clips (similar to the technique shown in Fig. 1) in the length of the query signal, with a hop size of hS. Then, the proposed segmentation process in Fig. 3 (b) is applied to each reference clip with the same parameter settings as the query signal, followed by ENF estimation from each STFT segment of each clip using the same procedure as in Fig. 2. Consequently, consecutive ground truth ENF signal portions, each in the same form as the query ENF signal, are obtained with a time difference of hS. Each succeeding row of this algorithm's output variable $\mathbf{G}$ represents successively the ground truth ENF portion obtained from each reference clip. It should be highlighted in Algorithm 2 that almost all fixed-size (wS) segments acquired for any clip, namely, the second to last fixed-size segments, are common to those for the subsequent clip. Accordingly, to avoid redundant ENF estimations and reduce the computational cost, the ENF samples obtained from these segments for any clip are copied to the ENF index values corresponding to those from the first to the penultimate fixed-size segments of the following clip (lines 35-36 in Algorithm 2). For the first clip, though, the ENF sample for every segment is computed independently of the others through the same process as in Algorithm 1.

The reason for suggesting Algorithm 2 is to acquire a ground truth signal most appropriate with the media ENF signal, estimated through the proposed scheme (i.e., Algorithm 1), in terms of estimation settings. This is significant for performing a reliable similarity test for applications, including time-of-recording verification. To achieve this, the media ENF signal goes through Pearson's correlation test with each ground truth ENF signal portion (i.e., each row of $\mathbf{G}$) rather than using the normalized cross-correlation. This is discussed in more detail in Section IV.

It may be questioned why not only fixed segment sizes are used across the reference signal instead of employing some smaller segment sizes required by the proposed technique because sufficiently large data are already available, and more data often leads to more accurate results. However, obtaining the ground truth ENF signal portions estimated in the same manner as the query signal may be more advantageous, considering that any ENF estimate is a product of the entire data in a segment. Section IV provides comparative results for the media time-stamping task when the ground truth ENF

signal is obtained using the standard technique and when it is acquired through the proposed framework.

IV. EXPERIMENTS AND RESULTS

This section evaluates the proposed method for different audio lengths and STFT segment sizes in the time-of-recording verification task by conducting experiments on the ENF-WHU dataset [43], [45], created in China, where the nominal ENF is 50 Hz. The ENF-WHU dataset consists of 130 audio files of various lengths between 4.5 and 16 minutes, each containing the ENF. A 24-hour reference data, that is, a reduced mains-power voltage signal, for each audio is also supplied with this dataset to acquire the ground truth ENF signal in the desired STFT parameter settings for the day each audio was recorded.

The highest ENF component in the audio recordings arises at the second ENF harmonic, that is, 100 Hz, based on the restricted frequency response range of their recorders [18], [19]. Therefore, the 100 Hz frequency band was used for the ENF signal extraction from each audio. To obtain the ground truth ENF signals from the reference signals, the 50 Hz frequency band was utilized, as it is the strongest ENF component for these signals.

The experiments were performed separately for the first 2-minute clips of all 130 recordings, the first 6-minute clips of 127 recordings (those longer than 6 minutes), and the first 10-minute clips of 82 recordings (those longer than 10 minutes). To estimate the ENF signal from each audio using the conventional STFT technique, fixed segment sizes of 16, 24, 32, 40, 48, 56, and 64 seconds were used separately. A one-second hop size (hS) was set for each set of segments. Therefore, a 1-sample/second ENF resolution was employed for each ENF signal. The same fixed-size segments, except for the 16 seconds, were also used for the second phase of the proposed adaptive segmentation scheme, with the same hop size. The proposed scheme was not considered for the fixed-size segment of 16 seconds because the smaller segment sizes (14 seconds and less) required in the second and third phases of the proposed technique may not be sufficiently appropriate to suppress the effect of noise on the ENF. For any specified fixed segment size (for the second phase), a range of sizes was set and tested individually for the minimum segment (for the first and third phases), depending on the desired amount of extension in the ENF signal. For instance, for a fixed segment size of 32 seconds, the smallest size was set to 30 seconds for an extension of one sample (by using a new segment of 30 seconds) at both ends, whereas it was set to 26 seconds for an extension of four samples (by using the additional segments of sizes 30, 28, 26, and 24 seconds at both ends. It should be recalled that there is a $2 \times hS$ difference between successive segments in the first and third phases of the proposed approach.

