applied sciences

MDPI

Article

Exploiting the Rolling Shutter Read-Out Time for ENF-Based Camera Identification

Ericmoore Ngharamike [1], Li-Minn Ang [1,*], Kah Phooi Seng [1,2] and Mingzhong Wang [1]

[1] School of Science, Technology, and Engineering, University of the Sunshine Coast, Petrie, QLD 4502, Australia; ericmoore.ngharamike@research.usc.edu.au (E.N.); jasmine.seng@xjtlu.edu.cn (K.P.S.); mwang@usc.edu.au (M.W.)
[2] School of AI and Advanced Computing, Xian Jiaotong Liverpool University, Suzhou 215123, China
* Correspondence: lang@usc.edu.au

Abstract: The electric network frequency (ENF) is a signal that varies over time and represents the frequency of the energy supplied by a mains power system. It continually varies around a nominal value of 50/60 Hz as a result of fluctuations over time in the supply and demand of power and has been employed for various forensic applications. Based on these ENF fluctuations, the intensity of illumination of a light source powered by the electrical grid similarly fluctuates. Videos recorded under such light sources may capture the ENF and hence can be analyzed to extract the ENF. Cameras using the rolling shutter sampling mechanism acquire each row of a video frame sequentially at a time, referred to as the read-out time (T_{ro}) which is a camera-specific parameter. This parameter can be exploited for camera forensic applications. In this paper, we present an approach that exploits the ENF and the T_{ro} to identify the source camera of an ENF-containing video of unknown source. The suggested approach considers a practical scenario where a video obtained from the public, including social media, is investigated by law enforcement to ascertain if it originated from a suspect's camera. Our experimental results demonstrate the effectiveness of our approach.

Keywords: electric network frequency (ENF); rolling shutter mechanism; read-out time, T_{ro}; camera forensics; camera sensor types; ENF extraction

Citation: Ngharamike, E.; Ang, L.-M.; Seng, K.P.; Wang, M. Exploiting the Rolling Shutter Read-Out Time for ENF-Based Camera Identification. *Appl. Sci.* **2023**, *13*, 5039. https://doi.org/10.3390/app13085039

Academic Editor: Roberto Saia

Received: 9 March 2023
Revised: 4 April 2023
Accepted: 11 April 2023
Published: 17 April 2023

1. Introduction

Digital multimedia materials, that is, audio, image, and video recordings, contain vast amounts of information which can be exploited in forensic investigations. In light of how digital manipulation techniques are always developing and how they may have far-reaching effects on different spheres of society and the economy, the field of digital forensics has seen an increasing growth in recent decades. In order to combat multimedia forgeries and ensure the authenticity of multimedia material, researchers have focused on developing new methods in the field of digital forensics.

Electric network frequency (ENF) has been utilized in recent years as a tool in forensic applications. Analysis of the ENF is a forensic tool used to verify the authenticity of multimedia recordings and spot any attempts at manipulation [1–4]. The ENF is the supply frequency of an electric power grid and varies in time around its nominal value of 60 Hz in North America and 50 Hz in Europe, Australia, and much of the rest of the world as a result of inconsistencies between power network supply and demand [1,5]. The nature of these inconsistencies can be observed to be random, unique per time, and usually quite the same across all locations connected by the same power grid. Consequently, an ENF signal recorded at any location in time, connected to a certain mains power can serve as a reference ENF signal for the entire region serviced by that power network for that period of time [6,7].

The ENF's fluctuating/instantaneous values over time is considered as an ENF signal. 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 [1,7,8]. This ENF signal can be estimated from the recordings using time-domain or frequency-domain techniques and utilized for various forensic and anti-forensic applications, such as time-stamp verification [5,8,9], audio/video authentication [10,11], location of recording estimation [12–14], power grid identification [15–20], and estimation of camera read-out time [21]. New studies have found that ENF analysis may be employed in other areas of multimedia signal processing, such as in audio and video record synchronization [22], historical audio recording alignment [23], and video synchronization without overlapping scenes [22].

Studies have recently shown that ENF signals can also be extracted from video recordings made under the illumination of a light source powered by a mains grid [5,24–27]. Fluorescent lights and incandescent bulbs used in indoor lighting fluctuate in light intensity at double the supply frequency, causing a nearly impossible to notice flickering that occurs in the illuminated environment. As a result, videos captured under indoor illumination settings using a camera may contain ENF signals. However most commonly/widely used cameras sensors do not capture light in the same manner. Charge-coupled device (CCD) sensors commonly associated with global shutter mechanisms capture all the pixels in a video frame at the same time/instant. Unlike CCDs, complementary metal oxide semiconductor (CMOS) sensors often have a rolling shutter mechanism that causes the sensor to scan the rows of each frame sequentially so that various rows of the same frame are exposed at slightly different instants [26,28].

One of the main issues confronted in the extraction of the embedded ENF signals from video recordings (particularly for videos captured with CCD cameras) is the problem of aliasing to DC due to insufficient/inadequate sampling rates. However, the authors of [25,29] demonstrated that the rolling shutter mechanism, although considered traditionally detrimental to the analysis of image and video, could be exploited to enhance the effective sampling rate for ENF extraction from video recording. Due to the sequential acquisition of the rows of a frame at distinct time instants, referred to as the read-out time T_{ro}, the rolling shutter can facilitate the extraction of ENF signals from video recordings by raising the effective sample rate by a factor of the number of rows to prevent the aliasing effect.

The T_{ro} is the amount of time it takes the camera to capture the rows of a single frame and it is a camera-specific parameter that can be leveraged to characterize CMOS cameras with rolling shutter mechanisms. The author of [30] proposed a method that can extract a flicker signal and the camera T_{ro} from a pirated video and utilize them to identify the LCD screen and camera used to create the pirated movie. Owing to the similarities between the flicker signal and the ENF signal, the authors of [21] appropriated the flicker-based technique in an ENF-based approach that can estimate the T_{ro} value of a camera that creates a video containing an ENF. Their experimental results showed that the approach could estimate the T_{ro} value with high accuracy.

In this work, inspired by [29], we present an approach that exploits the T_{ro} and the ENF to identify the camera used to record an ENF-containing video. Figure 1 shows the block diagram of our proposed method. Our approach considers a practical scenario where a video obtained from the public, including social media, is being investigated by law enforcement to ascertain if it originated from a suspect's camera. In our proposed approach, the ENF extraction method, adapted from [29], conducts an equal uniform sampling over time by returning zeros during the idle period at the end of every frame period to generate the row signals after certain preprocessing. The row signal/reference signal is passed through Fourier analysis (spectrogram) to generate the ENF traces. The ENF is then extracted by finding the dominant instantaneous frequency within a narrow band around the frequency of interest. Quadratic interpolation was utilized to refine the ENF estimate. This method attains a high signal-to-noise ratio (SNR) at the cost of the need to know the read-out time (T_{ro}) parameter, which is camera-model-dependent. Figure 2 shows the block diagram for estimating the ENF signal from video frames.

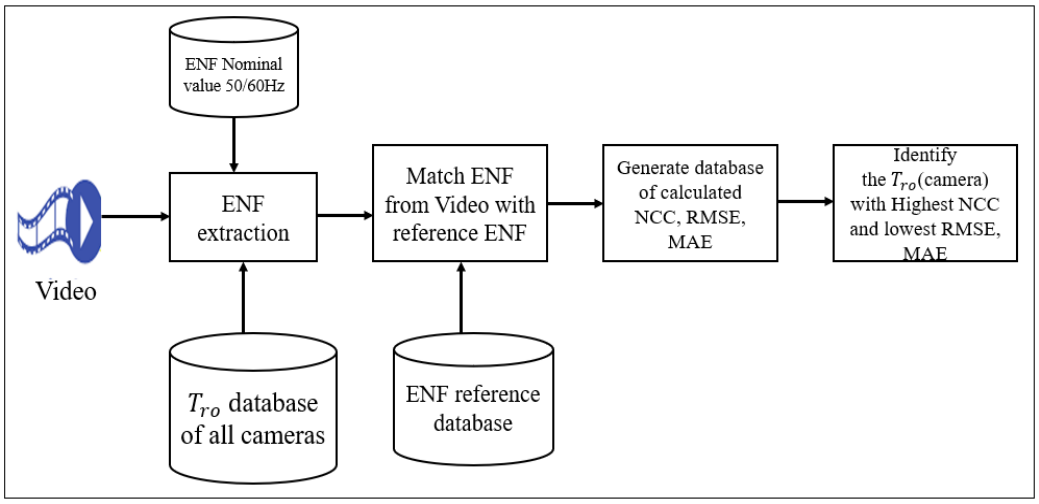

Figure 1. Block diagram of the proposed camera identification approach.

