"Seeing" ENF: Natural Time Stamp for Digital Video via Optical Sensing and Signal Processing *

Ravi Garg
University of Maryland
College Park, USA
ravig@umd.edu

Avinash L. Varna
University of Maryland
College Park, USA
varna@umd.edu

Min Wu
University of Maryland
College Park, USA
minwu@umd.edu

ABSTRACT

Electric Network Frequency (ENF) fluctuates slightly over time from its nominal value of 50 Hz/60 Hz. The fluctuations in the ENF remain consistent across the entire power grid even when measured at physically distant locations. The near-invisible flickering of fluorescent lights connected to the power mains reflect these fluctuations present in the ENF. In this paper, mechanisms using optical sensors and video cameras to record and validate the presence of the ENF fluctuations in fluorescent lighting are presented. Signal processing techniques are applied to demonstrate a high correlation between the fluctuations in the ENF signal captured from fluorescent lighting and the ENF signal captured directly from power mains supply. The proposed technique is then used to demonstrate the presence of the ENF signal in video recordings taken in various geographical areas. Experimental results show that the ENF signal can be used as a natural timestamp for optical sensor recordings and video surveillance recordings from indoor environments under fluorescent lighting. Application of the ENF signal analysis to tampering detection of surveillance video recordings is also demonstrated.

Categories and Subject Descriptors

I.5.4 [**Applications**]: Signal Processing; K.6.5 [**Security and Protection**]: Authentication

General Terms

Security, Algorithms, Experimentation.

Keywords

Video Authentication, Timestamp, Information Forensics, Electric Network Frequency.

*Area chair: Dr. Qibin Sun

MM'11, November 28–December 1, 2011, Scottsdale, Arizona, USA.

1. INTRODUCTION

In the modern era, a huge amount of digital information is available in the form of audio, image, video, and other sensor recordings. Stored on disks and other storage devices, this information has metadata describing the time and place of recording. However, digital tools can be used to modify the stored information. For example, digital editing softwares can be used to cut a clip from an original audio or insert a clip from one audio into another audio, or manipulate the metadata field to alter the recording date. Similar changes can also occur with video surveillance and other recordings. In the absence of any cryptographic protection and watermarking techniques during initial data acquisition, such modifications can be difficult to detect. Developing forensic tools to authenticate data using natural timestamp in the stored digital data presents an attractive direction to complement the existing technologies.

Exploration of Electric Network Frequency (ENF) is an emerging direction to authenticate digital recordings. ENF is the supply frequency of power distribution networks in a power grid. The nominal value of the ENF is 60 Hz in the United States and most parts of the Americas; and in Europe and Asia, 50 Hz is the generally accepted norm except for some parts of Japan and Taiwan. Digital devices, such as audio recorders, CCTV recorders and camcorders that are plugged into the power mains or are located near power sources, pick up the ENF signal because of interference from electromagnetic fields generated by power sources [3]. Battery operated devices are also capable of capturing the ENF signal from electromagnetic fields generated by nearby transmission lines or power equipment.

Importantly, the value of the ENF is not constant, fluctuating from the nominal value because of varying loads on the power grid [3]. For example, in Europe, these fluctuations make the ENF a continuous random variable typically between 49.90 Hz and 50.10 Hz [3]. Similarly, the ENF fluctuations in the United States vary between 59.90 Hz and 60.10 Hz [10]. The fluctuations from the nominal value are consistent across the power grid, and it has been shown that frequency fluctuations measured at locations physically far apart but connected on the same power grid have a similar value [3] [10]. Therefore, the ENF signal recorded at any location in the network of the power grid can serve as a representative signal for the ENF in the area covered by the power grid.

