Analysis of Rolling Shutter Effect on ENF-Based Video Forensics

Saffet Vatansever, Ahmet Emir Dirik, and Nasir Memon, *Fellow, IEEE*

Abstract—Electric network frequency (ENF) is a time-varying signal of the frequency of mains electricity in a power grid. It continuously fluctuates around a nominal value (50/60 Hz) due to changes in the supply and demand of power over time. Depending on these ENF variations, the luminous intensity of a mains-powered light source also fluctuates. These fluctuations in luminance can be captured by video recordings. Accordingly, the ENF can be estimated from such videos by the analysis of steady content in the video scene. When videos are captured by using a rolling shutter sampling mechanism, as is done mostly with CMOS cameras, there is an idle period between successive frames. Consequently, a number of illumination samples of the scene are effectively lost due to the idle period. These missing samples affect the ENF estimation, in the sense of the frequency shift caused and the power attenuation that results. This paper develops an analytical model for videos captured using a rolling shutter mechanism. This model illustrates how the frequency of the main ENF harmonic varies depending on the idle period length, and how the power of the captured ENF attenuates as idle period increases. Based on this, a novel idle period estimation method for potential use in camera forensics that is able to operate independently of video frame rate is proposed. Finally, a novel time-of-recording verification approach based on the use of multiple ENF components, idle period assumptions, and the interpolation of missing ENF samples is also proposed.

Index Terms—Electric network frequency (ENF), video forensics, multimedia forensics, camera forensics, rolling shutter, idle period, camera verification, time-of-recording, time-stamp.

I. INTRODUCTION

ENF (Electric Network Frequency) oscillates instantaneously between an upper and lower bound around a nominal value (50/60 Hz) owing to a continuous imbalance between generated power and consumed power [1]. In an interconnected mains power network, the temporal variation of ENF is expected to be the same across the whole region that the network spans [2].

ENF is inherently embedded in audio recordings due to the electromagnetic field or the acoustic mains hum

in the environment, and can be extracted from audio with time-domain or frequency-domain techniques [2]–[5]. In recent years, it was discovered that ENF is also embedded in video recordings of a scene illuminated by a mains-powered light source [6]. The illumination intensity of a mains-powered light source varies depending on ENF fluctuations in the power network and these variations in luminance can be measured from video recordings. Specifically, ENF can be extracted from videos by doing a careful analysis of subtle illumination alterations in steady content through consecutive video frames [6]–[10].

Since the ENF signal measured from any point in the power network can be used as a reference signal, as well as the fact that it can be extracted from digital media files has led the use of ENF in media forensics. Specifically, it can be exploited for a number of forensic and anti-forensic applications including time-of-recording verification [5], [7], [11], media authentication [12], [13], multimedia synchronization [14]–[16], power grid identification [17] and camera read-out time estimation [18].

Most consumer cameras equipped with CMOS sensors use a rolling shutter sampling mechanism to capture a video frame whereas those with CCD sensors employ a global shutter sampling mechanism. When using a global shutter mechanism, each pixel, so each row of a frame is exposed at the same time instance. Whereas with a rolling shutter, each row is captured at different time instances. This difference in the sampling mechanisms has led researchers to develop different approaches for ENF signal estimation from videos. The first approach [6] is aimed at the global shutter mechanism. It is based on averaging all the steady pixels in each frame of the video, i.e., one illuminations sample per frame. Consequently, it is unable to satisfy Nyquist criteria [19] and so is dependent on the alias ENF caused by the low sampling frequency as determined by the *video frame rate*. Another ENF estimation approach [10] is based on averaging steady super-pixels rather than all steady pixels within the frame, which leads the technique to be applicable to videos exposed in different sampling mechanisms. Nevertheless, since one illumination sample only is obtained per frame, this technique also relies on alias ENF. The other approach is designed for the rolling shutter mechanism and is based on averaging steady pixels in each row, i.e., one illumination sample per row [8], [9]. This method uses the advantage of the increased ENF sampling frequency of the rolling shutter mechanism, as determined by the *video frame rate × number of rows*. However, the rolling shutter approach brings with it the issue of idle time between two successive frames where there is no sampling done. Accordingly, some illumination samples are lost at the end

of each frame. Su et al. provide a fundamental understanding on how the main ENF harmonic is shifted to some other frequencies due to the idle period in videos captured using the rolling shutter mechanism [8]. They use a filter bank model formulation of the mechanism for their analysis.

In this work, the phenomenon introduced in [8] is further studied and an analytical model is derived to explore how the frequency of mains-powered illumination, and so ENF is shifted and is attenuated in relation to idle period length. Among the new ENF components caused by a particular idle period, the two with the most power are explored. Model predictions are verified by simulation results. Next, based on the model developed, a novel idle period estimation approach is proposed. To the best of our knowledge, the only work utilizing ENF for camera forensics is that of Hajj-Ahmad et al. [18]. They mainly rely on computing the time needed to read one frame based on vertical phase analysis [21], i.e., estimation of ENF phase for each row. For this purpose, the sequence of the mean luminance values of the ith rows over each frame is treated as a separate time series, leading to a sampling rate that equals the *video frame rate*. This does not satisfy the Nyquist Criteria [19]. Hence this approach depends on alias ENF and it is a challenge to apply this technique to videos where the frame rate is a divisor of nominal ENF, e.g., a frame rate of 25 fps with nominal ENF at 50 Hz in EU or of 30 fps at 60 Hz nominal ENF in the US. This is because alias ENF is obtained at 0 Hz DC component in such a condition. It may also be a challenge for the vertical phase analysis approach to operate in a noisy video, where the power of the captured ENF is weak for a reliable phase estimation.

Our model and analysis also leads to a novel time-of-recording verification technique that performs better than the current techniques [7]–[9]. A systematic search of possible ENF components that emerge as a result of the idle period, followed by idle period assumptions in each component and interpolation of missing samples for each assumption result in better quality ENF signal estimations. A better quality of ENF signal consequently leads to a better time-of-recording verification performance. In summary, the main contributions of this work include: (1) a model for where frequency of main ENF harmonic is shifted and how the power of the captured ENF is attenuated, depending on idle period length; (2) a novel idle period estimation technique targeted at camera forensics; (3) a novel time-of-recording verification method for videos captured using a rolling shutter mechanism.

II. THE IMPACT OF ROLLING SHUTTER ON ENF

In the global shutter mechanism, widely used in CCD sensors, an entire frame is exposed at one time instance, i.e., each row of pixels of a single frame is sampled simultaneously. Whereas in rolling shutter mechanism, mostly used in CMOS sensors, each row of pixels in a frame is captured sequentially at different time instances. Fig. 1 illustrates the timing for the two distinct procedures for a single frame [20]. The state-of-the-art ENF estimation techniques in the literature are basically designed by taking these sampling mechanisms into
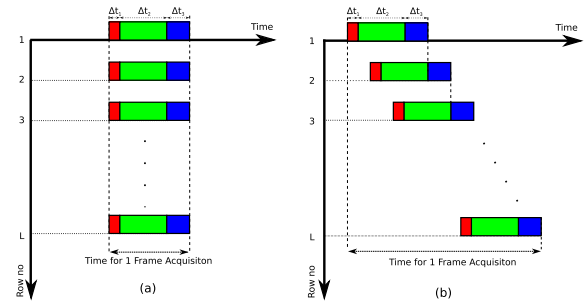

Fig. 1. Demonstration of (a) global shutter sampling mechanism - each row of a frame is exposed at the same time instance (b) rolling shutter sampling mechanism - each row of a frame is exposed at a different time instance [20]. Δt_1, Δt_2 and Δt_3 respectively denote reset, exposure, and readout time periods for one row.

consideration. The approach in [6] mainly relies on the global shutter phenomenon. In this method, one illumination sample is obtained per frame by averaging all steady pixels within the frame. Hence, the sampling frequency in this technique is essentially the *video frame rate*, which is much lower than twice the nominal ENF frequency. As this does not satisfy the Nyquist criteria [19], it has to work with alias ENF. However when the video frame rate is a divisor of the nominal ENF, alias ENF is observed at the 0 Hz DC component. So, it is a great challenge for this scheme to work under this condition. The super-pixel based ENF estimation technique in [10] is more generic and is able to operate in different sampling mechanisms. However, as in [6], this technique also depends on alias ENF since one illumination sample only is obtained per frame by steady super-pixels average within the frame.

The method in [8] and [9] is based on the rolling shutter phenomenon. It leads to a sampling frequency as high as *number of rows × camera frame rate*, as each row is treated as a different illumination sample. Consequently, since the number of samples is much higher than twice the ENF frequency (Nyquist rate), alias ENF is not a phenomenon for this scheme.