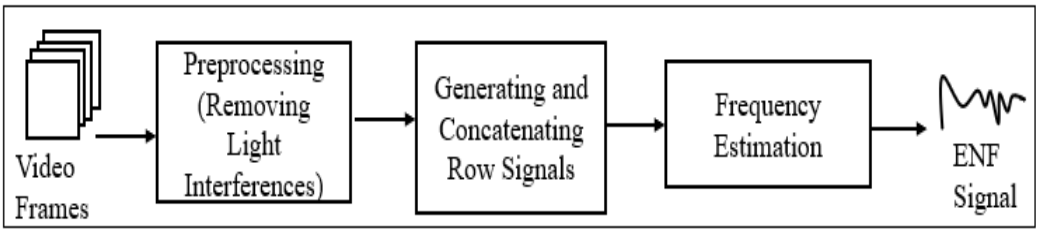

Figure 2. The process of ENF signal estimation from the frames of a video.

The T_{ro} is a sensitive parameter in the method and is employed to compute the number of zeros to be inserted in the idle period for the specific camera [30] or specific video resolution and frame rate [31] that captured the video before analyzing the video, which is critical in estimating an ENF signal with no distortion. Our proposed approach therefore employs a database of camera T_{ros} and ENF reference signals. For any ENF-containing video of an unknown source camera, our approach will analyze the video using different T_{ro} values. The estimated ENF signals are matched against the reference signals, and the performance metrics—normalized cross-correlation (NCC), root mean squared error (RMSE), and mean absolute error (MAE)—are then calculated with respect to the reference ENF and stored in a database. The T_{ro} (camera) with the highest NCC and lowest RMSE and MAE can then be identified. Our experimental analysis reveals that our proposed approach is very useful in identifying a source camera in the intended scenario.

We summarize the contribution of this paper as follows:

- We provide a review of ENF extraction from video recordings and the impact of the rolling shutter on ENF extraction.
- We propose a novel ENF-based approach to identify the rolling shutter camera used to capture an ENF-containing video of unknown source.

The remaining sections of this paper are structured as follows: Section 2 describes basic ENF concepts, reviews the ENF extraction method for video files, explains the impact of the rolling shutter, and discusses camera read-out time and ENF estimation. Section 3 presents the experiment conducted. Section 4 provides the results and discussion. Section 5 concludes the paper.

2. Related Works and Concepts

This section provides a review of basic concepts and relevant studies to aid the comprehension of the procedures involved in the extraction of ENF signals from videos. The relevance of the rolling shutter mechanism and the read-out time, T_{ro}, in camera forensics is also presented.

2.1. ENF Basics

The ENF is the supply frequency of an electric power grid. Generators rotating at a speed of 50 cycles per second in Europe, Australia, and most parts of the world and 60 cycles per second in North America generate an alternating current (AC) which travels along transmission lines [32]. In an ideal situation, the frequency is fixed at a nominal value of 50 Hz or 60 Hz, based on the region. However, as the generation of electricity is dependent on the demand for power, it must be produced proportionately. When the demand is high, the frequency drops, and when it is low, the frequency rises temporarily. The varying value of the frequency as the power generation and consumption rises and drops is regarded as the ENF signal. The following models may be used to describe the instantaneous voltage of the power grid:

$$\begin{aligned} W(t) &= \sqrt{2}W_0 \csc(\sigma(t)) \\ &= \sqrt{2}W_0 \cos\left(2\pi f_p + \theta_i(t) + \alpha_i\right) \\ &= \sqrt{2W_0} \cos\left(2\pi f_p t + 2\pi \int_0^t f_i(\tau) d\tau + \alpha_i\right. \end{aligned} \tag{1}$$

where f_p denotes the nominal frequency (50/60 Hz) and W_0 denotes mains effective voltage, while α_i denotes the initial phase offset [33]. $f_i(t)$ is the instantaneous variation from the nominal frequency, while $\theta_i(t)$ is the instantaneous phase that changes based on power imbalance caused by demand and supply. The instantaneous mains power frequency at time t can be stated from the equations above as:

$$f(t) = \frac{1}{2\pi} \frac{d\sigma(t)}{dt} = f_p + f_i(t) \tag{2}$$

As f_p remains unchanged, the ENF changes based on $f_i(t)$ fluctuations. Utilizing the model in [33], $f_i(t)$ can be expressed as:

$$f_i(t) = \frac{f_p}{2K}(E_s(t) - E_d(t)) \tag{3}$$

where $E_s(t)$ is the amount of power supplied, $E_d(t)$ is the total amount of power demanded plus losses, and K is an inertia constant. So, at each point in time, $f_i(t)$ and $f(t)$ change based on the difference between how much power is produced and how much is used.

2.2. Overview of ENF Extraction from Video

ENF extraction from audio recordings has been extensively researched. Various conventional frequency estimation techniques based on the short-time Fourier transform (STFT) [34] and subspace analysis [35,36] have been implemented to extract the ENF signals from media recordings. STFT is frequently utilized to analyze time-varying signals, including speech signals and ENF signals. With the STFT-based periodogram approach, the dominant instantaneous frequency can be extracted by finding the peak positions and increasing its accuracy using quadratic interpolation [37] and the weighted energy method [5].

In contrast, approaches based on subspace analysis, including multiple signal classification (MUSIC) [35] and estimation of signal parameters using rotational invariance techniques (ESPRIT) [36], leverage the signal subspace's orthogonal relationship to the noise subspace. ESPRIT implementations require lower computation and storage costs, which gives them an advantage over MUSIC [36]. Researchers have also developed dedicated approaches [38,39] and approaches that combine STFT and subspace analysis to accurately estimate the ENF signal [5]. However, some preprocessing is necessary prior to the use of the frequency estimation or tracking techniques mentioned above to accurately extract ENF signals from multimedia audio recordings.

In the case of video recordings, the temporal fluctuation in light intensity included in the video frames may be leveraged to extract the ENF signal. The variations in ENF in the grid network influences the intensity of illumination from any light source connected to

the power mains. As a result of the light source flickering at both the positive and negative cycles of AC current, the frequency of the light becomes twice the frequency of the mains power. The illumination signal may thus be seen as the absolute representation of the cosine function in Equation (1) [24]. For instance, for any video recording made under indoor illumination powered by 50 Hz power mains, since the polarity of the current changes at double the frequency of the mains power, the light flickers at 100 Hz.

In addition, decaying energy's higher harmonics frequently exist at integer multiples of 100 Hz when the mains power signal mildly deviates from a perfect sinusoid. In addition, the higher harmonics bandwidth is larger than the main component since the actual ENF signal of interest is a narrowband signal rather than a completely stable sinusoid. Consequently, the nth harmonic component bandwidth will be n times the ENF component bandwidth at 100 Hz. Hence, the Nyquist theorem states that a sampling rate of at least 200 Hz is required to reliably extract illumination frequency from recorded data [24].

Despite the fact that the majority of consumer cameras cannot offer such high frame sampling rates, the illumination frequency can still be estimated from an aliased frequency. Assuming that f_s is the sampling frequency of the camera and f_ℓ is the light-source illumination frequency, then the aliased illumination frequency f_a is expressed as [30]:

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

Therefore, when a 100 Hz illumination signal from a light source is sampled using a camera with a frame rate of 29.97 Hz, the aliased base frequency of the ENF will be obtained as 10.09 Hz, while the aliased second harmonic will be obtained as 9.79 Hz. The aliased effect as a result of an insufficient camera sampling rate is mainly associated with cameras with the global shutter mechanism.

The first work on ENF extraction from a video recording [5] calculated the average intensity of each frame of a white wall video to produce an intensity signal. This signal over time was then passed through a temporal bandpass filter with a passband matching the desired frequency to extract the ENF. Given that the video's content was largely consistent between frames, the existence of a considerable amount/value of energy in the frequency of interest was attributed to the ENF signal. In the second experiment with video recording with movement, directly averaging the pixel values of the whole frame may not have been an appropriate preprocessing step prior to carrying out frequency analysis because of the inconsistency in the content of each frame of the video. The authors, however, averaged the pixel intensities of relatively steady regions in the video where there was not much inconsistency and extracted the ENF signal from the average pixel intensity spectrogram. The extracted ENF was utilized for time-of-recording and tampering detection applications.