Fluctuations in the ENF have been used successfully to authenticate audio recordings in [3] [10]. These works demonstrate that the ENF signal is captured in audio recordings

and exhibit high correlation with the ENF signal recorded from the power mains supply at the corresponding time. It was also shown that the ENF signal can be used as a natural timestamp to authenticate audio recordings, both stand-alone and those associated with videos. All of these works require the active recording of the true ENF signal from the power grid to be used as the ground truth timing to help verify the authenticity of signal content. ENF signal is extracted from the audio in question and compared with the ENF patterns from the power grid stored in a database to determine the time of recording or authenticity of audio. This method also provides rough information about the geographical location in terms of the grid where recording occurred. In contrast to the active recording of the ENF signal, the phase continuity of the ENF signal is used to detect the region of tampering in [9]. Any sudden change in the phase of the ENF signal extracted from audio strongly indicates that the original recording has been modified. This method does not require the use of reference ENF to authenticate a given recording, though the time and the place of recording cannot be determined without matching against a database of reference ENF signals.

Although audio authentication using the ENF signal has been explored in recent years, to the best of our knowledge, no research has investigated the traces of the ENF in visual sensing data such as optical sensor recordings and video recordings, or to bind audio and visual tracks of recordings forensically. An important research question to address in this direction is whether the ENF signal is present and measurable in visual sensing data. Optical sensors and CCD sensors used in video cameras measure the intensity of photons falling on the sensor array and convert them to measurable quantities, such as electric current or further into digital image sequence. The photons generated from indoor fluorescent light sources flickering in accordance to the frequency of the electric current may induce the ENF signal in these recordings. Optical sensors with high temporal bandwidth can potentially measure changes in illumination from fluorescent lightings. On the other hand, CCD sensors used in video cameras capture visual data at the rate of 25/30 frames per second, which is considerably lower than the nominal ENF value of 50 Hz or 60 Hz. So, it is challenging to extract ENF fluctuations from such video recordings due to significant aliasing.

Our research starts with devising experiments to verify the presence of the ENF signal in optical sensor measurements under fluorescent lighting. After establishing the presence of ENF signals, we devise a methodology to extract the ENF signal from video recordings under fluorescent lighting. As will be shown later, the ENF signal in optical sensor and video recordings shows significant correlation with the reference ENF signal extracted from the power mains supply, which can be used to verify or estimate the time of recording. As information about the time of recording can facilitate verification of the integrity of the video recording, we also examine potential applications of the proposed technique to verify the integrity of surveillance videos. Another potential multimedia application of the ENF analysis that we study in this paper determines if the audio and the visual tracks in a video recording are actually captured together, or if the sound track from a different recording has been manually added to the visual track. ENF signals that are extracted from both visual and audio tracks can provide natural alignment and binding of audio-visual recordings.

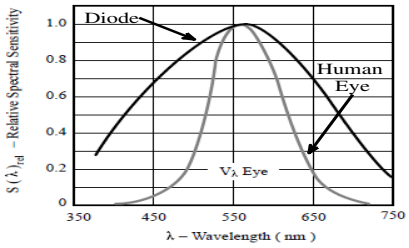


Figure 1: Spectral sensitivity of BPW21 photo diode (From: Vishay Semiconductor Web Catalog).

The rest of this paper is organized as follows. In Section 2, a method to detect the presence of the ENF signal from fluorescent lights is developed and validated through several experiments. In Section 3, the presence of ENF signals is established in video recordings in indoor environments illuminated by fluorescent lights, and the use of ENF signal as a natural timestamp for tampering detection of video recordings is demonstrated. Section 4 concludes the paper.

2. ENF MEASUREMENTS USING OPTICAL SENSORS

2.1 Background on Fluorescent Lamps

To understand the mechanism of ENF fluctuations in fluorescent lighting, we present a brief background on the operation of fluorescent lamps. A fluorescent lamp consists of a sealed tube containing a tungsten filament at both ends, an inert gas, and a small amount of mercury vapor. On passing electric current through this tube, electrons traveling in the form of electric current collide against the mercury atoms. If these electrons have enough kinetic energy, electrons in the mercury atoms become temporarily excited to higher energy levels. Unstable high energy levels cause electrons to return to lower energy levels, releasing photons with energy in the ultra-violet(UV) range. Since UV light is invisible to human eyes, the inner wall of the tube is coated with phosphor which absorbs UV light and emits light in the visible light range through another electron excitation process. Applying the electric current at the nominal ENF of 50 Hz/60 Hz to the tube heats the tungsten filament and generates free ions to establish a current path. As the current changes polarity at twice the rate of the nominal ENF, the current path inside the lamp turns on and off at 100/120 times per second. As a result, the lighting fluctuations illuminated by a fluorescent source is expected to be influenced by the ENF signal at a nominal value of 100 Hz/120 Hz.