A. Idle Period Effect: Variation in Frequency of Captured ENF Component

Although the rolling shutter mechanism can lead to an increase of sampling frequency and consequently avoids alias frequency, it brings with it the idle period issue. That is, some illumination samples in each frame are missed [8]. A representation of this phenomenon is depicted in Fig. 2. Here, M represents the number of samples, i.e., the number of rows that the rolling shutter mechanism is capable of capturing during one frame period. That is, the sampling mechanism has a capacity of sampling M rows over a frame period assuming no sample is lost, i.e., if the idle time is zero seconds. $M - L$ denotes the number of samples that are discarded during the idle period, so L is the number of remaining samples, i.e., the actual number of rows per frame ($L \leq M$). $x[n]$ is the time-series of the illumination samples, if there is no idle period. $y[n]$ is the output time-series of the remaining
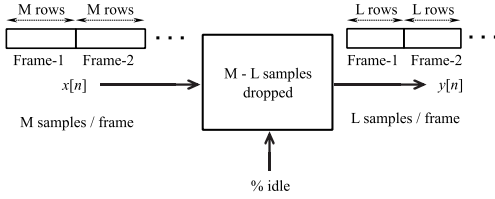
Fig. 2. Inherent reduction in captured luminance samples in a video due to the implementation of idle period at the end of each frame by camera: M is the number of luminance samples, i.e., one sample per row, that the rolling shutter mechanism is able to capture during one frame period presuming no idle period exists. $M - L$ is the number of samples discarded due to the idle period, and hence L is the actual number of samples, and so rows, per frame. $x[n]$ is the time-series of the illumination samples that can be obtained over the video duration, if there is no idle period. $y[n]$ is the actual time-series of the illumination samples as a result of abandoning the samples in the idle period of each frame.
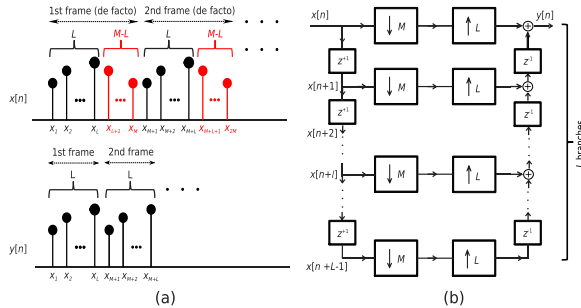

Fig. 3. Illustration of rolling shutter sampling mechanism [8] (a) in time domain; (b) based on poly-phase decomposition.
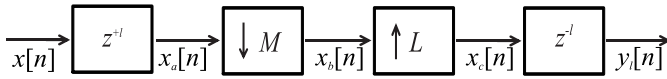

Fig. 4. The stages at the l_{th} branch of the poly-phase model in Fig. 3 (b).

illumination samples after dropping the samples in the idle period of each frame.

A time-domain representation of the above procedure is depicted in Fig. 3 (a) [8], and a realization of poly-phase decomposition [22] of this representation is illustrated in Fig. 3 (b) [8]. In this model, for simplicity, the anti-aliasing filter is omitted considering the narrow band nature of ENF. In each branch of the model, the input signal is shifted back in time, Eq. 1, followed by an M-fold down-sampling filter, Eq. 2. Next, an L-fold up-sampling filter is applied, Eq. 3. Then the signal is shifted forward in the appropriate order, Eq. 4. For the l_{th} branch, these stages are demonstrated in Fig. 4. Accordingly, DTFT (Discrete Time Fourier Transform) of $x_a[n]$, $x_b[n]$, $x_c[n]$ and $y_l[n]$ are respectively obtained as follows:

$$X_a\left(e^{j\omega}\right) = X\left(e^{j\omega}\right) e^{j\omega l} \tag{1}$$

$$X_b\left(e^{j\omega}\right) = \frac{1}{M} \sum_{m=0}^{M-1} X_a\left(e^{j\left(\frac{\omega}{M} - \frac{2\pi m}{M}\right)}\right)$$

$$= \frac{1}{M} \sum_{m=0}^{M-1} \left[X\left(e^{j\left(\frac{\omega}{M} - \frac{2\pi m}{M}\right)}\right) e^{j\left(\frac{\omega}{M} - \frac{2\pi m}{M}\right)l} \right] \tag{2}$$

$$X_c\left(e^{j\omega}\right) = X_b\left(e^{j\omega L}\right)$$

$$= \frac{1}{M} \sum_{m=0}^{M-1} \left[X\left(e^{j\left(\frac{\omega L}{M} - \frac{2\pi m}{M}\right)}\right) e^{j\left(\frac{\omega L}{M} - \frac{2\pi m}{M}\right)l} \right] \tag{3}$$

$$Y_l\left(e^{j\omega}\right) = X_c\left(e^{j\omega}\right) e^{-j\omega l} \tag{4}$$

Then, DTFT of the signal at the end of the lth branch, $Y_l\left(e^{j\omega}\right)$ can be expressed as follows:

$$Y_l\left(\omega\right) = \frac{1}{M} \left(\sum_{m=0}^{M-1} X\left(\frac{\omega L - 2\pi m}{M}\right) e^{j\frac{\omega L - 2\pi m}{M}l} \right) e^{-j\omega l} \tag{5}$$

It should be noted in Eq. 5 that $Y_l\left(e^{j\omega}\right)$ is shown in the form $Y_l\left(\omega\right)$ for simplicity. From this point on, $Y\left(e^{j\omega}\right)$, $X\left(e^{j\omega}\right)$ and other relevant frequency domain variables are shown in similar format. Consequently, when all the signals at the end of each branch of the poly-phase model are combined together, the output signal $Y\left(\omega\right)$ is formed as follows [8]:

$$Y\left(\omega\right) = \sum_{l=0}^{L-1} Y_l\left(\omega\right)$$

$$= \sum_{l=0}^{L-1} \frac{1}{M} \left(\sum_{m=0}^{M-1} X\left(\frac{\omega L - 2\pi m}{M}\right) e^{j\frac{\omega L - 2\pi m}{M}l} \right) e^{-j\omega l}$$

$$= \sum_{m=0}^{M-1} X\left(\frac{\omega L - 2\pi m}{M}\right) F_m\left(\omega\right) \tag{6}$$

where

$$F_m\left(\omega\right) = \frac{1}{M} \sum_{l=0}^{L-1} e^{-j\frac{\omega(M-L)+2\pi m}{M}l} \tag{7}$$

B. Proposed Model

In Eq. 6, F_m specifies the amount of attenuation in the ENF signal depending on the proportion of L to M. Accordingly, by disregarding F_m, the ω_y value making $\left| X\left(\frac{\omega_y L - 2\pi m}{M}\right) \right|$ the same as $\left| X\left(\omega_0\right) \right|$ is the shifted angular frequency for a specific idle period, where ω_0 is nominal angular electric frequency. To find ω_y value, the following cases should be analyzed:

Case 1:

$$\left| X\left(\omega_0\right) \right| = \left| X\left(\frac{\omega_y L - 2\pi m}{M}\right) \right| \tag{8}$$

Accordingly;

$$\omega_0 = \frac{\omega_y L - 2\pi m}{M} \tag{9}$$

ω_y and ω_0 can be expressed as follows:

$$\omega_y = \frac{2\pi f_y}{F_r L} \quad \text{and} \quad \omega_0 = \frac{2\pi f_0}{F_r M} \tag{10}$$

where f_0 is nominal illumination frequency (twice the ENF), f_y is the emerging shifted illumination frequency, and F_r is

video frame rate. By substituting ω and ω_0 equivalents, Eq. 9 can be rewritten as follows:

$$\frac{2\pi f_0}{F_r M} = \frac{2\pi f_y}{F_r L}\frac{L}{M} - \frac{2\pi m}{M}\frac{F_r}{F_r} \tag{11}$$

From the Eq. 11, for **Case 1**, f_y can be obtained as:

$$f_y = f_0 + m F_r \tag{12}$$

where, $f_y < \frac{F_r L}{2}$, which comes from the Nyquist theorem considering $F_r L$ is the sampling frequency. Accordingly, from the Eq. 12, the range of m for **Case 1** can be obtained as follows:

$$m < \frac{L}{2} - \frac{f_0}{F_r}, \quad m \in \mathbb{W} \tag{13}$$

Case 2: Denoting $X\left(\frac{\omega L - 2\pi m}{M}\right)$ as $\hat{X}(\omega)$, the following equivalents can be obtained based on the phenomenon that Fourier transform is an even function:

$$\left|\hat{X}(\omega_y)\right| = \left|\hat{X}(-\omega_y)\right| = \left|X\left(\frac{-\omega_y L - 2\pi m}{M}\right)\right| \tag{14}$$

Similarly,

$$\left|X(\omega_0)\right| = \left|X(-\omega_0)\right| \tag{15}$$

Then, the following expressions can be formed:

$$\left|X(-\omega_0)\right| = \left|X\left(\frac{-\omega_y L - 2\pi m}{M}\right)\right| \tag{16}$$

Accordingly,

$$\omega_0 = \frac{\omega_y L + 2\pi m}{M} \tag{17}$$

By substituting ω and ω_0 equivalents in Eq. 10, Eq. 17 can be rewritten as follows:

$$\frac{2\pi f_0}{F_r M} = \frac{2\pi f_y}{F_r L}\frac{L}{M} + \frac{2\pi m}{M}\frac{F_r}{F_r} \tag{18}$$