To acquire the ground truth signals for the audio-ENF signals that were extracted using the conventional segmentation scheme, each reference signal, as a whole, was applied the same segmentation and ENF estimation procedure as the audio, with identical STFT parameter settings. To obtain the ground truth signals for those estimated through the proposed scheme, each reference signal was first segmented into overlapping data portions of the same size as the audio clips, with a hop size of hS. Then, each data portion was individually put into the proposed segmentation scheme, the same as the audio (Fig. 3 (b)), followed by ENF estimation from each.

To test the similarity between each audio-ENF signal and the corresponding ground truth signal obtained through the traditional STFT technique, normalized cross-correlation (NCC) was used. Eq. 1 provides the expression for computing the kth NCC coefficient ($r(k)$) as follows:

$$r(k) = \frac{\sum_n \left[F(n) - \mu_F^k\right]\left[E(n-k) - \mu_E\right]}{\sqrt{\sum_n \left[F(n) - \mu_F^k\right]^2 \sum_n \left[E(n-k) - \mu_E\right]^2}} \quad (1)$$

where $E(n-k)$ denotes the audio ENF signal with k sample delays; μ_E is the mean value of E; $F(n)$ is the ground truth ENF signal (i.e., obtained by the standard STFT scheme); and μ_F^k represents the mean value of a clip of F, which starts from the kth sample and which is in the length of the test audio.

For the similarity analysis of the ENF signals that were estimated using the proposed technique, each audio-ENF was put into the Pearson correlation coefficient (PCC) test with each of their associated ground truth ENF portions one by one. Eq. 2 demonstrates the expression for calculating the ith PCC coefficient ($\rho(i)$) as follows:

$$\rho(i) = \frac{\sum_n \left[G^i(n) - \mu_G^i\right]\left[E(n) - \mu_E\right]}{\sqrt{\sum_n \left[G^i(n) - \mu_G^i\right]^2 \sum_n \left[E(n) - \mu_E\right]^2}} \quad (2)$$

where $G^i(n)$ denotes the ith ground truth ENF portion (i.e., the ith row of $\mathbf{G}$ in Algorithm 2), and μ_G^i is the mean of G^i. Here, no delay is applied to the $E(n)$, unlike that in Eq. 1, because both $E(n)$ and $G^i(n)$ are of the same length.

For each test with either similarity metric, the lag point (time index) of the maximum correlation coefficient was checked to determine whether it corresponded to the recording time of the query audio, and if it did, it was considered a correct match. More specifically, if the sum of the lag point and the initial time of the reference signal matched the initial time of the audio (within a tolerance of 15 seconds, in accordance with [17], [45], and [48]), it was concluded that the recording time was verified.

The rates of correct matches for the 2-minute audio-ENF signals that were extracted using the conventional STFT segmentation scheme ([6]) for fixed-segment sizes ranging from 16 to 64 seconds are given in line 0 in Table 1. Lines 1 to 24 in Table 1 show the outcomes for the ENF signals estimated using the proposed technique for ENF extensions of 1 to 24 samples from both ends, respectively. The minimum size of the new segments was set to 16 seconds in connection with the previously highlighted point regarding the inappropriateness of the smaller segment sizes.

TABLE 1. An evaluation of the proposed technique's effectiveness in time-stamping 130 2-minute audio clips when it is used both for query and reference signals for different STFT segment sizes.