The authors of [22,23,40] 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 the intensity signal which is passed through Fourier analysis to extract the ENF. The extracted ENF is further used for video synchronization applications.

Researchers have also employed image segmentation approaches for analyzing videos for ENF analysis. The authors of [24,26] proposed ENF estimation methods based on exploiting super-pixels. The method reported in [24] computed the average number of steady super-pixels instead of all steady pixels contained in a frame. The method was applied to a video dataset of 160 videos of different lengths recorded under different conditions using cameras that adopted both CMOS and CCD sensors. The method segmented each video into super-pixel regions and identified the steady points within each region throughout all the frames. The steady pixels of all regions of each video frame were then averaged to generate the intensity signal from which a "so-called ENF vector" was computed along each successive video frame of a particular shot. Then, the similarity of the computed ENF vectors was examined to detect the presence or absence of an ENF signal in the test video.

The method described in [26] is based on simple linear iterative clustering (SLIC) [41,42] for image segmentation. The authors employed the SLIC algorithm to generate regions of similar properties termed super-pixels whose average intensity exceeds a certain threshold to generate the mean intensity time series. The authors opined that, in those regions, the embedded ENF is not impeded by interference and noise, including shadows, textures, and brightness, resulting in more precise estimations irrespective of whether the test video is static or non-static. The method was applied to a public dataset of static and non-static CCD videos [43], 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 with particular characteristics to extract the ENF from non-static videos was proposed by the author of [27]. The method first employed a background subtraction algorithm named ViBe [44] to mitigate the deviation brought by movement and applied a differentiator filter at pixel level to eliminate the time-dependent mean value at each point before averaging the pixels. Luminance differences beyond a set threshold are suppressed, requiring only pixels classified as valid in both the current and previous frames to be processed. The frame-level signal is passed to a phase-locked loop-based FM demodulator [45] after preprocessing to extract the ENF, considering its autoregressive manner. The method is applied to estimate the time of recording of CCD videos, and the simulation results showed its effective performance. However, if the variations caused by movement affect a significant number of pixels in every frame, this method may be insufficient for extracting the ENF.

Some other extraction methods take advantage of the higher ENF sampling frequency of the rolling shutter mechanism, which is based on the number of rows multiplied by the video frame rate. However, the rolling shutter method introduces the problem of idle time between consecutive frames where no sampling is performed. Therefore, some light samples are lost during the idle time period which occurs at the end of every frame. The authors of [25] first leveraged the rolling shutter mechanism to address the inadequate sampling rate for ENF extraction from video recordings. They formulated an L-branch filter-bank model of the mechanism for their analysis and showed how the dominant ENF harmonic is shifted to different frequencies as a result of the idle period in videos recorded with the rolling shutter mechanism. To extract the ENF, the method ignores the idle time and 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 the ENF traces. The ENF is then extracted by finding the dominant instantaneous frequency within a narrow band around 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 might result in mild distortion to the estimated ENF traces [46].

To avert such distortion, the authors of [29] proposed a periodic zero-padding approach to deal with the idle time problem. The authors argued that instead of ignoring the idle period, an equally uniform sampling along time should be conducted by returning zeros to the end of each row signal until they reached the length of time that corresponded to the idle period before the concatenation. This zero-padding strategy is capable of producing ENF traces that are free of distortion, but it does need prior knowledge of the length of the idle period. The idle period can be determined using the camera read-out time, which is model-specific [21]. Experimental results demonstrated that the approach enhanced the SNR of the estimated ENF signal.

The authors of [28] introduced a MUSIC combining spectrum method that exploits the rolling shutter mechanism without knowing the idle time. Figure 3 shows the overall framework of the proposed method. Given a test video, the method computes row-by-row the average pixel intensity and concatenates these values in a singular time series, s[n], using a sampling rate of frame rate multiplied by the number of rows.

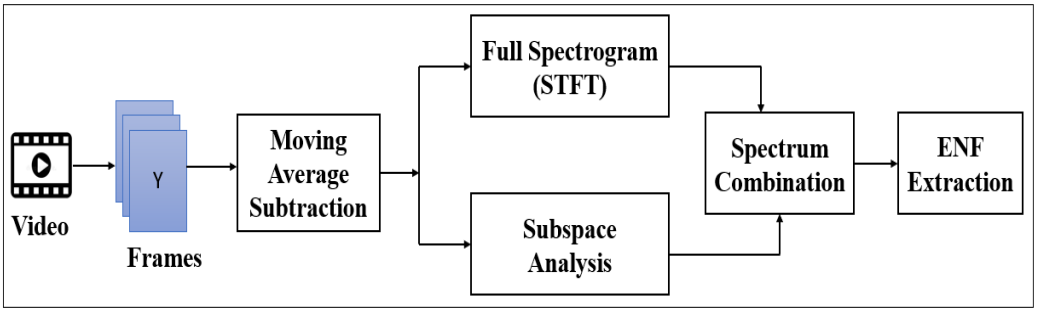

Figure 3. Framework of the proposed MUSIC approach.

The time-series signal, s[n], is then passed through the preprocessing stages, where the local average is subtracted to enable a more precise result, particularly when a video features movement. The signal, s[n], is further down-sampled to 1 kHz after applying an anti-aliasing filter. Using the resulting signal, s[n], MUSIC and Fourier domain analyses are conducted in parallel, and the respective spectra are combined to extract the ENF. The proposed method is applied to time-of-recording verification, and the experimental results showed its effectiveness even in the more difficult scenario of 1 min recordings of both static and non-static videos.

The authors of [47] designed a phase-based method to extract the ENF signal leveraging the rolling shutter mechanism. The method attempted to address the idle time problem by employing row-by-row samples individually at the frame level to avoid any discontinuity caused by the idle time. The proposed method is modeled as:

$$X_j^{\theta}[\ell] = A_j.\sin\left(2\pi\frac{f_j}{f_s}\ell + \varnothing_j\right) + d_j \tag{5}$$

With a parameter set $\theta = (A_j, f_j, \varnothing_j, d_j)$, where A_j is the amplitude, f_j is the light intensity instantaneous frequency, $\varnothing_j$ is the initial phase of the ENF signal when the initial row of the jth frame was obtained, and d_j is the DC value. $f_s = \frac{L}{T_{ro}}$ is the sampling rate, where L is the number of recorded samples during the read-out time and T_{ro} is the read-out time. The phases of the row signals are estimated by fitting/applying the suggested model to each row signal given the jth row signal's L-dimensional vector x_j. The value of the ENF is then calculated from the phase differences between successive row signals after certain optimizations steps.

2.3. Rolling Shutter Impact on ENF Extraction

When using the rolling shutter, the pixels of each frame are captured by scanning over a rectangular CMOS sensor row-by-row rather than recording all of the pixels in a frame at one time instance, as is done with a global shutter. Consequently, the ENF is embedded by successively recording the intensity of each frame row. The sequential capturing of rows of each frame enables a considerably faster sensing of the ENF signal as a result of the upscaling of the effective sampling rate by multiplying it with the number of rows of the sensed frame. The process of acquiring video using a camera with a rolling shutter mechanism is illustrated in Figure 4. In every frame period $T_c = \frac{1}{f_c}$, where f_c is the frame rate, every frame's row is successively sampled and captured for T_{ro} seconds, accompanied by T_{idle} seconds of inactivity before moving on to the next frame row [23,48].