From the Eq. 18, for **Case 2**, f_y results as:

$$f_y = f_0 - m F_r \tag{19}$$

From the Eq. 19, the set of m values for **Case 2**, for $f_y > 0$, yield the follows:

$$m < \frac{f_0}{F_r}, \quad m \in \mathbb{W} \tag{20}$$

It should be noted that the set of m values in Eq. 20 are for the f_y expression in Eq. 19 only. Similarly, the set of m values provided in Eq. 13 are for the f_y expression in Eq. 12 only. Accordingly, the following pair of equivalences are obtained:

$$f_y = \begin{cases} f_0 + m F_r, & m < \frac{L}{2} - \frac{f_0}{F_r}. \\ f_0 - m F_r, & m < \frac{f_0}{F_r}. \end{cases} \tag{21}$$

where $0 < f_y < \frac{L F_r}{2}$. Any of the ENF components of f_y having relatively high SNR (Signal-to-Noise Ratio) can be used to estimate the ENF variations along time. By using the corresponding pair of equivalences in Eq. 21, the m value
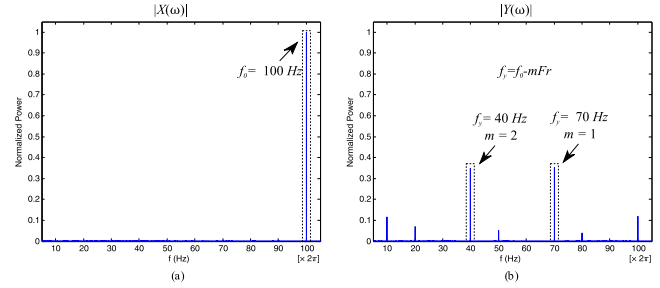

Fig. 5. Frequency spectrum of a video with 30 fps captured in EU, where nominal ENF is 50 Hz, provided that (a) no idle period is applied by the camera: the only ENF component appears at 100 Hz frequency (main ENF harmonic) (b) 45% idle period is applied: new ENF components emerge in some other frequencies. The power of the captured ENF noticeably reduces due to the idle period.

making $\left|Y(\omega)\right|$ maximum can consequently be obtained as follows:

$$m_p = \underset{m}{\arg\max} \left| X_m \left(\frac{\omega L + 2\pi m}{M}\right) \times F_m(\omega)\right| \tag{22}$$

The corresponding f_y for the resulting m_p, which can be computed from the Eq. 12 or Eq. 19, is the strongest emerging ENF component for a particular idle period, i.e., $\frac{M-L}{M} \times 100$. Fig. 5 (a) illustrates frequency spectrum for a video with 30 fps captured in EU, where nominal ENF is 50 Hz, provided that no idle period is applied by the camera. As can be seen from the figure, the only ENF component appears at 100 Hz frequency (main ENF harmonic). Fig. 5 (b) provides emerging ENF components if 45% idle period is applied by the camera. Here, the strongest ENF component appears at 70 Hz frequency ($m = 1$), and the second strongest one appears at 40 Hz ($m = 2$). It can be highlighted that the power of 40 Hz and 70 Hz ENF components are very close since this idle period is a transition point for the greatest ENF components. Fig. 5 (b) also depicts that the power of the captured ENF noticeably reduces due to idle period. Here, the power is normalized. It is also notable that the greater ENF harmonics and emerging ENF components are disregarded.

Fig. 6 (a) illustrates the variation in frequency of the strongest ENF component depending on the proportion of the idle period length (in %) for a video with 30 fps captured in EU (50 Hz mains power grid). Accordingly, the strongest ENF component, 100 Hz for illumination, is replaced with 70 Hz, 40 Hz and 10 Hz respectively for 15%, 45%, 75% idle periods. It can also be noticed from the figure that, the power of the captured ENF decreases as the idle period increases. These outcomes are also validated by simulation results. As can be seen from Fig. 6 (b), the simulation findings are almost the same as those obtained via the proposed model. The slight differences are most likely caused by additional ENF components emerging due to the effect of the 2nd illumination harmonic, 200 Hz, which comes from the absolute cosine form of the illumination signal. Although there may be such extra ENF components, it is validated from both the simulation and the proposed model results that the strongest
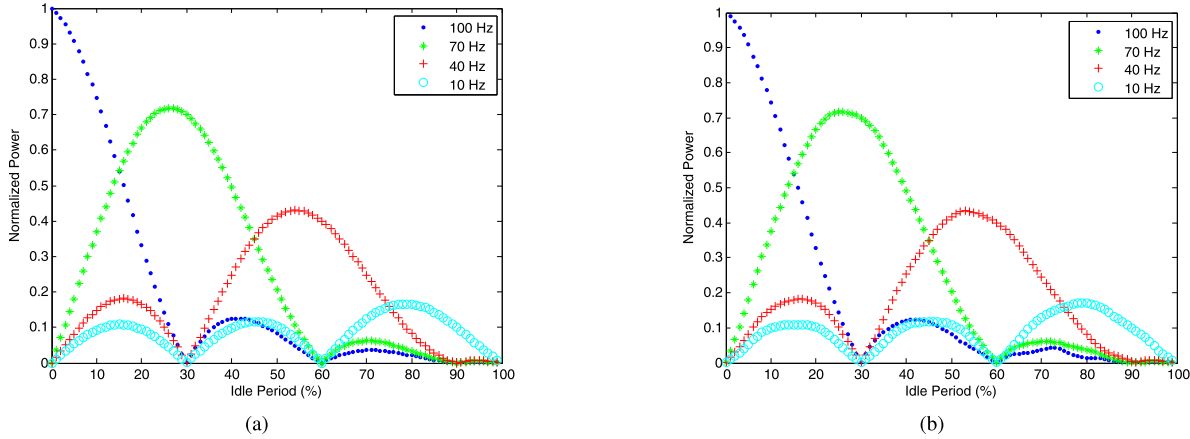
Fig. 6. Variation in frequency of main ENF harmonic vs. idle period for a video with 30 fps captured in EU. (a) The proposed analytical model, (b) Simulation.

two ENF components at any idle period emerge from the 1st illumination harmonic.

III. IDLE PERIOD ESTIMATION

Camera forensics has become a significant field of research for various applications such as verification query image/video pairs can be attributed to the same source camera, and verification the source of a suspicious image/video [23], [24]. To the best of our knowledge, the only work exploiting ENF for camera forensics is that of Hajj-Ahmad *et al.* [18]. They basically compute camera read-out time per frame based on vertical phase analysis [21]. Although the technique they introduced, i.e., vertical phase analysis, may be an effective tool for some cases, it has some significant limitations. First, it is dependent on alias ENF, and so is inapplicable to videos, where the frame rate is a divisor of nominal ENF (0 Hz alias ENF). Second, it may be a challenge for their method to operate on noisy videos due to difficulty in computation of the ENF phase correctly in such conditions. In the subsections below, this state-of-the-art approach is discussed first. Then a novel idle period estimation approach is proposed, which can overcome the limitations of the state-of-the-art. Next, experimental work is provided for a performance evaluation.

A. State-of-the-Art Approach

This subsection, first, briefly discusses how camera read-out time can be estimated via vertical phase analysis [18], [21]. Then, an adaptation of this technique to idle period computation is presented. Results from the adapted technique, including some common properties of idle period are illustrated via experiments conducted on videos with still content (a wall-scene), which are all known to contain ENF. Then the limitations of this approach are explored.

Hajj-Ahmad et al. technique targets cameras with a rolling shutter mechanism. It basically computes the time needed to read one frame (frame read-out time), T_{ro} by estimating phase of ENF for each row along the video frames, i.e., vertical phase analysis as follows:

$$T_{ro} = \frac{L\tilde{\omega}_b}{2\pi \tilde{f}_e} \qquad (23)$$

where L is the number of rows in a frame, $\tilde{\omega}_b$ is vertical radial frequency and $\tilde{f}_e$ is the ENF component that oscillates around the nominal frequency. $\tilde{\omega}_b$ can be computed from the slope of the vertical phase line, e.g., Fig. 7, which can be obtained via the estimation of the phase of ENF, $\Phi[l]$ ($l \in \{1, 2, 3, \ldots, L\}$, where L is number of rows) for each single row [18]. $\Phi[l]$ can be extracted through the Fourier Transform of the lth-row time-series that can be obtained by using the lth-row mean intensity of each frame of the video. Since the sampling frequency of this time-series is simply the video frame rate (and hence Nyquist theorem is not satisfied), peak search in the Fourier domain is made around alias ENF. That is, the property of the increased sampling frequency (*number of rows × camera frame rate*) of the rolling shutter, discussed in II, is invalid. What this approach mainly exploits is the ENF phase shift emerged between consecutive rows due to distinct time of sampling of each row by rolling shutter.