ENF extension (in samples)	Sizes of fixed STFT segments (in seconds) and true decision rates (in %)						
	16	24	32	40	48	56	64
0 ([6])	53.08	57.69	54.62	45.38	23.08	11.54	6.15
1	-	58.46	54.62	46.15	27.69	14.62	5.38
2	-	57.69	56.92	48.46	29.23	15.38	6.15
3	-	58.46	57.69	49.23	30.77	17.69	6.92
4	-	56.92	58.46	51.54	33.08	17.69	8.46
5	-	-	60.00	53.08	36.92	16.15	10.00
6	-	-	60.77	52.31	38.46	18.46	10.00
7	-	-	**61.54**	53.85	40.00	20.00	10.77
8	-	-	60.00	56.92	41.54	23.08	11.54
9	-	-	-	56.15	44.62	23.08	13.85
10	-	-	-	58.46	46.92	26.15	13.85
11	-	-	-	58.46	49.23	26.15	13.85
12	-	-	-	56.92	51.54	28.46	13.85
13	-	-	-	-	54.62	32.31	16.15
14	-	-	-	-	54.62	36.15	19.23
15	-	-	-	-	55.38	37.69	20.77
16	-	-	-	-	51.54	40.00	22.31
17	-	-	-	-	-	42.31	25.38
18	-	-	-	-	-	43.08	26.92
19	-	-	-	-	-	46.92	29.23
20	-	-	-	-	-	46.15	30.77
21	-	-	-	-	-	-	33.08
22	-	-	-	-	-	-	31.54
23	-	-	-	-	-	-	34.62
24	-	-	-	-	-	-	36.92

- Line 0 refers to the results obtained by using the traditional STFT technique [6].
- Lines 1 to 24 represent the outcomes achieved by using the proposed STFT scheme.
- Red and bold black denote the best performances obtained by the conventional and proposed methods, respectively.

Therefore, not all the extensions in the table are suitable. As shown in Table 1, the proposed technique increases the rate of correct matches for almost every fixed-segment size compared to the traditional scheme. The best match rates for 2-minute clips were obtained as 57.69% and 61.54% for the fixed-segment sizes of 24 and 32 seconds, respectively, using the classical method and the proposed technique with seven samples of ENF extension from both ends. It should be recalled that a fixed-size segment represents every segment for the standard segmentation, although it refers to those in the second phase of the proposed scheme. Similar experiments were repeated for the 6-minute and 10-minute audio clips. As shown in Table 2 and Table 3, both the conventional and proposed techniques achieved their best results using a fixed-segment size of 64 seconds for each set of clip lengths. While the proposed method succeeded in a true match rate of 87.40% and 93.90%, respectively, for 6-minute and 10-minute clips in a variety of ENF extensions, the standard method achieved 85.04% and 91.46% performance for these clips, respectively.

However, segmenting a reference signal into overlapping data portions of the same size as the audio and applying an ENF estimation process for every single portion using the introduced segmentation scheme (recall Algorithm 2) is computationally inefficient. Moreover, it may not always be possible or available to work with a reduced mains voltage signal to obtain a ground truth ENF through the proposed technique; that is, an existing ground truth ENF signal, already estimated using the conventional technique, may have to be used. Accordingly, the proposed scheme should be applied only to query media. Table 4, Table 5, and Table 6 present the outcomes in such a scenario for the same 2-minute, 6-minute, and 10-minute audio clips, respectively. As can be seen from Table 5, and Table 6, the results for the 6-minute and 10-minute clips are comparable to those in Tables 2 and 3. However, the acquired results for the 2-minute clips, shown in Table 4, are not as good as those in Table 1; that is, worse than what was achieved when the ground truth was computed through the adaptive segmentation scheme. However, even in this case, the outcomes of the adopted technique (lines 1 to 24 in Table 4) are still superior to those obtained when the audio-ENF and ground truth ENF were both acquired using the conventional approach (line 0 in Table 4). The proposed technique obtained the best results as 59.23%, 87.40%, and 91.46% true match rates, respectively, for 2-minute clips with 32-second segments, 6-minute clips with 64-second segments, and 10-minute clips with 64-second segments. For the top results settings, the proposed scheme increased the operational time by approximately 0.02 seconds, 0.03 seconds, 0.06 seconds, and 0.9 seconds, respectively, for the 2-minute clips, 6-minute clips, and 10-minute clips, compared to the standard segmentation.

The above experiments show that while the proposed method can work effectively in media of various lengths, it is most efficient in short media. This outcome is actually what was expected, given that the proposed strategy mitigates the adverse effects of ENF pattern similarities that are more pronounced in shorter ENF signals.

TABLE 2. An evaluation of the proposed technique's effectiveness in time-stamping 127 6-minute audio clips when it is used both for query and reference signals for different STFT segment sizes.