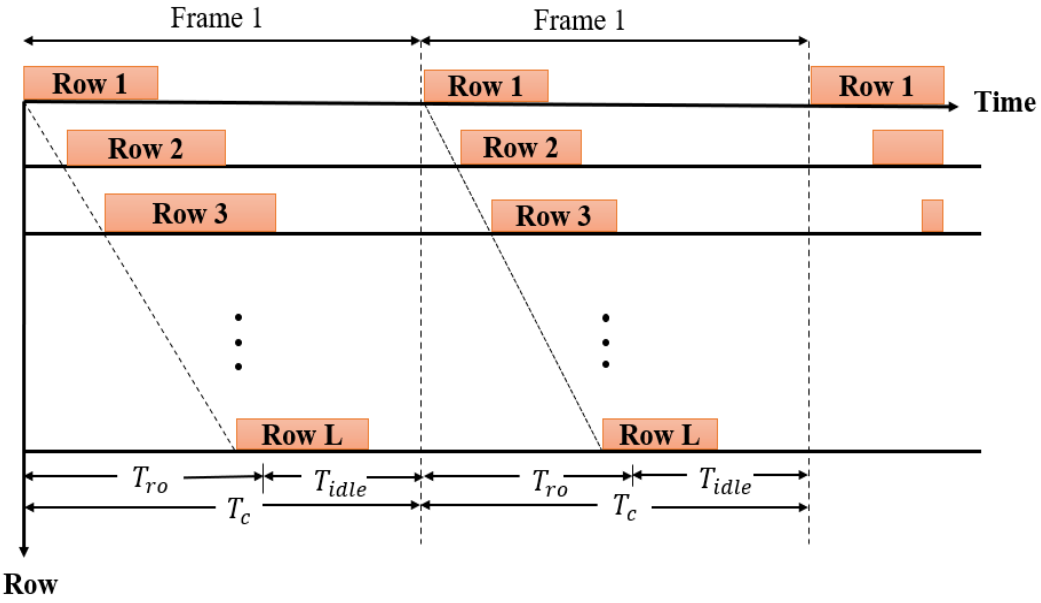

Figure 4. Rolling shutter mechanism video acquisition process [48].

The T_{ro} is referred to as the read-out time of the camera, which is the length of time it takes to obtain the rows of a frame. Every camera has a unique T_{ro} value [21,31]. Furthermore, the work in [31] demonstrates that inside a camera device, the T_{ro} might change based on the resolution or frame rate of a video. Due to the exposure of pixels in different rows at different times, and their simultaneous display during playback, the rolling shutter can generate skew, blur, and other visual distortions, particularly with objects moving rapidly and in quick flashes of light [49]. As a result of the associated distortions, the rolling shutter's sequential read-out process has long been seen as unfavorable to image/video quality. Conversely, recent research has demonstrated that techniques such as computational photography and computer vision can be employed to exploit the rolling shutter mechanism [50,51].

2.4. Camera Read-Out Time (T_{ro}) and ENF Estimation

The T_{ro} is the duration during which a camera captures the rows of a video frame. It is camera-specific and is not often listed in the user manual or specification catalog [46]. The T_{ro} is a sensitive parameter that can be exploited for several applications, including camera forensics. Camera forensics has become an important area of study for a variety of applications, including verifying whether a query image and video pair originated from one source camera and verifying the origin of a questionable image or video [52,53].

The author of [30] proposed a method that can extract a flicker signal and the camera T_{ro} from a pirated video and utilize them to identify the LCD screen and camera used to create the pirated movie. The authors of [21] presented a method that exploits the ENF and the T_{ro} for camera forensics of rolling shutter cameras. They essentially estimate the T_{ro} of the camera for each frame using vertical phase analysis modeled as:

$$T_{ro} = \frac{L\omega b}{2\pi f_e} \tag{6}$$

where L denotes the frame's number of rows, ω_b represents the vertical radial frequency, which is calculated from the slope of the vertical phase line, and f_e represents the ENF component that oscillates around the nominal frequency. The slope of the vertical phase line may be acquired by estimating the ENF phases $\phi[\ell](\ell \in \{1, 2, 3, 4, 5, ..., L\}$ for every one of the rows. $\phi[\ell]$ is extracted by applying the Fourier transform to the time series of the ℓth row, which can be generated by calculating the mean intensity of the lth row of every video frame. The authors of [29] made use of the T_{ro} to compute the number of zeros

that correspond to the idle time for the specific camera [30] or the specific video resolution and frame rate [31] that captured the video, which enabled a better estimation of the ENF signal with no distortion compared to the method described in [25], where the T_{ro} was not exploited.

3. Experiment

We obtained a video dataset recorded with an iPhone 6s back camera under electric lighting in an indoor environment using different frame rates. The videos were static-scene videos recorded in Raleigh, USA, where the ENF nominal value is 60Hz. The power mains signals recorded concurrently with the videos were also acquired and served as the ground-truth ENF signals. We selected seven cameras for the experiments and obtained their T_{ro} values reported in [21], as shown in Table 1. In a real-life scenario, our approach will need a database of the T_{ro} values of all cameras and a database of ENF reference signals. For this study, we used videos whose camera frame heights (L) = 480, frame rates = 23.0062, and T_{ro}s = 19.8 ms. Each T_{ro} value was used with our adapted method [29] to analyze the video. The T_{ro} that corresponds to the camera used to record the video will lead to the estimation of an ENF that best matches the reference signal. In our adapted ENF estimation method, T_{ro} is a sensitive parameter and is employed to compute the number of zeros to be inserted in the idle period before analyzing the video, which is critical in estimating an ENF signal with no distortion. Three performance metrics: the normalized cross-correlation (NCC), the root mean squared error (RMSE), and the mean absolute error (MAE) are used to evaluate the (dis)similarity between the estimated ENF signal $\{v\}_{k=1}^{m}$ and the reference signal $\{r\}_{k=1}^{m}$. The NCC is evaluated as:

$$\frac{\sum_{k=1}^{M} r_k^c v_k^c}{\sqrt{\left(\sum_{k=1}^{M} r_k^{c2} \sum_{k=1}^{M} v_k^{c2}\right.}} \tag{7}$$

where $r_k^c = r_k - \frac{1}{M}\sum_n r_n$, $v_k^c = v_k - \frac{1}{M}\sum_n v_n$, and the variables with overhead bars are respective sample means. The RMSE is evaluated as:

$$\sqrt{(\sum_{k=1}^{M}(r_k - v_k)^2 / M)} \tag{8}$$

and the MAE is evaluated as:

$$\sum_{k=1}^{M} |r_k - v_k| / M \tag{9}$$

Table 1. Cameras and parameters used in our experiment.

Camera ID	Model	L	T_{ro} (ms)
1	iPhone 6s back camera	480	19.8
2	Sony Cybershot DSC-RX 100 II	1080	13.4
3	iPhone 5 front camera	720	22.9
4	iPhone 5 back camera	1080	27.4
5	Sony Handycam HDR-TG1	1080	14.6
6	Canon SX230-HS	240	18.2
7	iPhone 6	1080	30.9

When a test ENF signal $\{v_k\}$ is examined against the reference ENF signal $\{r_k\}$, the (dis)similarity is evaluated using $\{\hat{v}_k\}$ instead of $\{v_k\}$, $\hat{v}_k \stackrel{\text{def}}{=} \hat{\beta}_1 v_k + \hat{\beta}_0$, and $(\hat{\beta}_0, \hat{\beta}_1)$ are the least-square estimates when $\{v_k\}$ is regressed on $\{r_k\}$. This measure will guarantee that the RMSE and MAE metrics can be compared directly to the fluctuations in the reference ENF signal.

4. Results and Discussion

When a video is analyzed using the true T_{ro} value of the camera used to capture it together with the true ENF nominal value where it is captured, the extracted ENF will be a near-to-perfect match with the reference signal, which shows that the video originated from the camera. Figure 5 shows the analysis performed using a T_{ro} value of 19.8 ms, which is the T_{ro} value of the iPhone 6s used to record the video. The extracted ENF provides a better match to the reference signal compared to the analysis results in Figures 6 and 7, where the T_{ro} values of 13.4 ms (Sony Cybershot DSC-RX 100 II) and 30.9 ms (iPhone 6) were used, respectively. The performance of the estimated ENF signals against the reference signal at different T_{ro} values were also evaluated in terms of NCC, RMSE, and MAE, as shown in Table 2. The results further show that the T_{ro} value corresponding to the camera used to capture the video under analysis will lead to the extraction of an ENF signal with the highest correlation and the lowest error rate relative to the reference signal.

Table 2. (Dis)similarity between the extracted ENF and the reference ENF.