Vertical phase analysis can easily be adapted to perform idle period computation. The adapted technique mainly utilizes 2 key features, namely the number of sinusoidal cycles that would be acquired per frame if there was no idle period and the actual number of sinusoidal cycles that is acquired per frame, i.e., in the presence of idle period. Fig. 7 (a) and Fig. 7 (b) illustrate the vertical phases estimated respectively for a 480 P video (video-1), and a 720 P video (video-2), recorded by Canon PowerShot SX230HS model camera with 30 fps in Turkey, where the nominal illumination frequency is 100 Hz. From the figures, the proportion of the idle period per frame (in %), R_{T_I}, can be computed with the use of the following equation:

$$R_{T_I} = 100 - \frac{f_I}{F_r} \times \frac{100}{N_c} \qquad (24)$$

where N_c is the actual number of sine cycles, i.e., triangles (between $-\pi$ and $+\pi$) obtained through vertical phase analysis. f_I denotes the nominal illumination frequency, and F_r is video frame rate. $\frac{f_I}{F_r}$ results in the number of sinusoidal waves that would be acquired per frame if there was no idle period. Accordingly R_{T_I} values for video-1 and video-2 are obtained as 48% and 38% respectively. From these findings, it can be
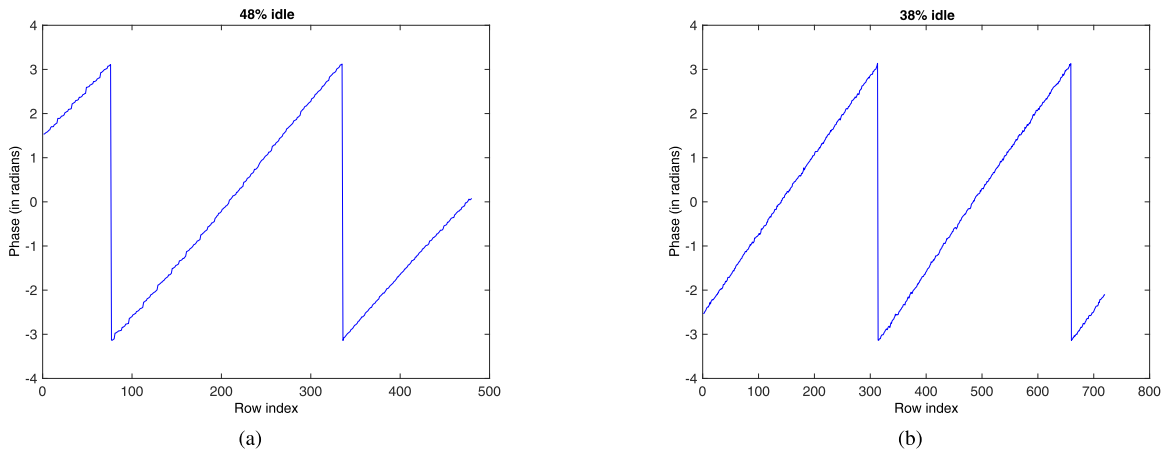
Fig. 7.   (a) Vertical phase analysis for a 480 P wall-scene video, video-1, corresponding to about 48% idle period. (b) Vertical phase analysis for a 720 P wall-scene video, video-2, corresponding to about 38% idle period. Both videos are recorded at 30 fps by CanonPowerShot SX230 HS model camera in Turkey, where illumination frequency is 100 Hz (twice the nominal ENF (50 Hz)).
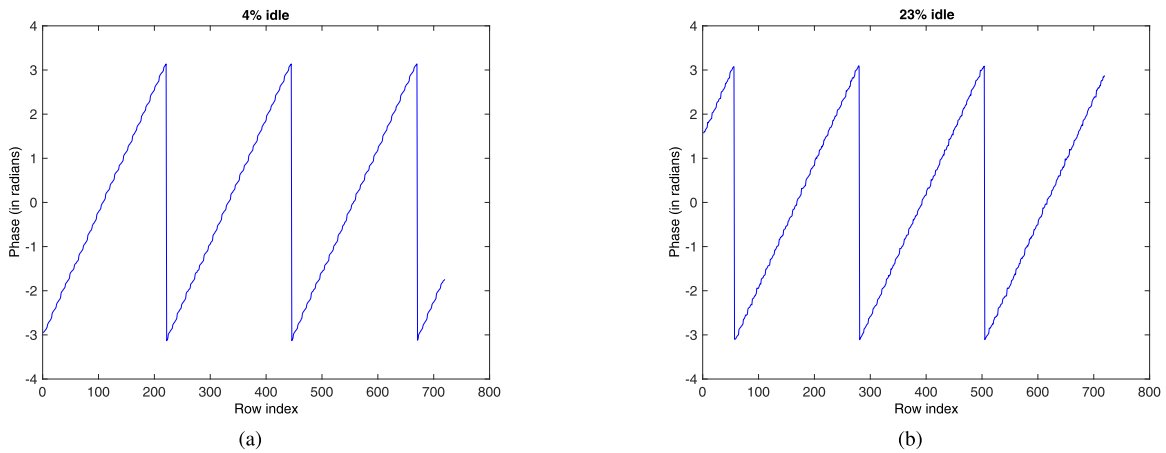

Fig. 8.   (a) Vertical phase analysis for a 30 fps wall-scene video, video-3, corresponding to about 4% idle period. (b) Vertical phase analysis for a 23.976 fps wall-scene video, video-4, corresponding to about 23% idle period. Both videos are recorded in 720 P resolutions by Nikon D3100 model camera in Turkey, where illumination frequency is 100 Hz (twice the nominal ENF (50 Hz)).

deduced that video resolution may a factor affecting the idle period.

Fig. 8 (a) and Fig. 8 (b) illustrate the vertical phases estimated respectively for a 30 fps video (video-3), and a 23.976 fps video (video-4), which are recorded by Nikon D3100 model camera in 720 P in Turkey. Accordingly, R_{T_I} values for video-3 and video-4 are obtained as 4% and 23% respectively. These outcomes show that video frame rate is another factor affecting the idle period.

Although vertical phase analysis technique can provide estimations of idle period time, it has some limitations. First, when the power of the captured ENF in a video is weak due to various reasons including moving content, long idle period and unusual characteristic of some light sources, it may be a challenge to estimate vertical phase correctly. Such an example is illustrated in Fig. 9 (a), computed from a wall-scene (still content) video illuminated by a CFL (compact fluorescent) bulb. As can be seen from the figure, it is really tough for the phase to be estimated in such a condition. It should be noted that spectral responses of different type of illumination are different and CFL bulb illumination is one that may contribute
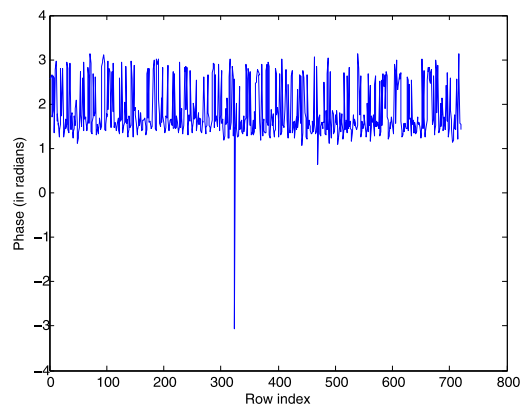
Fig. 9.   Vertical phase analysis for a 23.976 fps video, video-5 - it is a great challenge to compute the phase for this video, which is captured under illumination of **CFL** bulb by Nikon D3100 camera in Turkey.

to the quality of ENF negatively [25]. Another limitation of the vertical phase method is that the vertical phase approach is dependent on alias ENF. Therefore, it is also a great challenge for this approach to work for videos recorded with a frame
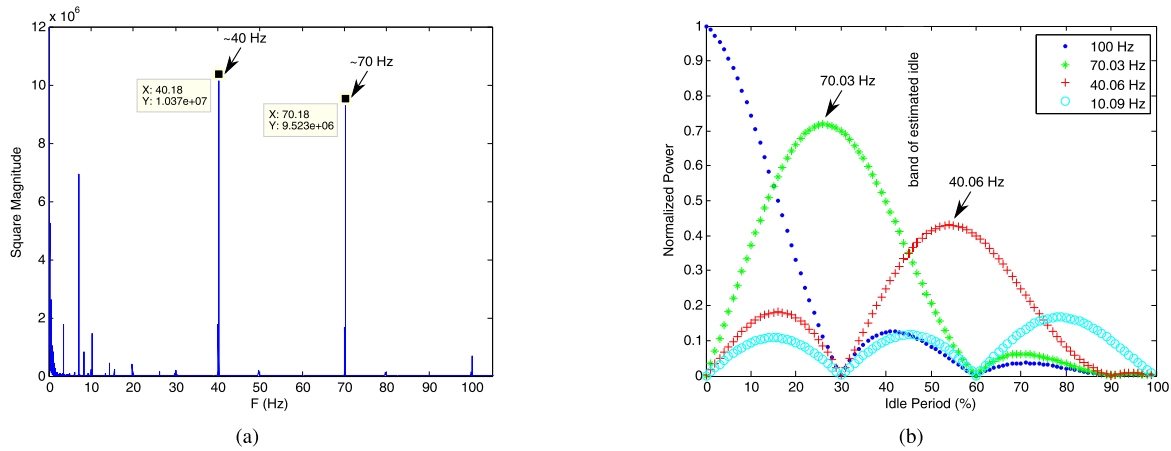
Fig. 10.   (a) Frequency spectrum for the video with 30 fps in 480P resolutions captured by CanonPowerShot SX230 HS, video-1. (b) The reference model - variation in frequency of main ENF harmonic vs. idle period for the same video. The estimated idle is in the range between 45% and 50% (The measured idle [18] is 48%).

rate that is a divisor of the nominal illumination frequency. It should be recalled that alias ENF is obtained at 0 Hz in such a condition and it is extremely thorny to make an estimate of the ENF from this DC component.

B. Proposed Approach

In this subsection, a novel idle period estimation method for videos containing an ENF signal, that can handle the limitations of the state-of-the-art is proposed. The proposed method relies on the analytical model developed in section II-B and so is for videos exposed by a rolling shutter mechanism. According to the model, the nominal illumination harmonic, 100/120 Hz (twice the nominal ENF) is shifted to some other frequency depending on the idle period length as illustrated in Fig. 6. As the idle period increases, captured power of the strongest ENF component decreases, while the second strongest increases. For the proposed approach, the locations of these two ENF components as well as the power ratio between them form the key features.