ENF extension (in samples)	Sizes of fixed STFT segments (in seconds) and true decision rates (in %)						
	16	24	32	40	48	56	64
0 ([6])	70.08	74.80	77.17	80.31	84.25	84.25	85.04
1	-	73.23	76.38	80.31	84.25	84.25	85.83
2	-	75.59	76.38	80.31	84.25	84.25	85.83
3	-	75.59	76.38	81.10	84.25	84.25	85.83
4	-	76.38	76.38	80.31	84.25	84.25	86.61
5	-	-	76.38	78.74	84.25	84.25	86.61
6	-	-	76.38	77.95	84.25	84.25	86.61
7	-	-	76.38	77.95	84.25	84.25	86.61
8	-	-	76.38	77.95	83.46	84.25	86.61
9	-	-	-	78.74	82.68	85.04	**87.40**
10	-	-	-	78.74	81.89	85.04	**87.40**
11	-	-	-	78.74	81.89	85.04	**87.40**
12	-	-	-	78.74	81.89	84.25	**87.40**
13	-	-	-	-	81.89	83.46	**87.40**
14	-	-	-	-	81.89	81.89	**87.40**
15	-	-	-	-	81.89	82.68	**87.40**
16	-	-	-	-	81.89	82.68	86.61
17	-	-	-	-	-	81.89	85.83
18	-	-	-	-	-	81.89	84.25
19	-	-	-	-	-	81.89	84.25
20	-	-	-	-	-	83.46	85.04
21	-	-	-	-	-	-	85.04
22	-	-	-	-	-	-	85.04
23	-	-	-	-	-	-	85.83
24	-	-	-	-	-	-	85.83

- Line 0 refers to the results obtained by using the traditional STFT technique [6].
- Lines 1 to 24 represent the outcomes achieved by using the proposed STFT scheme.
- Red and bold black denote the best performances obtained by the conventional and proposed methods, respectively.

TABLE 3. An evaluation of the proposed technique's effectiveness in time-stamping 82 10-minute audio clips when it is used both for query and reference signals for different STFT segment sizes.

ENF extension (in samples)	Sizes of fixed STFT segments (in seconds) and true decision rates (in %)						
	16	24	32	40	48	56	64
0 ([6])	74.39	76.83	84.15	84.15	86.59	91.46	91.46
1	-	76.83	85.37	84.15	86.59	91.46	91.46
2	-	78.05	86.59	84.15	85.37	91.46	91.46
3	-	79.27	85.37	84.15	86.59	91.46	91.46
4	-	80.49	85.37	84.15	86.59	91.46	92.68
5	-	-	85.37	84.15	86.59	91.46	92.68
6	-	-	85.37	84.15	86.59	91.46	92.68
7	-	-	85.37	84.15	87.80	91.46	**93.90**
8	-	-	85.37	84.15	87.80	91.46	**93.90**
9	-	-	-	84.15	87.80	92.68	**93.90**
10	-	-	-	84.15	89.02	92.68	**93.90**
11	-	-	-	84.15	89.02	92.68	**93.90**
12	-	-	-	84.15	90.24	92.68	**93.90**
13	-	-	-	-	90.24	91.46	**93.90**
14	-	-	-	-	90.24	91.46	**93.90**
15	-	-	-	-	90.24	91.46	**93.90**
16	-	-	-	-	90.24	91.46	**93.90**
17	-	-	-	-	-	91.46	**93.90**
18	-	-	-	-	-	91.46	**93.90**
19	-	-	-	-	-	91.46	**93.90**
20	-	-	-	-	-	91.46	**93.90**
21	-	-	-	-	-	-	**93.90**
22	-	-	-	-	-	-	**93.90**
23	-	-	-	-	-	-	**93.90**
24	-	-	-	-	-	-	92.68

- Line 0 refers to the results obtained by using the traditional STFT technique [6].
- Lines 1 to 24 represent the outcomes achieved by using the proposed STFT scheme.
- Red and bold black denote the best performances obtained by the conventional and proposed methods, respectively.

However, the performance may decrease slightly for some ENF sample extensions because of the uncertainties associated with the noise effect arising from the additional data segments of smaller sizes. Determining the optimal ENF extension for a specified fixed-segment size may be challenging because it may differ for different query clips

TABLE 4. An evaluation of the proposed technique's effectiveness in time-stamping 130 2-minute audio clips when it is used only for the query signals, i.e., not for the reference signals, for different STFT segment sizes.