2.2 Measuring Optical Signal under Fluorescent Light Sources

The first step of our study investigates the presence of the ENF signal in fluorescent lighting fluctuations. The typical wavelength of light emitted by fluorescent lighting occurs in the visible range (400nm-700nm). To record the light intensity, we use photo diodes that produce electric current when light falls on the diodes. We choose BPW21 photo-diodes that have light sensitivity of $9nA/lx$ and high

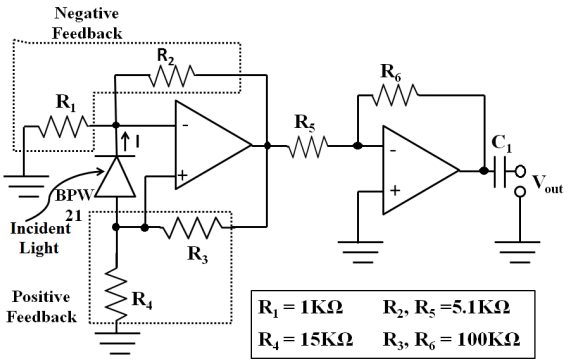
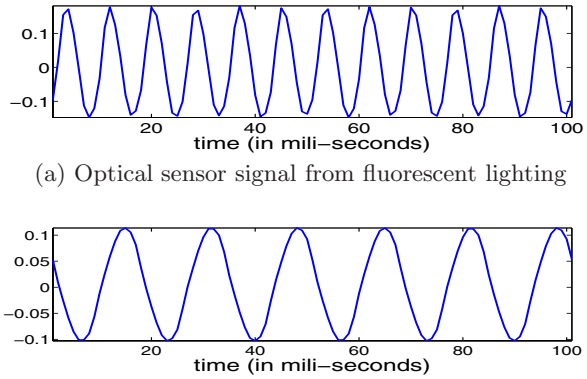


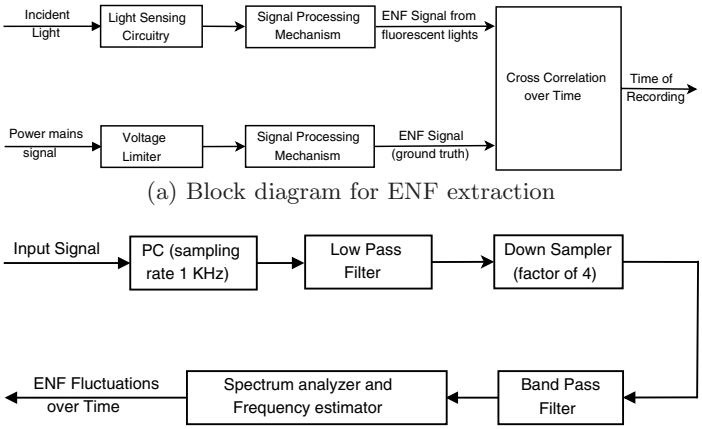
Figure 2: Light sensing circuitry.



(a) Optical sensor signal from fluorescent lighting

(b) Power mains signal

Figure 3: 100**ms segments of ENF signals.**



(a) Block diagram for ENF extraction

(b) Details of the signal processing module in (a)

Figure 4: Block diagram for estimating the time of recording of sensor signal.