The proposed approach consists of a number of operational steps. First, the reference analytical model for the test video showing the variation in frequency of the most powerful ENF component depending on the possible idle period is derived based on the video frame rate and the nominal ENF in the region, where the video is recorded, as described in section II-B. Following this, the time-series for illumination variation throughout the video is formed by concatenating the row illumination samples, one illumination sample per row via the steady pixels average, similar to the procedure in Fig. 3 (a). For the details of how to form this time series, the reader is referred to [8] and [9]. Discrete Fourier Transform of this time-series is then computed. Based on the derived analytical model, the strongest two ENF components in the test video are located in the frequency (Fourier) domain. The ratio of the power of the located ENF components are also computed. Next, the corresponding idle point for these ENF components and their power ratio is located using the analytical model. Considering the noise factor, a small band in the vicinity of this point is assigned as candidate idle period

range. The middle point of this range is taken as the estimated idle period.

Unlike to the state-of-the-art [18] which treats the sequence of mean luminance values of the ith rows of a sequence of frames of the video as a separate time series, the proposed approach concatenates the illumination samples of all rows of all consecutive frames. Hence, the sampling rate for the proposed approach is as high as *video frame rate × number of rows* as discussed in section II, which is much higher than the Nyquist Criteria, which leads the proposed approach to operate on videos with any frame rate, i.e., no alias ENF.

The operational procedure of the proposed algorithm is clarified with the experiments in the next subsections. Besides, results demonstrating how the proposed method is able to handle the limitations of the state-of-the art are provided.

C. Experiments With Videos With Still Content

In this subsection, the proposed idle period estimation approach is tested on wall-scene videos that are used in section III-A. The results are compared with that of the vertical phase analysis technique. Besides, a video with 25 fps (a divisor of the nominal ENF, 50 Hz) is tested, where the vertical phase analysis approach is unable to work.

A comparison between the frequency spectrum of a 480 P wall-scene video at 30 fps that was recorded by a Canon PowerShot SX230HS (video-1), and the corresponding model is shown in Fig. 10 (a) and (b), respectively. According to Fig. 10 (a), 40 Hz ENF component has the highest power and 70 Hz ENF component is the second highest. The power ratio between them is around 1.1. The corresponding ENF components in the model, Fig. 10 (b) are located between idle periods of 45% and 50%. Referring back to Fig. 7 (a), the idle period of the same video is estimated as 48% based on vertical phase analysis.

Similarly, a comparison between the frequency spectrum of a 720 P wall-scene video at 30 fps that was recorded by a Canon PowerShot SX230HS (video-2), and the corresponding
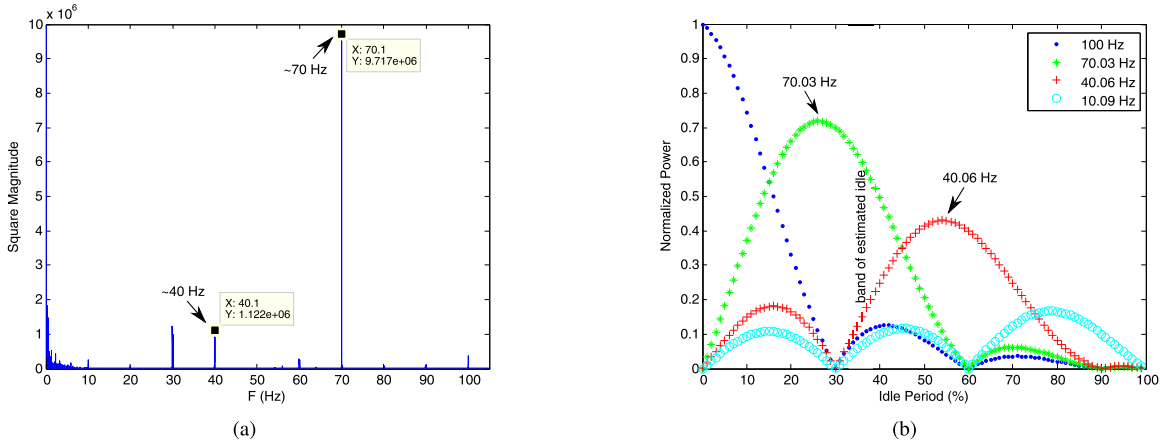
Fig. 11. (a) Frequency spectrum for the video with 30 fps in 720P resolutions captured by CanonPowerShot SX230 HS, video-2. (b) The reference model - variation in frequency of main ENF harmonic vs. idle period for the same video. The estimated idle is in the range between 35% and 40% (The measured idle [18] is 38%).

Fig. 12. (a) Frequency spectrum for the video with 30 fps in 720P resolutions captured by Nikon D3100, video-3. (b) The reference model - variation in frequency of main ENF harmonic vs. idle period for the same video. The estimated idle is in the range between 0% and 5% (The measured idle [18] is 4%).
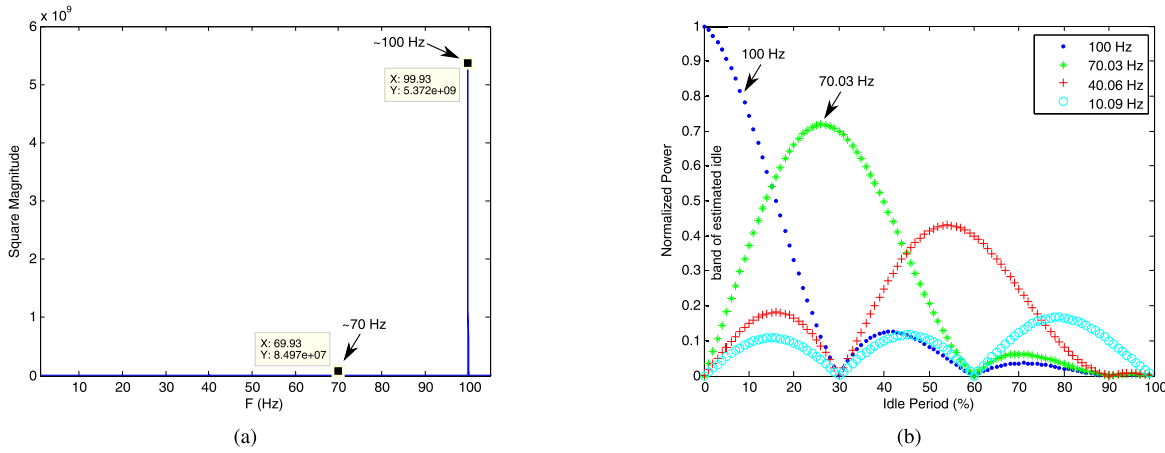
model are shown in Fig. 11 (a) and (b), respectively. In Fig. 11 (a), 70 Hz ENF component is the highest and 40 Hz ENF component is the second highest. The power ratio between them is around 8.5. The corresponding ENF components in the model, Fig. 11 (b) are located between idle periods of 35% and 40%. Referring back to Fig. 7 (b), the idle period of the same video was estimated as 38% based on vertical phase analysis. It is notable that as the proposed approach is independent of video resolution, as the reference model illustrations for video-1 and video-2 are the same as shown in Fig. 10 (b) and Fig. 11 (b).

Similar experiments and comparisons were conducted for a 720P video at 30 fps (video-3), and for a 720P video at 23.976 fps (video-4), recorded by a Nikon D3100. The estimated idle period for video-3 is between 0% and 5%, whereas it is between 20% and 25% for video-4 as shown in Fig. 12 and Fig. 13, respectively. Referring back to Fig. 8 (a) and Fig. 8 (b), the idle period of the same videos were estimated as 4% and 23% respectively for video-3 and video-4 based on the vertical phase method. It is notable that as the proposed approach is dependent on the video frame rate,

reference model illustrations for video-3 and video-4 are not the same as shown in Fig. 12 (b) and Fig. 13 (b).

In Fig. 14 (a), the frequency spectrum of another 23.976 fps wall-scene video, 720 P, recorded by a Nikon D3100 is provided. This video was recorded under CFL bulb illumination, in which the vertical phase analysis failed to work as discussed in section III-A, Fig. 9. When compared to the corresponding model illustration in Fig. 14 (b), the idle period for this video is estimated between 20% and 25%, which is the same as that for video 4 illustrated in Fig. 13. This outcome indicates that the proposed approach is more robust than the state-of-the-art for a noisy video where the power of the ENF signal is weak. This may be due to the reason that the power ratio between the strongest two ENF components is likely to be preserved, even though the power of the each component is reduced.