ENF extension (in samples)	Sizes of fixed STFT segments (in seconds) and true decision rates (in %)						
	16	24	32	40	48	56	64
0 ([6])	53.08	57.69	54.62	45.38	23.08	12.31	5.38
1	-	58.46	54.62	45.38	26.92	15.38	4.62
2	-	56.92	56.92	47.69	29.23	14.62	5.38
3	-	58.46	56.92	50.00	28.46	15.38	6.15
4	-	56.92	57.69	51.54	30.00	16.15	8.46
5	-	-	**59.23**	53.08	32.31	15.38	6.92
6	-	-	**59.23**	53.85	34.62	17.69	6.92
7	-	-	56.92	54.62	36.92	20.77	6.15
8	-	-	**59.23**	54.62	37.69	22.31	6.15
9	-	-	-	56.92	40.00	21.54	6.92
10	-	-	-	54.62	40.00	21.54	6.15
11	-	-	-	54.62	40.77	21.54	8.46
12	-	-	-	53.85	46.15	23.08	7.69
13	-	-	-	-	46.92	22.31	10.00
14	-	-	-	-	43.85	23.08	10.00
15	-	-	-	-	44.62	26.15	10.77
16	-	-	-	-	44.62	29.23	13.85
17	-	-	-	-	-	30.77	14.62
18	-	-	-	-	-	32.31	13.08
19	-	-	-	-	-	33.08	13.85
20	-	-	-	-	-	33.08	13.08
21	-	-	-	-	-	-	14.62
22	-	-	-	-	-	-	13.85
23	-	-	-	-	-	-	19.23
24	-	-	-	-	-	-	18.46

- Line 0 refers to the results obtained by using the traditional STFT technique [6].
- Lines 1 to 24 represent the outcomes achieved by using the proposed STFT scheme.
- Red and bold black denote the best performances obtained by the conventional and proposed methods, respectively.

TABLE 5. An evaluation of the proposed technique's effectiveness in time-stamping 127 6-minute audio clips when it is used only for the query signals, i.e., not for the reference signals, for different STFT segment sizes.

ENF extension (in samples)	Sizes of fixed STFT segments (in seconds) and true decision rates (in %)						
	16	24	32	40	48	56	64
0 ([6])	70.08	74.80	77.17	80.31	84.25	84.25	85.04
1	-	73.23	76.38	80.31	84.25	84.25	85.83
2	-	75.59	76.38	80.31	84.25	84.25	85.83
3	-	75.59	76.38	81.10	84.25	84.25	85.83
4	-	76.38	76.38	80.31	84.25	84.25	86.61
5	-	-	76.38	78.74	84.25	84.25	86.61
6	-	-	76.38	77.95	84.25	84.25	86.61
7	-	-	76.38	77.95	84.25	84.25	86.61
8	-	-	76.38	77.95	83.46	84.25	86.61
9	-	-	-	78.74	82.68	85.04	**87.40**
10	-	-	-	78.74	81.89	85.04	**87.40**
11	-	-	-	78.74	81.89	85.04	**87.40**
12	-	-	-	78.74	81.89	84.25	**87.40**
13	-	-	-	-	81.89	83.46	**87.40**
14	-	-	-	-	81.89	81.89	**87.40**
15	-	-	-	-	81.89	82.68	**87.40**
16	-	-	-	-	81.89	82.68	86.61
17	-	-	-	-	-	81.89	85.83
18	-	-	-	-	-	81.89	84.25
19	-	-	-	-	-	81.89	84.25
20	-	-	-	-	-	82.68	84.25
21	-	-	-	-	-	-	84.25
22	-	-	-	-	-	-	85.04
23	-	-	-	-	-	-	85.04
24	-	-	-	-	-	-	85.04

- Line 0 refers to the results obtained by using the traditional STFT technique [6].
- Lines 1 to 24 represent the outcomes achieved by using the proposed STFT scheme.
- Red and bold black denote the best performances obtained by the conventional and proposed methods, respectively.

of different media depending on the recording conditions. Nevertheless, preferring the largest ENF extension possible may be a wise choice because a longer ENF signal becomes more distinct and more useful in most cases. The selection of an ideal fixed segment size may also be challenging. However, based on the experimental results, smaller segment

TABLE 6. An evaluation of the proposed technique's effectiveness in time-stamping 82 10-minute audio clips when it is used only for the query signals, i.e., not for the reference signals, for different STFT segment sizes.