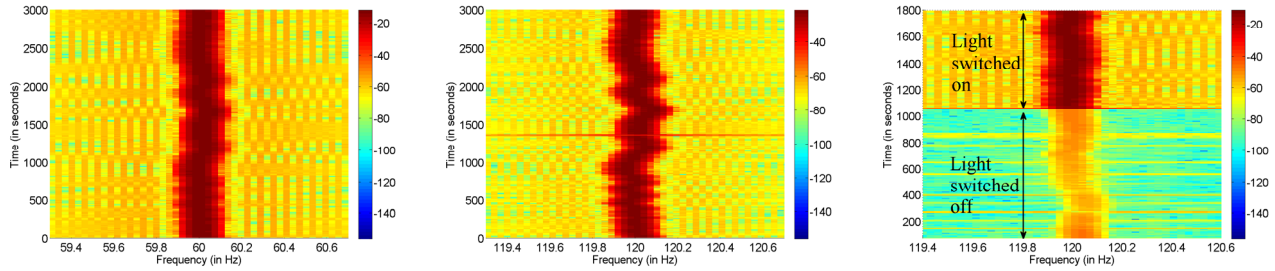
spectral sensitivity in the visible range, as shown in Fig. 1. Using the diode with high spectral sensitivity in the visible range ensures that any infra-red signal generated from nearby heating sources does not interfere with the measurements obtained using the diode. This type of diode has been used in the power engineering literature to verify the presence of inter-harmonic flicker in fluorescent lighting [2]. The BPW21 diodes are extremely sensitive and can capture light shining at an angle or even weak ambient light. As discussed later in Section 2.4, when the fluorescent lamp directly above the sensor is switched off, the ENF signal is still recorded due to the low ambient light present in the corner of the room where recordings happen. The current generated by the photo diodes is on the order of μA, too weak to be directly recorded by common current measuring devices. We use a current-to-voltage amplifier to convert the current to measurable voltage levels. The electric circuit used for this amplification purpose is shown in Fig. 2. In the first stage, a combination of a positive and a negative feedback loop is applied to obtain a robust amplification. Use of the positive feedback loop increases the gain factor and produces less noisy amplification as discussed in [5]. In the second stage, a negative feedback loop is applied to further amplify the output from the first stage. For the given values of circuit parameters in Fig. 2, the overall gain of this circuit from input photo current I induced by incident light to output voltage V_{out} is $8 \times 10^6 \Omega$. The signal obtained after amplification is given as input to a PC sound card and sampled at 1 KHz.

To study the correlation between the ENF signal in fluorescent lighting captured using optical sensors and the ENF signal in power mains supply, we record the ground truth ENF signal directly from the power mains of the electrical supply. We first use a step-down transformer to convert the power supply voltage level to 5V. The transformer output is then given to a voltage divider to limit the current and voltage at safe levels so that it can be given as an input to a PC sound card. A similar mechanism was also used to capture the ENF signal in audio forensics [3]. These experiments are conducted in the state of Maryland (which is part of the eastern power grid of the United States), and the ENF signal captured by optical sensors is expected at around 120 Hz as discussed in Sec. 2.1. To validate the presence of the ENF signal at 120 Hz, Fig. 3(a) shows a time domain plot of a 100ms segment of the optical sensor signal captured under fluorescent lighting and recorded by a PC. The signal can be clearly seen oscillating at a fundamental frequency of about 120 Hz or an equivalent fundamental period of 8.33ms, suggesting that fluorescent lights flicker in accordance with the ENF. The corresponding power mains signal is shown in Fig. 3(b) oscillating at a frequency of about 60 Hz or an equivalent fundamental period of 16.66ms.

To extract the ENF signal, the recorded signals are given as an input to an anti-aliasing low-pass filter with passband of 125 Hz. The resulting signal is downsampled by a factor of 4 to reduce the sampling rate from 1 KHz to 250 Hz, and then passed through an equiripple band-pass filter with a narrow passband centered at 120 Hz for the optical sensor signal and 60 Hz for the power mains signal.