In Fig. 15 (a) and (b), respectively a comparison between frequency spectrum of 25 fps video in 720 P (video-6), recorded by a Nikon D3100 in a region where nominal ENF is 50 Hz and the corresponding model are provided. Based on the argument in this section, the idle period is estimated between 20% and 25%, which is a similar result to the case for
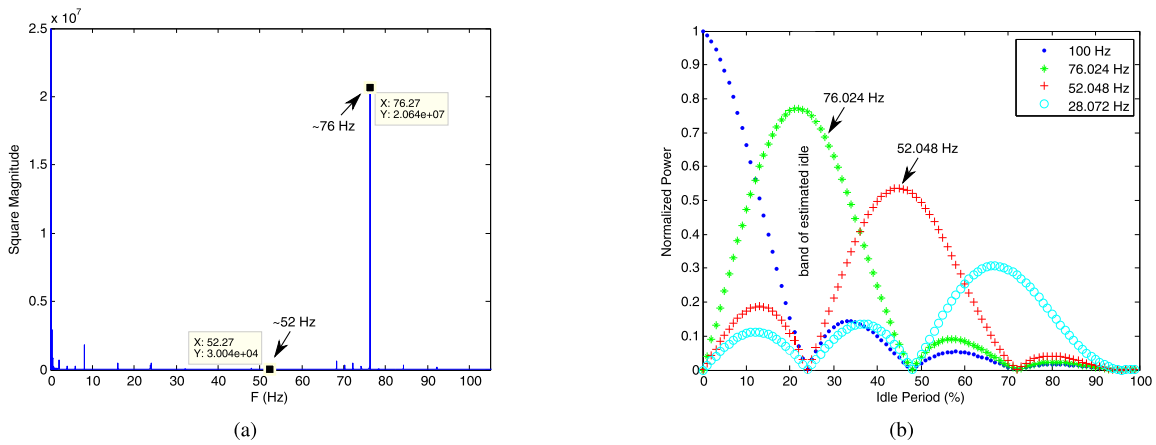
Fig. 13. (a) Frequency spectrum for the video with 23.976 fps in 720P resolutions captured by Nikon D3100, video-4. (b) The reference model - variation in frequency of main ENF harmonic vs. idle period for the same video. The estimated idle is in the range between 20% and 25% (The measured idle [18] is 23%).
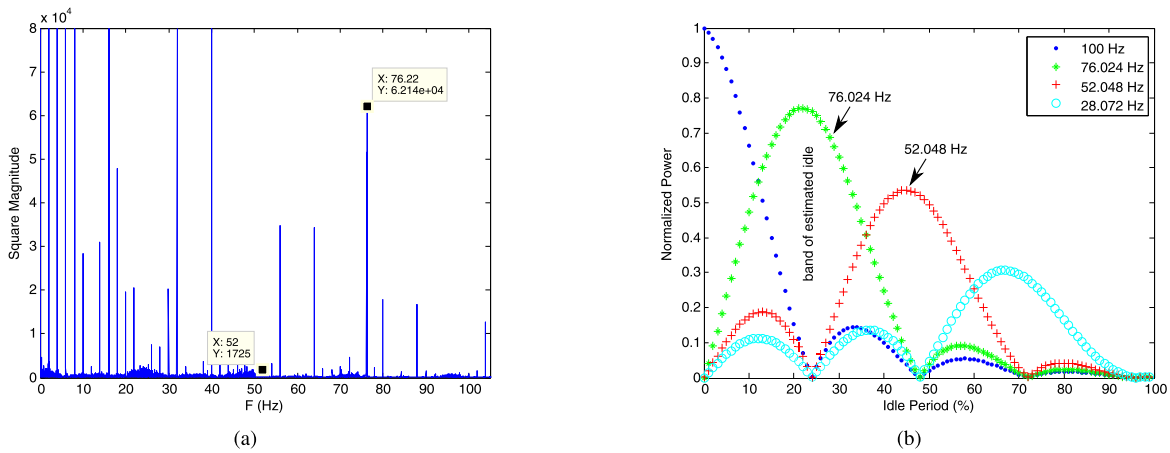
Fig. 14. (a) Frequency spectrum for the another 23.976 fps in 720P video by Nikon D3100, video-5 (captured under **CFL** bulb). (b) The reference model - variation in frequency of main ENF harmonic vs. idle period for the same video. The estimated idle is in the range between 20% and 25% (The measured idle [18] is 23%).
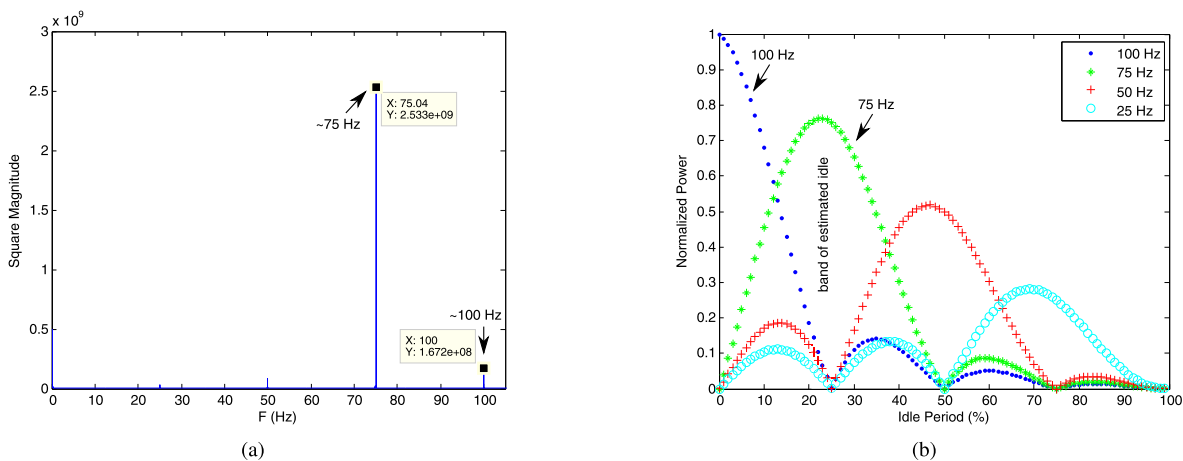
Fig. 15. (a) Frequency spectrum for the video with 25 fps in 720P resolutions captured by Nikon D3100, video-6. (b) The reference model - variation in frequency of main ENF harmonic vs. idle period for the same video. The estimated idle period is in the range between 20% and 25%. Since the video frame rate is half the nominal ENF, i.e., 0 Hz alias ENF, idle period cannot be computed via [18] for this case.

23.976 fps video, as shown in Fig. 13 (a). However this time, when compared to Fig. 13 (a), the proportion of the power of the strongest ENF component to the second strongest is much lower, which corresponds to a smaller idle period. It is reasonable that idle period should be inversely proportional to video frame rate since the total time per frame should

TABLE I

TABULATED FINDINGS OF THE PROPOSED MODEL FOR VARIATION IN FREQUENCY OF MAIN ENF HARMONIC VS. IDLE PERIOD FOR A 29.97 fps VIDEO RECORDED IN EU

Idle Period (%)	H_1 (Hz)	H_2 (Hz)	P_{H_1}/P_{H_2}
0	100	0	$\approx \infty$
5	100	70	5.0
10	100	70	2.0
15	100, 70	100, 70	1.0
20	70	100	2.0
25	70	100	5.0
30	70	100, 40	505.6
35	70	40	5.0
40	70	40	2.0
45	70, 40	70, 40	1.0
50	40	70	2.0
55	40	70	5.0
60	40	10, 70	498.5
65	40	10	4.9
70	40	10	2.0
75	40, 10	40, 10	1.0
80	10	40	2.0
85	10	40	5.1
90	10	40	334.0
95	10	40	6.9

TABLE II

TABULATED FINDINGS OF THE PROPOSED MODEL FOR VARIATION IN FREQUENCY OF MAIN ENF HARMONIC VS. IDLE PERIOD FOR A 25 fps VIDEO RECORDED IN EU

Idle Period (%)	H_1 (Hz)	H_2 (Hz)	P_{H_1}/P_{H_2}
0	100	0	$\approx \infty$
5	100	75	4.0
10	100	75	1.5
15	75	100	1.5
20	75	100	4.0
25	75	100	$\approx \infty$
30	75	50	4.0
35	75	50	1.5
40	50	75	1.5
45	50	75	4.0
50	50	100	$\approx \infty$
55	50	25	4.0
60	50	25	1.5
65	25	50	1.5
70	25	50	4.0
75	25	75	$\approx \infty$
80	25	50	6.0
85	25	50	3.5
90	25	50	2.7
95	25	50	2.2

be smaller for videos with greater frame rate. Hence, it is an indication that the idle period for video-6 is closer to 20% than that for video-4. This experiment shows that the proposed method works also for videos with a frame rate that is a divisor of the nominal ENF. Although highlighted in the previous sections that vertical phase analysis cannot work under such a frame rate, i.e., a divisor of nominal ENF, it may be argued that testing a video with a frame rate in close proximity may provide an approximation to the expected idle period. However, most consumer cameras provide very few options for frame rates. It should also be noted that neither the frame read-out time nor the idle period is provided by camera manufacturers. This is why all the comparisons in this subsection and also in the next are made by using vertical phase analysis.