ENF extension (in samples)	Sizes of fixed STFT segments (in seconds) and true decision rates (in %)						
	16	24	32	40	48	56	64
0 ([6])	74.39	76.83	84.15	84.15	86.59	91.46	91.46
1	-	76.83	85.37	84.15	86.59	91.46	91.46
2	-	78.05	86.59	84.15	85.37	91.46	91.46
3	-	79.27	85.37	84.15	86.59	91.46	91.46
4	-	80.49	85.37	84.15	86.59	91.46	92.68
5	-	-	85.37	84.15	86.59	91.46	92.68
6	-	-	85.37	84.15	86.59	91.46	92.68
7	-	-	85.37	84.15	87.80	91.46	**93.90**
8	-	-	85.37	84.15	87.80	91.46	**93.90**
9	-	-	-	84.15	87.80	92.68	**93.90**
10	-	-	-	84.15	89.02	92.68	**93.90**
11	-	-	-	84.15	89.02	92.68	**93.90**
12	-	-	-	84.15	89.02	92.68	**93.90**
13	-	-	-	-	89.02	91.46	**93.90**
14	-	-	-	-	89.02	91.46	**93.90**
15	-	-	-	-	89.02	91.46	**93.90**
16	-	-	-	-	89.02	91.46	**93.90**
17	-	-	-	-	-	91.46	**93.90**
18	-	-	-	-	-	91.46	**93.90**
19	-	-	-	-	-	91.46	**93.90**
20	-	-	-	-	-	91.46	**93.90**
21	-	-	-	-	-	-	**93.90**
22	-	-	-	-	-	-	**93.90**
23	-	-	-	-	-	-	**93.90**
24	-	-	-	-	-	-	**93.90**

- Line 0 refers to the results obtained by using the traditional STFT technique [6].
- Lines 1 to 24 represent the outcomes achieved by using the proposed STFT scheme.
- Red and bold black denote the best performances obtained by the conventional and proposed methods, respectively.

TABLE 7. An assessment of the proposed approach when used in conjunction with current ENF enhancement strategies.

Clip length (in min.)	Segment size (in sec.)	Number of audio	Adopted technique and true decision rates (in %)			
			[43]	[43]+Proposed	[45]	[45]+Proposed
10	64	82	100	100	97.56	98.78
6	64	127	89.76	89.76	91.34	93.70
2	32	130	55.38	63.08	70.77	74.62
2	24	130	60.00	63.85	74.62	77.69

sizes for shorter media and larger segment sizes for longer media are expected typically to perform better.

V. DISCUSSIONS AND EXTENSIONS

As previously stated, the objective of the proposed STFT segmentation scheme in Section III is not to improve the quality (i.e., accuracy) of the ENF signal to be estimated from media but rather to increase its distinctiveness by expanding it through the extraction of extra ENF samples from additional data segments that are not considered by the conventional strategy. Because the main procedure is in the data segmentation stage of the STFT, the suggested technique can be used in conjunction with any pre-process, intermediate process (during ENF estimation from any segment), or post-process ENF enhancement techniques proposed in the literature. In other words, it can be integrated into any STFT-based ENF signal refinement strategy to further improve the performance of ENF-based forensic applications.

Table 7 provides an evaluation of the proposed technique, using the ENF-WHU dataset, when combined with the robust filtering algorithm (RFA) [43] and enhanced maximum likelihood estimator (E-MLE) [45] that were proposed recently. The 10-minute and 6-minute clips were tested for a fixed segment size of 64 seconds because both the conventional and proposed techniques achieved their best performance for this setting in the experiments in Section IV. The 2-minute clips were experimented with for both 32 and 24 seconds of fixed segment sizes because the best performances for this set of clips were obtained in different segment sizes for the proposed and traditional schemes, that is, 32 seconds for the proposed technique and 24 seconds for the classical method. Each method used the same ground truth ENF signals obtained using the standard STFT technique. As evident from the table (i.e., Table 7), when integrated with the RFA [43], the proposed segmentation scheme considerably raised the true match rate for 2-minute clips, from 55.38% to 63.08% for 24-second segments and from 60.00% to 63.85% for 32-second segments, compared to the RFA alone. For the 6-minute and 10-minute clips, the performance did not change. When combined with the E-MLE [45], the proposed technique increased the true