Camera Model	T_{ro} (ms)	NCC	RMSE	MAE
iPhone 6s back camera	19.8	0.970	0.0016	0.0012
Sony Cybershot DSC-RX 100 II	13.4	0.940	0.0023	0.0018
iPhone 5 front camera	22.9	0.961	0.0017	0.0013
iPhone 5 back camera	27.4	0.934	0.0023	0.0018
Sony Handycam HDR-TG1	14.6	0.952	0.0020	0.0015
Canon SX230-HS	18.2	0.921	0.0018	0.0014
iPhone 6	30.9	0.944	0.0022	0.0016

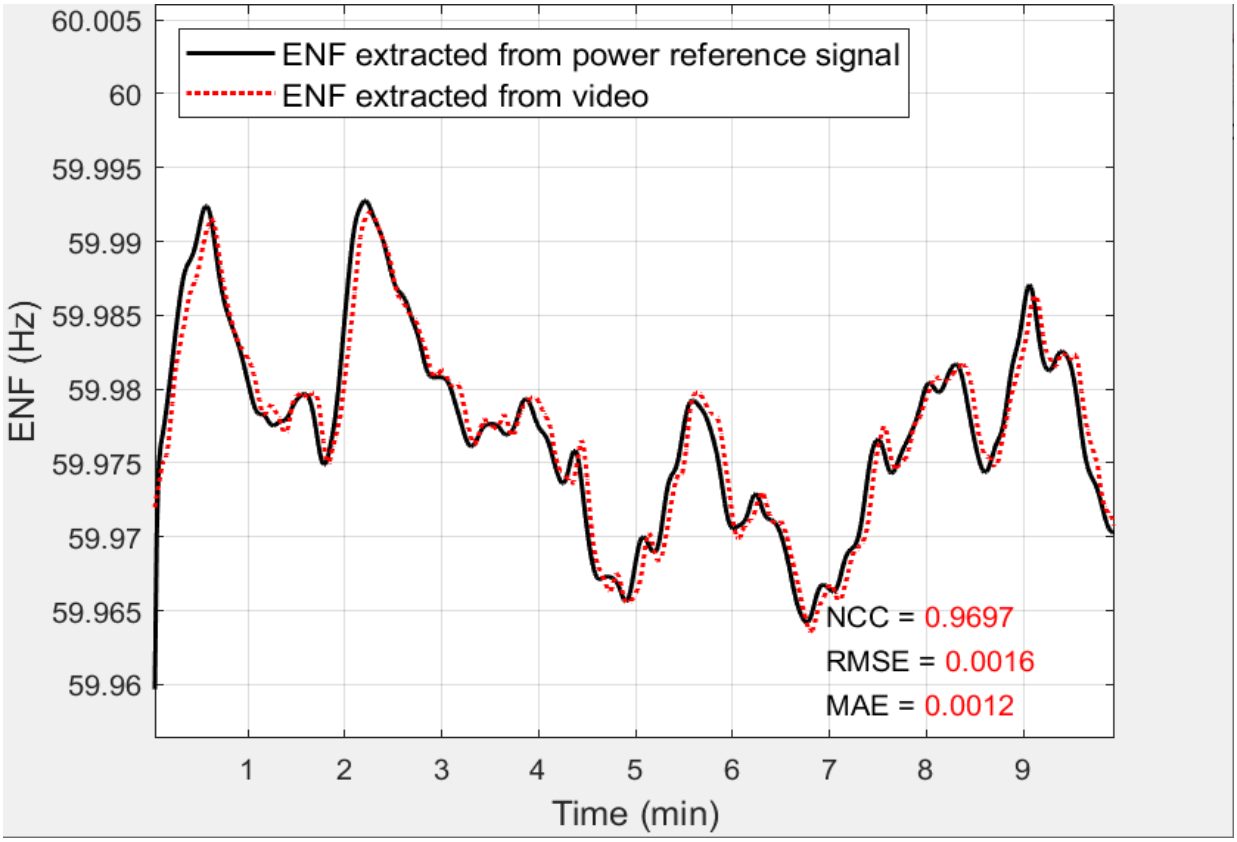

Figure 5. Sample of the ENF signal extracted from a video file (red) using a T_{ro} value of 19.8 ms (iPhone 6s) and matched against the reference signal (black). The bottom right corner shows the measure of similarity and dissimilarity between the extracted signal and the reference signal.

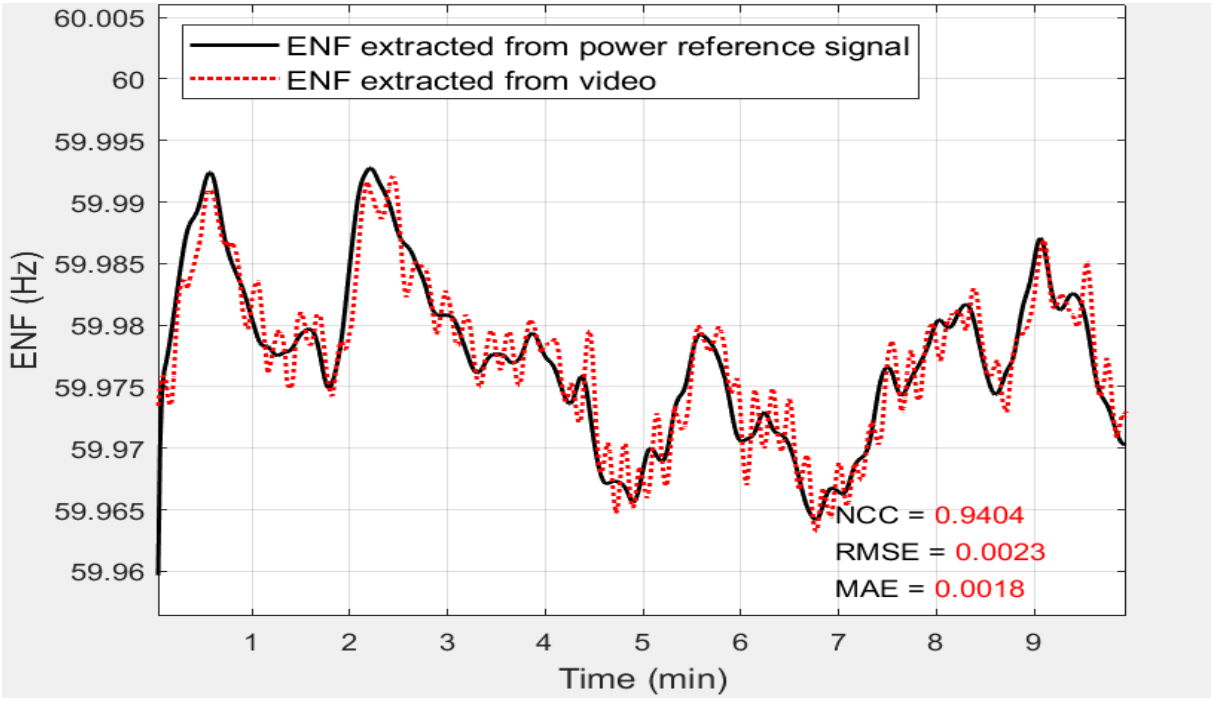

Figure 6. Sample of the ENF signal extracted from a video file (red) using a T_{ro} value of 13.4 ms (Sony Cybershot DSC-RX 100 II) and matched against the reference signal (black). The bottom right corner shows the measure of similarity and dissimilarity between the extracted signal and the reference signal.