D. Experiments With Videos With Moving Content

In this subsection, the performance of the proposed idle period estimation technique is evaluated by conducting experiments on videos with moving content that are captured using 5 different fixed cameras. Each dataset for each camera included 6 different videos, half of which were indoor and the other half outdoor. All the videos were captured in Turkey, where the nominal ENF is 50 Hz. Since the power of the estimated ENF signal in a video, and so the performance of the proposed approach may be affected from the type of mains-powered light source, and the power of illumination emitted by mains-powered light source in relative to that of the
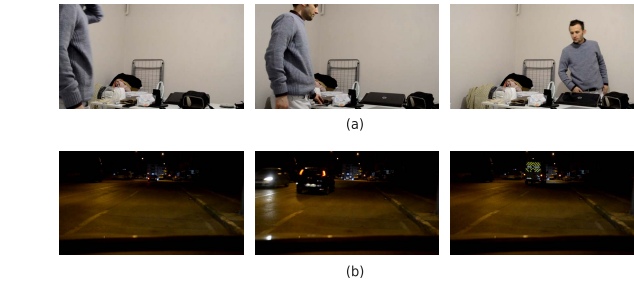
Fig. 16. Sample frames from videos with moving content (a) an indoor video, (b) an outdoor video.

non-mains-powered ones in the scene, various settings were used to build the dataset. Both the indoor and the outdoor videos were recorded under different light sources such as LED bulb, CFL bulb and sodium vapor light. Most of the outdoor videos were captured in poorly illuminated streets. Sample frames from 2 exemplary videos for indoor and outdoor scenes respectively are provided in Fig. 16 (a) and (b).

First, the results for each of the videos from two different cameras were analyzed providing the estimated key parameters. The first dataset included 480 P videos at 29.97 fps recorded by a Canon PowerShot SX230HS model camera, and the second dataset had 25 fps video recorded at 720 P by a Nikon D3100 model camera. Table I and Table II show the largest 2 ENF components and the power ratio between them, at idle periods of multiples of 5%, for 29.97 fps and for 25 fps videos respectively, which are extracted from the analytical

TABLE III

THE SETTINGS AND ESTIMATED OUTCOMES OF EACH VIDEO USED FOR EVALUATION OF THE PROPOSED IDLE PERIOD ESTIMATION METHOD

Camera Model	Vid. no	Res. (P)	Fr. rate (fps)	H_1 (Hz)	H_2 (Hz)	P_{H_1}/P_{H_2}	Estimated idle (%)	Expected idle (%)	Err (%)
Canon SX230HS	1	480	29.97	40	70	1.9	47.5	48.0	**0.5**
Canon SX230HS	2	480	29.97	70	40	1.0	45	48.0	**3.0**
Canon SX230HS	3	480	29.97	70	40	1.1	47.5	48.0	**0.5**
Canon SX230HS	4	480	29.97	40	70	1.27	47.5	48.0	**0.5**
Canon SX230HS	5	480	29.97	40	70	1.21	47.5	48.0	**0.5**
Canon SX230HS	6	480	29.97	40	70	1.4	47.5	48.0	**0.5**
Nikon D3100	7	720	25	75	100	3.5	17.5	21.5	**4.0**
Nikon D3100	8	720	25	75	100	3.6	17.5	21.5	**4.0**
Nikon D3100	9	720	25	75	100	2.1	17.5	21.5	**4.0**
Nikon D3100	10	720	25	50	100	1.6	No match	21.5	-
Nikon D3100	11	720	25	75	100	10.9	22.5	21.5	**1.0**
Nikon D3100	12	720	25	75	50	3.2	32.5	21.5	11.0

TABLE IV

IDLE PERIOD ESTIMATION STATISTICS FOR DIFFERENT CAMERAS FOR 720 P AND 29.97 fps VIDEOS

Camera Model	Expected idle (%) [18]	Estimated idle (%)	Average Estimation Error
GoPro Hero 4	72.69	72.50	1.66
Nikon D3100	4.30	5.00	4.16
Nikon P100	40.55	40.00	2.91
Canon SX230HS	38.00	37.50	2.50
Canon SX220HS	40.59	37.50	5.00

model illustrations in Fig. 12 and in Fig. 15, respectively. They form the reference parameters for comparisons with those of the test videos. Table III provides the main metadata for each test video as well as the location of the largest two ENF components and the power ratio between them obtained via Fourier analysis and from the proposed analytical model. Table III also provides the estimated idle period, the expected idle period and the relative error for each video. Estimated idle period is obtained based on the search of the corresponding match between the reference parameters and the test video parameters. The expected idle period for the 30 fps videos was obtained from still-content videos in section III-A by using vertical phase analysis [18]. The expected idle period for the 25 fps video set is obtained in section III-C mainly based on the proposed approach. Since a 25 fps video set is a divisor of nominal ENF, the vertical phase analysis approach is unable to work for this condition. It should be emphasized also that idle period information is unlikely to be provided by the camera manufacturers. According to the Table III, the idle period of 10 videos out of 12, i.e., 83%, were successfully estimated in the vicinity of $\pm 5\%$ of the expected idle period. The idle period of 1 video was estimated around the vicinity of $\pm 10\%$ of the expected idle period, though the test fails for 1 video, where no match was found.

Second, idle period estimation statistics for the videos in the same settings, yet captured by 5 different cameras, 6 videos for each camera, were explored. The cameras were GoPro Hero 4, Nikon D3100, Nikon P100, Canon SX230HS and Canon SX220HS. Each video by each camera model was captured at 720 P and at 29.97 fps. Table IV provides the median value of estimated idle periods for the videos by each camera and average estimation error based on the estimated median. Accordingly, the median value for the each camera is very close to the expected idle period, and the average estimation error is in close vicinity of the median value. From the table, it can also be noticed that the idle period for Nikon P100, Canon SX230HS and Canon SX220HS cameras are very close to each other. Hence, different cameras may have similar or the same idle period.

One possible application of the idle period in camera forensics may be verifying if 2 videos were produced by different cameras. That is, if the idle periods of the two videos are very similar, it may not be an indication that they are captured with the same camera. However, if the idle periods of the videos are distinct from each other, it is most likely that the videos were captured by different cameras.

IV. A NOVEL TIME-OF-RECORDING VERIFICATION APPROACH

In this section, a novel time-of-recording verification technique for videos exposed by rolling shutter mechanism is proposed. It is based on a systematic search of possible ENF components that emerge as a result of idle period, followed by idle period assumptions in each component and interpolation of missing samples for each assumption.

According to the analytical model introduced in section II-B, the strongest ENF components in an ENF containing video can be located. However, for some videos, these ENF components may contain traces of a non-ENF signal in addition to the ENF. This non-ENF signal may reduce the quality of the ENF signal estimated from these components. On the other hand, the third strongest ENF component, the fourth strongest ENF component and so on, whose power is very low compared to the strongest two, may contain a pure ENF trace. In addition to consideration of the other ENF components, the proposed method also relies on idle period assumptions and interpolation of missing samples for each chosen assumption. Consequently, the use of multiple ENF components, the idle period assumptions in each component, followed by interpolation of missing ENF samples for the each assumption can lead to better quality ENF signal estimations. A higher quality of estimated ENF signal generally leads to obtain a higher correlation coefficient value with the ground-truth ENF, hence resulting in a better performing time-of-recording verification. To the best of our knowledge, no state-of-the-art technique including [7]–[9] rely on idle period assumption and interpolation of missing samples for ENF estimation from videos sampled by the rolling shutter. Hence it may be a challenge for these techniques to obtain a good quality of ENF signal for some cases, including compressed videos and videos with moving content.

The operational procedure of the proposed time-of-recording verification method consists of several steps. First, the time-series for illumination variation throughout the video period is formed via concatenation of the illumination samples of all rows of all frames as in section III-B. Following this, the possible ENF components that may emerge as a result of the idle period are established based on the analytical model. For the each ENF component, the idle period of the video is presumed to be consecutively 5%, 10%, … 95%. For the each idle period presumption, missing illumination samples in each frame are interpolated. In reality, it is a great challenge to estimate so many missing samples properly. However, as the ENF signal does not show abrupt changes, an approximation can be made. The interpolation technique used in this work is based on simply averaging the illumination samples of the previous and of the next frame. From the each of the interpolated time-series, ENF signal can be estimated by utilization of any time domain or frequency domain techniques discussed in [7]. In this work, ENF is computed via Short Time Fourier Transform (STFT) using 20 seconds time windows with 19 seconds overlaps, followed by quadratic interpolation, which results in 1-second temporal ENF resolution [10]. Each estimated ENF signal from the each interpolated time-series is compared with the ground-truth ENF data via normalized cross correlation operation. The resulting peak correlation coefficient and the corresponding lag point for each operation are recorded. All the resulting peak correlation coefficients and the corresponding lag points are analyzed with the use of an appropriate metric proposed in section IV-A, i.e., Metric 3 or Metric 4. If the final lag point yielded by the applied metric corresponds to the difference between the given time-of-recording and the beginning time of the ground-truth ENF, it is concluded that the given video time-of-recording is true.

A. Experiments With Videos With Still Content

The proposed time-stamp verification algorithm was applied on a wall-scene video dataset that was created by Nikon D3100 and Canon PowerShot SX230HS model cameras in Turkey, where nominal ENF is 50 Hz. With the use of each camera, 2 videos with 29.97 fps in 720 P were taken under each of 3 different light sources, namely LED, Halogen and CFL (Compact Fluorescent) bulbs. Each of 12 original videos were compressed via FFMPEG with different compression ratios (150k, 100k, 50k) as well as YouTube. Accordingly, a total of 48 compressed wall-scene videos were created. The performance evaluation of the proposed method was conducted by using the following metrics:

Metric 1 [7]: Peak correlation coefficient ($\rho_{H_i} > TH_c$) at nominal illumination frequency, $H = 100$ Hz, without idle period assumption, $i = \%0$.

Metric 2 [8]: Maximum of peak correlation coefficients ($\rho_{H_i} > TH_c$) obtained for a variety of ENF components, $H \in \{10, 40, 70, 100, 130, 160, 190, 200\}$ Hz, without idle assumption, $i = \%0$.