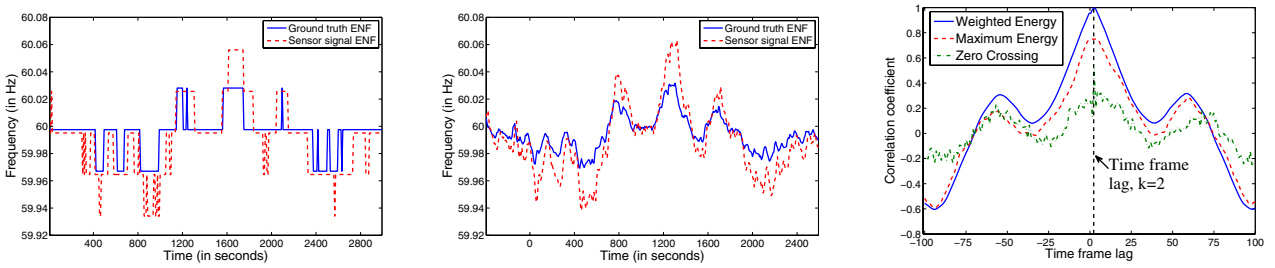
2.3 Correlation between optical sensor signal and power mains signal

In this section, we describe the methods used to compare the ground truth ENF signal captured from the power mains with the ENF signal from fluorescent lighting captured using optical sensors. After pre-processing recorded signals using the method described in Section 2.2, we find the dominant



(a) Power mains ENF signal (b) Optical sensor ENF signal (c) Optical sensor ENF signal

Figure 5: (a),(b) ENF fluctuations of signals captured in Experiment 1; (c) ENF for sensor signal in Experiment 2, including a weak signal for first 17 minutes from the ambient light, when the main light is switched off.



(a) Maximum Energy (b) Weighted Energy (c) NCC Coefficient, $\rho(k)$

Figure 6: (a),(b) ENF fluctuations for Experiment 1 measured using different frequency domain methods; (c) NCC coefficient, $\rho(k)$, as a function of time frame lag k.

instantaneous frequency in each signal to measure the fluctuations in ENF as a function of time. For this purpose, we explore different frequency estimators in time and frequency domains.

Among time domain techniques, we examine the zero-crossing method to estimate the frequency fluctuations in the ENF signal [8]. We divide the signal into overlapping frames of 16 seconds each with an overlap factor of 50%, and count the number of times the signal crosses zero in each frame. The dominant instantaneous frequency in each frame is recorded as one half of the zero-crossing count. This method is easy to compute, but does not give a high estimation accuracy for slowly varying signals. Hence, we also use spectrogram based methods to analyze fluctuations in the ENF signal with higher resolution.

Spectrogram is a common frequency domain analysis tool that can be used to determine the dominant instantaneous frequency of the signal as a function of time [4]. To obtain the spectrogram of the ENF signal, we divide the signal into overlapping frames of 16 seconds each with an overlap factor of 50%. A high resolution FFT of 8192 points is taken for each frame, providing a frequency resolution of approximately 0.03 Hz. After obtaining the spectrogram of the ENF signal, we use the following frequency domain measures to find the dominant instantaneous frequency in the ENF signal:

(a) Maximum energy: In this method, the frequency corresponding to the maximum energy in each time bin of the spectrogram is identified. The resulting one dimensional signal represents the dominant instantaneous frequency as a function of time. This method is susceptible to frequency outliers caused by sudden changes in energy due to factors such as variations in ambient lighting. For example, movement near light sensors can cause transient changes in the maximum energy.

(b) Weighted Energy: The weighted energy method records the weighted average frequency in each time bin of the spectrogram. Average frequency is obtained by weighing frequency bins around the ENF value with the energy present in the corresponding frequency bin as

$$F(n) = \frac{\sum_{l=L1}^{L2} f(n,l)|S(n,l)|}{\sum_{l=L1}^{L2} |S(n,l)|}, \qquad (1)$$

where $L1 = \lfloor \frac{(f_{ENF}-0.5)N_F}{fs} \rfloor$ and $L2 = \lceil \frac{(f_{ENF}+0.5)N_F}{fs} \rceil$; fs and N_F are the sampling frequency and the number of FFT points used to compute the spectrogram; $f(n,l)$ and $S(n,l)$ are the frequency and the energy in the l^{th} frequency bin of the n^{th} time frame of the spectrogram of the recorded signal, respectively. Since we are interested in estimating the dominant instantaneous frequency around a known frequency value, i.e., around the nominal value of the ENF; the value of l is chosen to include the band within ± 0.5 Hz of the ENF frequency of interest.Weighting brings robustness against outliers, so this method provides a more accurate

estimate of the instantaneous frequencies for a given ENF signal than the maximum energy method.