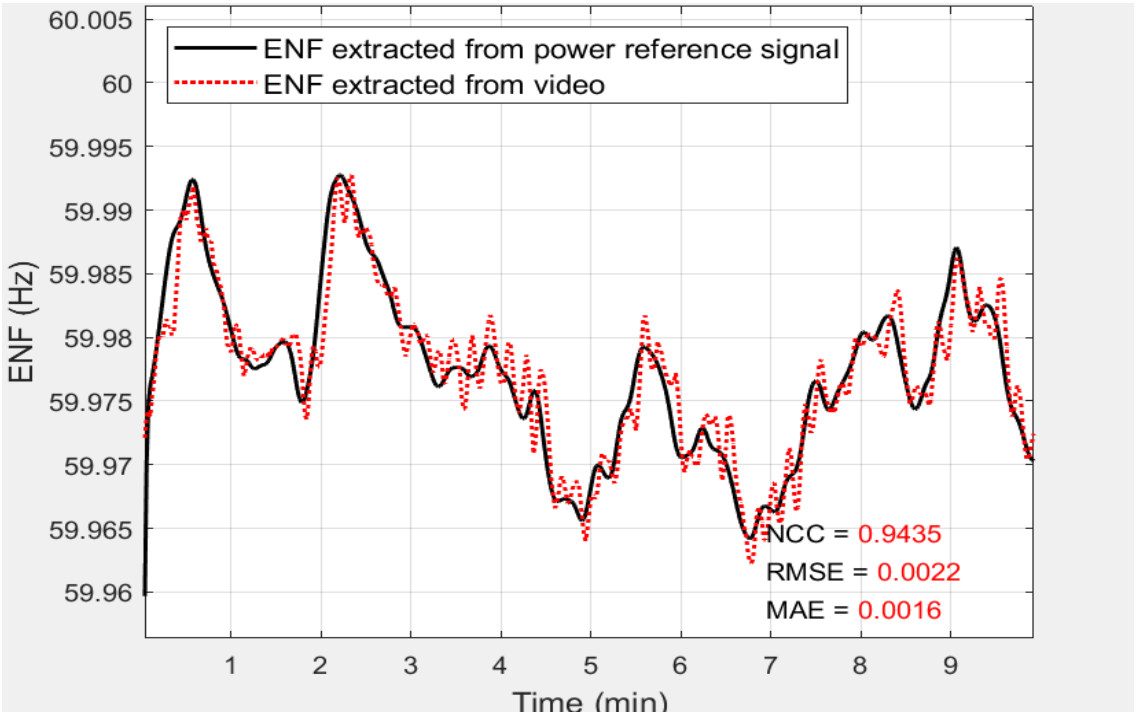

Figure 7. Sample of the ENF signal extracted from a video file (red) using a T_{ro} value of 30.9 ms (iPhone 6) and matched against the reference signal (black). The bottom right corner shows the measure of similarity and dissimilarity between the extracted signal and the reference signal.

Figure 8 shows a plot of the correlations of the ENF signals extracted using different T_{ro} values. It can be observed that the T_{ro} of the iPhone 6s (19.8 ms) which was used to capture the video used in the analysis produced an ENF signal with the highest correlation.

It can also be seen in Figures 9 and 10 that the signal extracted using the correct T_{ro} (the T_{ro} of the camera that captured the video) had the lowest error rate.

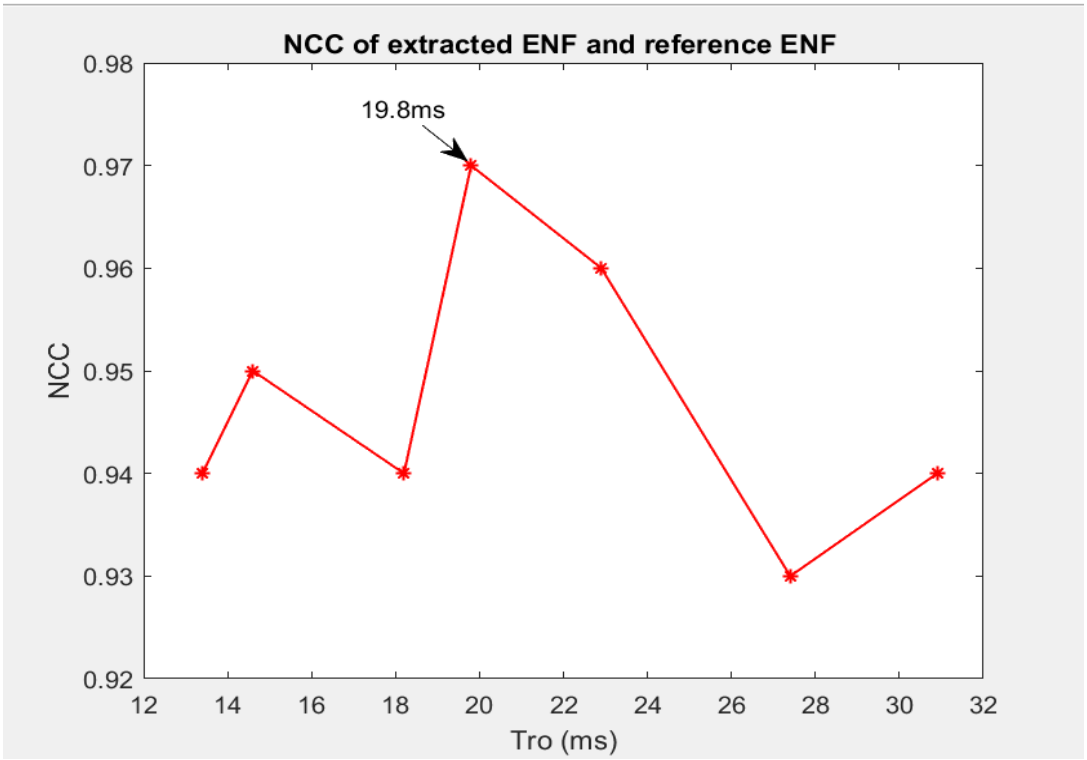

Figure 8. NCC of the extracted ENF and the reference ENF.

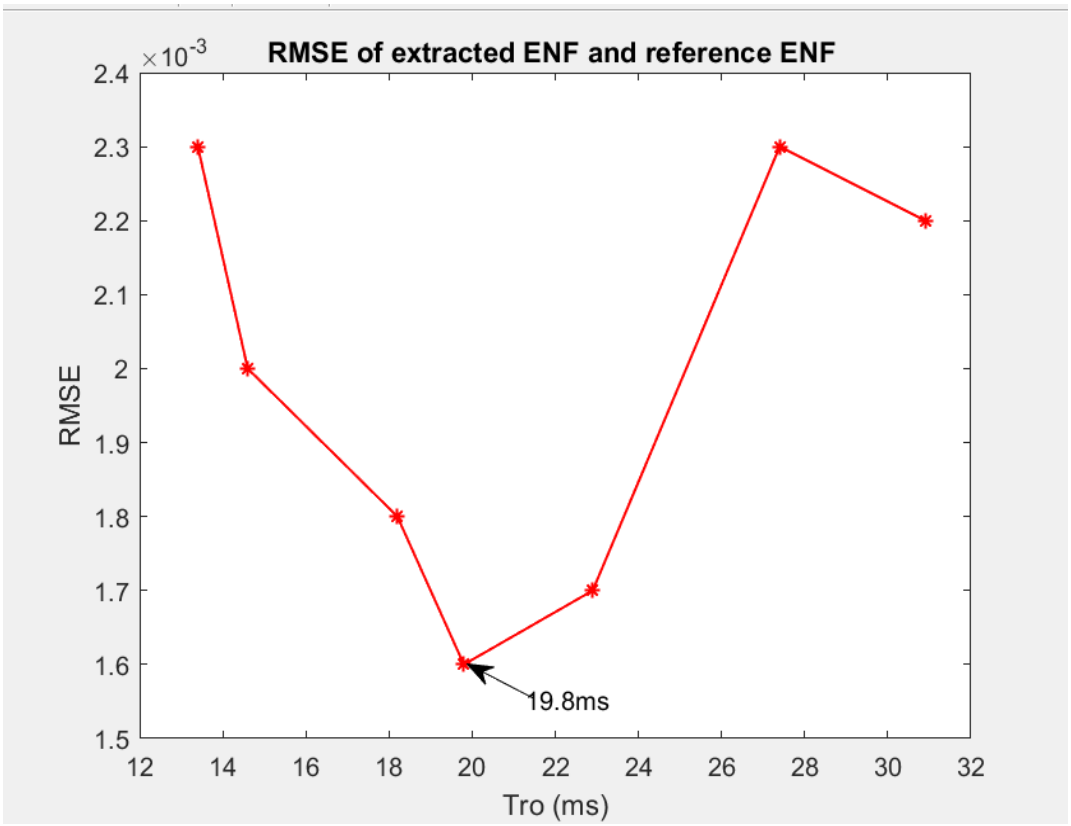

Figure 9. RMSE of the extracted ENF and the reference ENF.