match rate against the single E-MLE for any settings: from 70.77% to 74.62% for 2-minute clips with 24-second segments, from 74.62% to 77.69% for 2-minute clips with 32-second segments, from 91.34% to 93.70% for 6-minute clips with 64-second segments and from 97.56% to 98.78% for 10-minute clips with 64-second segments. Consequently, the proposed segmentation scheme is noticeably effective in further improving the performance of existing ENF enhancement strategies.

Regarding the computational cost, the operational time for the 2-minute clips with 24-second segments, 2-minute clips with 32-second segments, 6-minute clips with 64-second segments, and 10-minute clips with 64-second segments was increased by approximately 0.86, 1.00, 5.16, and 9.78 seconds, respectively, when the proposed segmentation scheme was combined with the RFA [43], compared to that when the RFA was used with the standard segmentation. When the proposed technique was integrated with the E-MLE [45], the operational time rose by approximately 0.47, 0.88, 2.82, and 3.43 seconds for 2-minute clips with 24-second segments, 2-minute clips with 32-second segments, 6-minute clips with 64-second segments, and 10-minute clips with 64-second segments, respectively, in comparison with that when the E-MLE was used with the standard scheme. All the computations were performed through a 10th-generation Intel i5 processor.

It should be noted that the proposed technique exploited, for the above experiments, the largest ENF extension possible for each adopted segment size using a minimum segment size of 16 seconds, as discussed and recommended in Section IV.

VI. CONCLUSION

This work proposed an enhanced STFT segmentation scheme to estimate the ENF signal effectively for use in media forensics. Compared to the conventional STFT, the presented method ingeniously constitutes additional data segments at the beginning and end of the media to achieve extra ENF estimates for any selected STFT parameters, resulting in an expanded and, thus, more distinct ENF signal. To build a ground truth ENF signal with settings equivalent to the audio ENF signal, an adaptation of the proposed technique to a reference signal was also introduced. Large-scale time-stamp verification tests were conducted to evaluate the proposed method, using the ENF-WHU audio dataset for various clip lengths and STFT segment sizes. Experimental results demonstrate that the proposed approach outperforms the traditional STFT scheme. The experiments also showed that when integrated with any existing STFT-based ENF-accuracy enhancement strategy, the proposed method is considerably effective in further boosting the performance.

While the proposed technique leads to more effective ENF signal estimations, it has some disadvantages or challenges. First, it increases the time complexity owing to additional computations for the new segments. In particular, it is computationally inefficient when applied to a large reference signal to obtain a ground truth ENF signal with settings identical to the media ENF for the similarity tests. Fortunately, using a ground truth ENF signal obtained through the standard segmentation (i.e., exploiting the proposed scheme for media only) was an excellent trade-off to avoid the high computational cost burden with a slight performance drop. Second, although the proposed work was introduced for STFT segment sizes longer than 16 seconds because the shorter segment sizes were assumed to be unreliable for accurate ENF estimations, there may be cases where a shorter segment size is more practical, specifically for short media. In such circumstances, the proposed method may be ineffective. Lastly, determining the most appropriate size for the ENF extension is a challenge for the proposed method. It varies mainly depending on the media itself (i.e., how noisy it is), the length of the media, and the selected segment size. A potential area of future research may be investigating the ideal ENF extension depending on the clip length and selected STFT parameters.

REFERENCES

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

[2] 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.

[3] C. Grigoras, A. Cooper, and M. Michalek, *Best Practice Guidelines for ENF Analysis in Forensic Authentication of Digital Evidence*, document FSAAWG-BPM-ENF-001, Forensic Speech and Audio Analysis Working Group, 2009, pp. 1–10.

[4] 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.

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

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

[7] M. Huijbregtse and Z. Geradts, "Using the ENF criterion for determining the time of recording of short digital audio recordings," in *Proc. 3rd Int. Workshop Comput. Forensics*, 2009, pp. 116–124.

[8] E. B. Brixen, "Techniques for the authentication of digital audio recordings," in *Proc. Audio Eng. Soc., Conv. 122*, May 2007, pp. 1–8.