Metric 3 (Proposed): Maximum of peak correlation coefficients ($\rho_{H_i} > TH_c$) obtained for a variety of ENF components, $H \in \{10, 40, 70, 100, 130, 160, 190, 200\}$ Hz for 30 fps videos recorded in EU, with a variety of idle period assumptions, $i \in \%\{0, 5, 10, \ldots, 95\}$, and interpolation accordingly.

Metric 4 (Proposed): Maximum of normalized Euclidean distance (d_g) to ($n_{l_g}, \rho_{n_{l_g}}$).

$$d_g = \sqrt{n_{l_g}{}^2 + \rho_{n_{l_g}}{}^2} \tag{25}$$

n_{l_g} : Number of lags in a lag group. A lag group is composed of lags being the same in a specified tolerance.

$\rho_{n_{l_g}}$: The greatest peak correlation coefficient obtained for each n_{l_g} ($\rho_{n_{l_g}} > TH_c$).

Metric 3 and Metric 4 are described based on the proposed approach in this work. Whereas Metric 1 and Metric 2 are based on the argument in [7] and [8], respectively, which are provided for comparisons. It should also be noted that TH_c is selected as 0.94 based on empirical analysis.

Table V provides estimated results for each metric for the original wall-scene videos. In the table, TD represents true decision rate, FD depicts false decision rate and ND is no decision. As can be seen from the table, each metric shows 100% true decision rate. However, the performance noticeably changes for the compressed forms of the videos as can be seen in Table VI. The performance of Metric 1 and Metric 2 are lower than Metric 3 and Metric 4. Metric 4 outperforms with a true decision rate 70.83% in relative to 45.83%, 56.25% an 68.75%, respectively for Metric 1, Metric 2, Metric 3.

B. Experiments With Videos With Moving Content

In this subsection, the performance of the proposed time-of-recording verification algorithm is evaluated with the use of

TABLE V

TIME-OF-RECORDING VERIFICATION PERFORMANCE OF THE PRESENTED METRICS FOR ORIGINAL WALL-SCENE VIDEOS

Method	Metric 1 [7]	Metric 2 [8]	Metric 3	Metric 4
TD (%)	100	100	100	100
FD (%)	0	0	0	0
ND (%)	0	0	0	0

TD = True Decision, FD = False Decision, ND = No Decision

TABLE VI

TIME-OF-RECORDING VERIFICATION PERFORMANCE OF THE PRESENTED METRICS FOR THE COMPRESSED WALL-SCENE VIDEOS

Method	Metric 1 [7]	Metric 2 [8]	Metric 3	Metric 4
TD (%)	45.83	56.25	68.75	70.83
FD (%)	0	4.17	4.17	2.08
ND (%)	54.17	39.58	27.08	27.08

TD = True Decision, FD = False Decision, ND = No Decision

TABLE VII

TIME-OF-RECORDING VERIFICATION PERFORMANCE OF THE PRESENTED METRICS FOR VIDEOS WITH MOVING CONTENT

Method	Metric 1 [7]	Metric 2 [8]	Metric 3	Metric 4
TD (%)	62.50	75.00	79.16	83.33
FD (%)	8.33	8.33	12.50	8.33
ND (%)	29.16	16.66	8.33	8.33

TD = True Decision, FD = False Decision, ND = No Decision

the same dataset described in section III-D, i.e., 24 videos with moving content captured in different environments. The same metrics introduced in IV-A are used to test effectiveness of the proposed approach on these dataset. The only difference in the way the proposed approach is applied for the 25 fps videos is that {25, 50, 75, 100, 125, 150, 175, 200} Hz ENF components are used instead of {10, 40, 70, 100, 130, 160, 190, 200} based on the analytical model proposed in section II-B. Other than that, all the steps are applied identically. Table VII provides the estimated results for each metric. Accordingly, Metric 3 and Metric 4 outperform Metric 1 and Metric 2 for this video dataset as well with true decision rates 79.16% and 83.33%, respectively. The true decision rates for Metric 1 and Metric 2 respectively are 62.5% and 75%.

Experiments on videos with both still content and moving content show that exploitation of multiple ENF components, along with idle period assumptions in each component and interpolation of missing ENF samples for the each assumption, yield better results than the state-of-the-art. Although the proposed approach leads to comparable outcomes with metric 3 and with metric 4, metric 4 outperforms.

V. CONCLUSION

In this work, a comprehensive analysis of the rolling shutter mechanism is provided for ENF based video forensics. First, an analytical model is presented illustrating how the frequency of the main ENF harmonic is replaced with new ENF components depending on the length of the idle period per frame. The model also reveals that the power of the captured ENF signal is inversely proportional to idle period length. Hence, in the presence of noise, videos with long idle periods may be tough for ENF based forensic analysis. Second, with the use of this model, a novel idle period estimation approach is proposed for camera forensics. Unlike the state-of-the-art, i.e., vertical phase analysis, which relies on alias ENF, the proposed method is able to operate also on videos, where the frame rate is a divisor of nominal ENF. Thirdly, a novel time-of-recording verification technique is introduced, based on utilization of multiple ENF components as well as idle period assumptions in each component and interpolation of missing samples. This approach can result in higher quality ENF signal estimations, consequently leading the time-of-recording verification performance to increase. Experiments show that the proposed approach outperforms those in the literature.

Although the proposed idle period estimation method can be used as an auxiliary feature in camera forensics, it cannot be ignored that there may be videos with the same idle period, even if they were captured by different camera models. Besides, the proposed technique requires some more research to have better precision. The ENF components that emerge due to the second and the third illumination harmonics, i.e., 200/240 Hz (EU/US) and 300/360 Hz (EU/US) can be taken into consideration. Including these harmonics may also lead to the performance of the proposed time-stamp verification method to improve. Plus, the use of a better interpolation technique, may further augment the performance.

REFERENCES

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

[2] C. Grigoras, "Digital audio recording analysis–the electric network frequency criterion," *Int. J. Speech Lang. Law*, vol. 12, no. 1, pp. 63–76, Aug. 2005.

[3] C. Grigoras, A. Cooper, and M. Michalek, "Best practice guidelines for ENF analysis in forensic authentication of digital evidence," *Forensic Speech Audio Anal. Work. Group*, vol. 1, pp. 1–10, Aug. 2009.

[4] J. Chai, F. Liu, Z. Yuan, R. Conners, and Y. Liu, "Source of ENF in battery-powered digital recordings," *Audio Eng. Soc. Conv.*, vol. 135, p. 9007, Oct. 2013.

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

[6] R. Garg, A. 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*, Dec. 2011, pp. 23–32.

[7] 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.

[8] 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.

[9] 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," Univ. Maryland, College Park, MD, USA, Tech. Rep., Dec. 2015.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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*, 2014, pp. 13–18.

[15] 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.

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

[17] 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.

[18] 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.

[19] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing: Principles, Algorithms, and Applications*, 4th ed. London, U.K.: Pearson Education, 2007.

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

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

[22] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.

[23] H. T. Sencar and N. Memon, *Digital Image Forensics: There is More to a Picture than Meets the Eye*. New York, NY, USA: Springer, 2013.

[24] D. Shullani, M. Fontani, M. Iuliani, O. A. Shaya, and A. Piva, "Vision: A video and image dataset for source identification," *J. Inf. Secur.*, vol. 2017, no. 1, p. 15, Oct. 2017. doi: 10.1186/s13635-017-0067-2.

[25] S. Vatansever and A. E. Dirik, "The effect of light source on ENF based video forensics," *Uludag Univ. J. Fac. Eng.*, vol. 22, no. 1, pp. 53–64, 2017. doi: 10.17482/uumfd.305193.

Saffet Vatansever received the B.Sc. degree in electronics and communications engineering from Yildiz Technical University, Bursa, Turkey, and the M.Sc. degree in mechatronics engineering from Newcastle University, Newcastle upon Tyne, U.K. He is currently pursuing the Ph.D. degree in electronics engineering with Uludag University, Bursa. He is currently a Research Assistant with the Department of Mechatronics, Bursa Technical University, Bursa. His main research area is multimedia forensics.

Ahmet Emir Dirik received the B.Sc. and M.Sc. degrees in electronics engineering from Uludag University, Bursa, Turkey, and the Ph.D. degree in electrical engineering from the Polytechnic Institute, New York University Tandon School of Engineering, Brooklyn, NY, USA, in 2010. He is currently an Associate Professor with the Department of Computer Engineering, Uludag University. His research interests include multimedia forensics and signal processing.

Nasir Memon received the M.Sc. and Ph.D. degrees in computer science from the University of Nebraska. He is currently a Professor with the Department of Computer Science and Engineering, New York University (NYU) Tandon School of Engineering, the Director of the OSIRIS Laboratory, and a Founding Member of the Center for Cyber Security. He is also with the Collaborative Multidisciplinary Initiative of several schools within NYU. He has authored over 250 articles in journals and conference proceedings. He holds a dozen patents in image compression and security. His research interests include digital forensics, biometrics, data compression, network security, and usable security. He has received several awards, including the Jacobs Excellence in Education Award and several best paper awards. He is a fellow of the IEEE and SPIE. He was the Editor-in-Chief of the IEEE TIFS. He has been on the editorial boards of several journals.