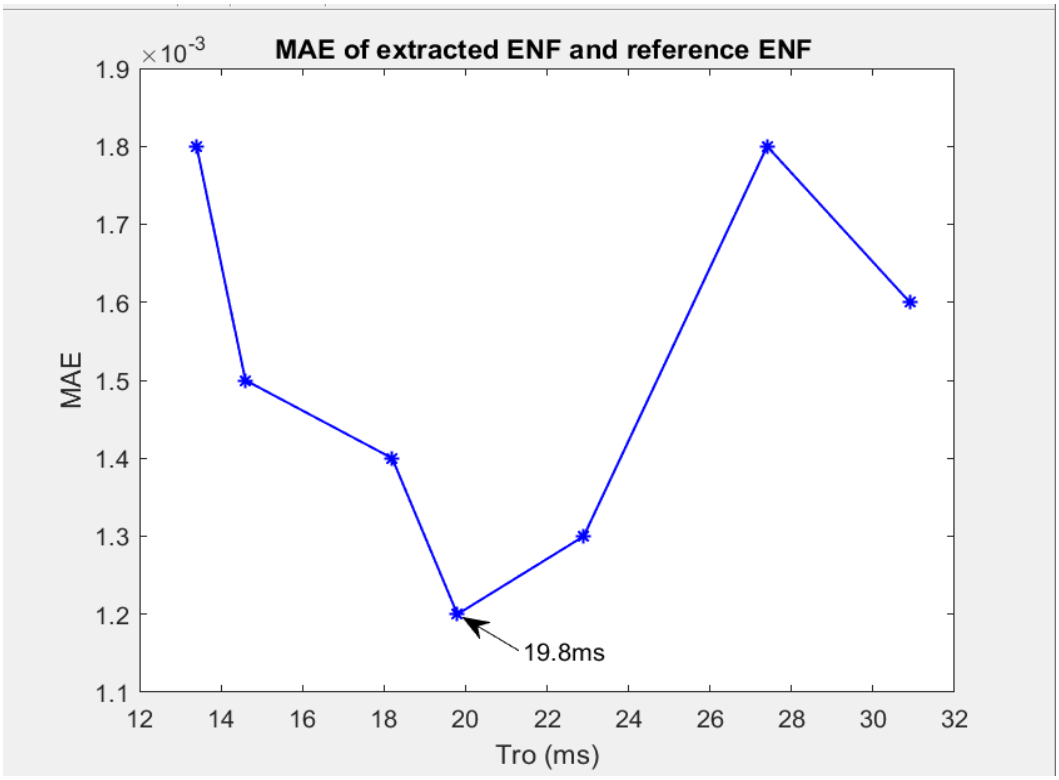

Figure 10. MAE of the extracted ENF and the reference ENF.

Our results show that if a T_{ro} database of all cameras and the database of ENF reference signals are obtained, an ENF-containing video of unknown source can be analyzed using our approach to trace the source camera.

5. Conclusions

In this study, we have presented an approach that exploits the T_{ro} and the ENF to trace an ENF-containing video to its source camera. We adapted an ENF estimation method in which the T_{ro} is a sensitive parameter and is employed to compute the number of zeros to be inserted in the idle period for the specific camera that captured a video or the specific resolution and frame rate of the video before analyzing it, which is critical in estimating an ENF signal with no distortion. The T_{ro} value that corresponds to the camera used to record the video will lead to the estimation of an ENF that best matches the reference signal. In essence, our approach is based on the idea that, given an ENF-containing video from an unknown source camera, we can apply different T_{ro} values to analyze the video and the T_{ro} value (camera) that leads to the extraction of an ENF signal with the highest correlation and lowest error margin when compared with the reference ENF can be said to have produced the video. We performed experiments using the T_{ro} values of seven cameras and the results validate our idea. Our approach could prove very useful in a practical scenario where a video obtained from the public, including social media, is being investigated by law enforcement to ascertain if it originated from a suspect's camera. The limitation of our proposed approach is the computation cost required to calculate the T_{ro} values for the ENF extraction process and the matching of the extracted ENF against a large ENF reference database. Our approach can be further studied using several videos and more camera T_{ro} values to examine its consistency in achieving useful performance.

Author Contributions: Conceptualization, E.N., L.-M.A., K.P.S. and M.W.; Methodology, E.N. and L.-M.A.; Software, E.N.; Writing—original draft, E.N.; Writing—review & editing, E.N, L.-M.A., K.P.S. and M.W.; Supervision, L.-M.A., K.P.S. and M.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Grigoras, C. Applications of ENF criterion in forensic audio, video, computer and telecommunication analysis. *Forensic Sci. Int.* **2007**, *167*, 136–145. [CrossRef] [PubMed]
2. Jeon, Y.; Kim, M.; Kim, H.; Kim, H.; Huh, J.H.; Yoon, J.W. I'm Listening to your Location! Inferring User Location with Acoustic Side Channels. In Proceedings of the 2018 World Wide Web Conference, Geneva, Switzerland, 23–27 April 2018; pp. 339–348.
3. Rodríguez, D.P.N.; Apolinário, J.A.; Biscainho, L.W.P. Audio authenticity: Detecting ENF discontinuity with high precision phase analysis. *IEEE Trans. Inf. Forensics Secur.* **2010**, *5*, 534–543. [CrossRef]
4. Lin, X.; Kang, X. Supervised audio tampering detection using an autoregressive model. In Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 5–9 March 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 2142–2146.
5. Garg, R.; Varna, A.L.; Hajj-Ahmad, A.; Wu, M. "Seeing" ENF: Power-signature-based timestamp for digital multimedia via optical sensing and signal processing. *IEEE Trans. Inf. Forensics Secur.* **2013**, *8*, 1417–1432. [CrossRef]
6. Sanders, R.W. Digital audio authenticity using the electric network frequency. In Proceedings of the Audio Engineering Society Conference: 33rd International Conference: Audio Forensics-Theory and Practice. Audio Engineering Society, Denvor, CO, USA, 5–7 June 2008.
7. Grigoras, C. Digital audio recording analysis–the electric network frequency criterion. *Int. J. Speech Lang. Law* **2005**, *12*, 63–76. [CrossRef]
8. Fechner, N.; Kirchner, M. The humming hum: Background noise as a carrier of ENF artifacts in mobile device audio recordings. In Proceedings of the 2014 Eighth International Conference on IT Security Incident Management & IT Forensics, Münster, Germany, 12–14 May 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 3–13.
9. Bykhovsky, D.; Cohen, A. Electrical network frequency (ENF) maximum-likelihood estimation via a multitone harmonic model. *IEEE Trans. Inf. Forensics Secur.* **2013**, *8*, 744–753. [CrossRef]
10. Hua, G.; Zhang, Y.; Goh, J.; Thing, V.L. Audio authentication by exploring the absolute-error-map of ENF signals. *IEEE Trans. Inf. Forensics Secur.* **2016**, *11*, 1003–1016. [CrossRef]
11. Savari, M.; Wahab, A.W.A.; Anuar, N.B. High-performance combination method of electric network frequency and phase for audio forgery detection in battery-powered devices. *Forensic Sci. Int.* **2016**, *266*, 427–439. [CrossRef] [PubMed]
12. Yao, W.; Zhao, J.; Till, M.J.; You, S.; Liu, Y.; Cui, Y.; Liu, Y. Source location identification of distribution-level electric network frequency signals at multiple geographic scales. *IEEE Access* **2017**, *5*, 11166–11175. [CrossRef]
13. Narkhede, M.; Patole, R. Acoustic scene identification for audio authentication. In *Soft Computing and Signal Processing*; Wang, J., Reddy, G., Prasad, V., Reddy, V., Eds.; Springer: Singapore, 2019; pp. 593–602.
14. Zhao, H.; Malik, H. Audio recording location identification using acoustic environment signature. *IEEE Trans. Inf. Forensics Secur.* **2013**, *8*, 1746–1759. [CrossRef]
15. Ohib, R.; Arnob, S.Y.; Arefin, R.; Amin, M.; Reza, T. ENF Based Grid Classification System: Identifying the Region of Origin of Digital Recordings. *Criterion* **2017**, *3*, 5.
16. Despotović, D.; Knežević, M.; Šarić, Ž.; Zrnić, T.; Žunić, A.; Delić, T. Exploring Power Signatures for Location Forensics of Media Recordings. In Proceedings of the IEEE Signal Processing Cup, Shanghai, China, 20–25 March 2016.
17. Sarkar, M.; Chowdhury, D.; Shahnaz, C.; Fattah, S.A. Application of electrical network frequency of digital recordings for location-stamp verification. *Appl. Sci.* **2019**, *9*, 3135. [CrossRef]
18. El Helou, M.; Turkmani, A.W.; Chanouha, R.; Charbaji, S. ANovel ENF Extraction Approach for Region-of-Recording Verification of Media Recordings. *Forensic Sci. Int.* **2005**, *155*, 165.
19. Zhou, H.; Duanmu, H.; Li, J.; Ma, Y.; Shi, J.; Tan, Z.; Wang, X.; Xiang, L.; Yin, H.; Li, W. Geographic Location Estimation from ENF Signals with High Accuracy. In Proceedings of the IEEE Signal Processing Cup, Shanghai, China, 20–25 March 2016; pp. 1–8.
20. Hajj-Ahmad, A.; Garg, R.; Wu, M. ENF-based region-of-recording identification for media signals. *IEEE Trans. Inf. Forensics Secur.* **2015**, *10*, 1125–1136. [CrossRef]
21. Hajj-Ahmad, A.; Berkovich, A.; Wu, M. Exploiting power signatures for camera forensics. *IEEE Signal Process. Lett.* **2016**, *23*, 713–717. [CrossRef]
22. Su, H.; Hajj-Ahmad, A.; Wu, M.; Oard, D.W. Exploring the use of ENF for multimedia synchronization. In Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy, 4–9 May 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 4613–4617.