[9] J. Chai, F. Liu, Z. Yuan, R. Conners, and Y. Liu, "Source of ENF in battery-powered digital recordings," in *Proc. Audio Eng. Soc. Conv. 135*, 2013, pp. 1–7.

[10] 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 Manag., IT Forensics (IMF)*, May 2014, pp. 3–13.

[11] 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.

[12] 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*, Nov. 2011, pp. 23–32.

[13] 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.

[14] 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)*, Oct. 2014, pp. 5367–5371.

[15] M. Wu, A. Hajj-Ahmad, and H. Su, "Techniques to extract ENF signals from video image sequences exploiting the rolling shutter mechanism; and a new video synchronization approach by matching the ENF signals extracted from soundtracks and image sequences," U.S. Patent 9 916 857 B2, Dec. 10, 2015.

[16] 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.
[17] 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.
[18] 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.
[19] 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.
[20] 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.
[21] S. Vatansever, "A thorough investigation into the ENF reconstruction in videos exposed by rolling shutter," *IEEE Access*, vol. 11, pp. 96330–96342, 2023.
[22] L. Xu, G. Hua, H. Zhang, L. Yu, and N. Qiao, "'Seeing' electric network frequency from events," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 18022–18031.
[23] L. Xu, G. Hua, H. Zhang, and L. Yu, "'Seeing' ENF from neuromorphic events: Modeling and robust estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Apr. 9, 2024, doi: 10.1109/TPAMI.2024.3386813.
[24] 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.
[25] 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.
[26] 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.
[27] 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.
[28] S. Vatansever, A. E. Dirik, and N. Memon, "ENF based robust media time-stamping," *IEEE Signal Process. Lett.*, vol. 29, pp. 1963–1967, 2022.
[29] S. Vatansever, "Modern techniques in forensic analysis of multimedia signals," Ph.D. dissertation, Dept. Electron. Eng., Bursa Uludag Univ., Bursa, Turkey, 2019.
[30] 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.
[31] 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.
[32] 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 Experiences*, Nov. 2014, pp. 13–18.
[33] 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.
[34] 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.
[35] 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.
[36] 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.
[37] 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.
[38] E. Ngharamike, L.-M. Ang, K. P. Seng, and M. Wang, "Exploiting the rolling shutter read-out time for ENF-based camera identification," *Appl. Sci.*, vol. 13, no. 8, p. 5039, Apr. 2023.
[39] G. Hua, Q. Wang, D. Ye, H. Zhang, G. Wang, and S. Xia, "Factors affecting forensic electric network frequency matching—A comprehensive study," *Digit. Commun. Netw.*, Jan. 2023, doi: 10.1016/j.dcan.2023.01.009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352864823000226
[40] E. Ngharamike, L.-M. Ang, K. P. Seng, and M. Wang, "ENF based digital multimedia forensics: Survey, application, challenges and future work," *IEEE Access*, vol. 11, pp. 101241–101272, 2023.
[41] 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.
[42] 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.
[43] G. Hua and H. Zhang, "ENF signal enhancement in audio recordings," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 1868–1878, 2020.
[44] 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.
[45] 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.
[46] C. Korgialas, C. Kotropoulos, and K. N. Plataniotis, "Leveraging electric network frequency estimation for audio authentication," *IEEE Access*, vol. 12, pp. 9308–9320, 2024.
[47] 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.
[48] 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.

ALI BERK YALINKILIÇ received the B.Sc. and M.Sc. degrees in mechatronics engineering from Bursa Technical University, Bursa, Türkiye, in 2019 and 2024, respectively. He is currently working in a software company. His research interests include electronics, data science, artificial intelligence, and forensics.

SAFFET VATANSEVER (Member, IEEE) received the B.Sc. degree in electronics and communications engineering from Yıldız Technical University, Bursa, Türkiye, the M.Sc. degree in mechatronics engineering from the University of Newcastle upon Tyne, Newcastle, U.K., and the Ph.D. degree in electronics engineering from Bursa Uludağ University, Bursa. He is currently an Assistant Professor with the Department of Mechatronics, Bursa Technical University, Bursa. His research interests include multimedia forensics, signal processing, and machine learning.