23. Su, H.; Hajj-Ahmad, A.; Wong, C.W.; Garg, R.; Wu, M. ENF signal induced by power grid: A new modality for video synchronization. In Proceedings of the 2nd ACM International Workshop on Immersive Media Experiences, Orlando, FL, USA, 7 November 2014; pp. 13–18.
24. Vatansever, S.; Dirik, A.E.; Memon, N. Detecting the presence of ENF signal in digital videos: A superpixel-based approach. *IEEE Signal Process. Lett.* **2017**, *24*, 1463–1467. [CrossRef]
25. Su, H.; Hajj-Ahmad, A.; Garg, R.; Wu, M. Exploiting rolling shutter for ENF signal extraction from video. In Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP), Paris, France, 27–30 October 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 5367–5371.
26. Karantaidis, G.; Kotropoulos, C. An Automated Approach for Electric Network Frequency Estimation in Static and Non-Static Digital Video Recordings. *J. Imaging* **2021**, *7*, 202. [CrossRef] [PubMed]
27. Fernández-Menduiña, S.; Pérez-González, F. Temporal localization of non-static digital videos using the electrical network frequency. *IEEE Signal Process. Lett.* **2020**, *27*, 745–749. [CrossRef]
28. Ferrara, P.; Sanchez, I.; Draper-Gil, G.; Junklewitz, H.; Beslay, L. A MUSIC Spectrum Combining Approach for ENF-based Video Timestamping. In Proceedings of the 2021 IEEE International Workshop on Biometrics and Forensics (IWBF), Rome, Italy, 6–7 May 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 1–6.
29. Choi, J.; Wong, C.W. ENF signal extraction for rolling-shutter videos using periodic zero-padding. In Proceedings of the 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 2667–2671.
30. Hajj-Ahmad, A.; Baudry, S.; Chupeau, B.; Doërr, G. Flicker forensics for pirate device identification. In Proceedings of the 3rd ACM Workshop on Information Hiding and Multimedia Security, Portland, OR, USA, 17–19 June 2015; pp. 75–84.
31. Vatansever, S.; Dirik, A.E.; Memon, N. Analysis of rolling shutter effect on ENF-based video forensics. *IEEE Trans. Inf. Forensics Secur.* **2019**, *14*, 2262–2275. [CrossRef]
32. Gemayel, T.E.; Bouchard, M. A Parametric Autoregressive Model for the Extraction of Electric Network Frequency Fluctuations in Audio Forensic Authentication. *J. Energy Power Eng.* **2016**, *10*, 504–512. [CrossRef]
33. Bollen, M.H.; Gu, I.Y. *Signal Processing of Power Quality Disturbances*; Mohamed, E.E., Ed.; John Wiley & Sons: New York, NY, USA, 2016.
34. Haykin, S. *Advances in Spectrum Analysis and Array Processing*, 3rd ed.; Pentice-Hall, Inc.: Hoboken, NJ, USA, 1995.
35. Schmidt, R. Multiple emitter location and signal parameter estimation. *IEEE Trans. Antennas Propag.* **1986**, *34*, 276–280. [CrossRef]
36. Roy, R.; Kailath, T. ESPRIT-estimation of signal parameters via rotational invariance techniques. *IEEE Trans. Acoust. Speech Signal Process.* **1989**, *37*, 984–995. [CrossRef]
37. Smith, J.O.; Serra, X. PARSHL: An analysis/synthesis program for non-harmonic sounds based on a sinusoidal representation. In Proceedings of the 1987 International Computer Music Conference, ICMC, Champaign/Urbana, IL, USA, 23–26 August 1987; pp. 290–297.
38. Dosiek, L. Extracting electrical network frequency from digital recordings using frequency demodulation. *IEEE Signal Process. Lett.* **2014**, *22*, 691–695. [CrossRef]
39. Hua, G.; Zhang, H. ENF signal enhancement in audio recordings. *IEEE Trans. Inf. Forensics Secur.* **2019**, *15*, 1868–1878. [CrossRef]
40. Vidyamol, K.; George, E.; Jo, J.P. Exploring electric network frequency for joint audio-visual synchronization and multimedia authentication. In Proceedings of the 2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), Kerala, India, 6–7 July 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 240–246.
41. Achanta, R.; Shaji, A.; Smith, K.; Lucchi, A.; Fua, P.; Süsstrunk, S. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 2274–2282. [CrossRef] [PubMed]
42. Stutz, D.; Hermans, A.; Leibe, B. Superpixels: An evaluation of the state-of-the-art. *Comput. Vis. Image Underst.* **2018**, *166*, 1–27. [CrossRef]
43. Fernández-Menduiña, S.; Pérez-González, F. ENF Moving Video Database. *Zenodo* **2020**. [CrossRef]
44. Barnich, O.; Van Droogenbroeck, M. ViBe: A universal background subtraction algorithm for video sequences. *IEEE Trans. Image Process.* **2010**, *20*, 1709–1724. [CrossRef]
45. Lindsey, W.C.; Chie, C.M. A survey of digital phase-locked loops. *Proc. IEEE* **1981**, *69*, 410–431. [CrossRef]
46. Hajj-Ahmad, A.; Wong, C.W.; Choi, J.; Wu, M. Power Signature for Multimedia Forensics. In *Multimedia Forensics. Advances in Computer Vision and Pattern Recognition*; Sencar, H.T., Verdoliva, L., Memon, N., Eds.; Springer: Singapore, 2022; pp. 235–280.
47. Han, H.; Jeon, Y.; Song, B.K.; Yoon, J.W. A phase-based approach for ENF signal extraction from rolling shutter videos. *IEEE Signal Process. Lett.* **2022**, *29*, 1724–1728. [CrossRef]
48. Choi, J.; Wong, C.W.; Su, H.; Wu, M. Analysis of ENF Signal Extraction from Videos Acquired by Rolling Shutters. 2022. Available online: https://www.techrxiv.org/articles/preprint/Analysis_of_ENF_Signal_Extraction_From_Videos_Acquired_by_Rolling_Shutters/21300960 (accessed on 3 December 2022).
49. Liang, C.K.; Chang, L.W.; Chen, H.H. Analysis and compensation of rolling shutter effect. *IEEE Trans. Image Process.* **2008**, *17*, 1323–1330. [CrossRef]
50. Ait-Aider, O.; Bartoli, A.; Andreff, N. Kinematics from lines in a single rolling shutter image. In Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 17–22 June 2007; IEEE: Piscataway, NJ, USA, 2007; pp. 1–6.

51. Gu, J.; Hitomi, Y.; Mitsunaga, T.; Nayar, S. Coded rolling shutter photography: Flexible space-time sampling. In Proceedings of the 2010 IEEE International Conference on Computational Photography (ICCP), Cambridge, MA, USA, 29–30 March 2010; IEEE: Piscataway, NJ, USA, 2010; pp. 1–8.
52. Sencar, H.T.; Memon, N. Digital image forensics. In *Counter-Forensics: Attacking Image Forensics*; Springer: New York, NY, USA, 2013; pp. 327–366.
53. Shullani, D.; Fontani, M.; Iuliani, M.; Shaya, O.A.; Piva, A. VISION: A video and image dataset for source identification. *EURASIP J. Inf. Secur.* **2017**, *1*, 1–16. [CrossRef]