Compared to the time-domain zero-crossing method, the spectrogram based methods for estimating instantaneous frequency provide significantly higher accuracy. After estimating the frequency fluctuations in ENF signals as a function of time, we find the Normalized Cross-Correlation (NCC) coefficient between the two signals as

$$\rho(k) = \frac{\sum_{n=1}^{N}(F_s(n)-\mu_s)(F_m(n-k)-\mu_m)}{\sqrt{\sum_{n=1}^{N}(F_s(n)-\mu_s)^2 \sum_{n=1}^{N}(F_m(n-k)-\mu_m)^2}}, \quad (2)$$

where $F_s(n)$ and $F_m(n)$ is the dominant instantaneous frequency in nth time frame of the ENF signal in optical sensor recording and power mains recording, respectively; μ_s and μ_m are the mean values of $F_s(n)$ and $F_m(n)$, respectively; and N is the total number of frames in optical sensor signal. The peak in the value of $\rho(k)$ represents the estimated delay of k frames in the time of recording between the optical sensor signal and the power mains signal. The block diagram representation of the overall system used to extract the ENF fluctuations over time and compare them is shown in Fig. 4.

One of the advantage of using a correlation based metric as compared to the Euclidean distance based metrics is that the NCC coefficient depends only on the shape of the ENF signal, whereas distance based metrics rely on the absolute value of the signal. If frequency offset exists between the ground truth ENF signal and sensor ENF signal, then distance based metrics may produce wrong estimates of the time of recording. Frequency offset is generally unknown in advance and it may not be possible to compensate for the offset in the distance metric calculations. Use of a correlation based metric to compare the two ENF signal avoids this limitation of distance based metric [6].

2.4 Results and Discussions

In this section, we describe different experiments conducted to determine the presence of the ENF signal under fluorescent lighting. A light sensing circuitry shown in Fig. 2 is assembled, and its output signal is given as an input to a PC sound card to record it. Power mains signal is also recorded in parallel as the ground truth ENF signal.

Experiment 1: In our first experiment, a fluorescent light is switched on near the light sensing circuitry, and data is recorded for a duration of 50 minutes. Power mains signal recording is started with a delay of 15 seconds from the optical sensor recording in order to determine the accuracy of our method in estimating the time of recording. Spectrograms of the ENF signals obtained from the ground truth signal and the optical sensor signal around the frequencies of interests are shown in Fig 5(a) and Fig. 5(b), respectively. From these figures, we observe that the fluctuations in the ENF signal captured under fluorescent lighting is similar to the fluctuations in the ground truth ENF signal. High correlation between these fluctuations can be seen in the instantaneous frequencies obtained using the maximum energy and the weighted energy frequency estimation methods, as shown in Fig. 6(a) and Fig. 6(b), respectively.

As previously discussed, the fundamental component of the ENF signal from sensors is present at a frequency of 120 Hz. In order to facilitate examination of the ground truth ENF signal and the sensor ENF signal on the same scale, we subtract a DC value of 60 Hz from the sensor ENF signal to obtain the plots shown in Fig. 6(a) and Fig. 6(b). The plot of NCC coefficient, $\rho(k)$, for different frequency estimation methods described in Section 2.3 is shown in Fig. 6(c). The maximum value of the NCC coefficient occurs at the lag of two time frames for all three methods, verifying that the delay between optical sensor signal recording and power mains signal recording is approximately 16 seconds. The NCC coefficient found using the zero-crossing method is miniscule compared to the frequency domain methods because of the low precision of the zero-crossing method for frequency estimation. The NCC coefficient obtained using the maximum energy method is slightly smaller than that obtained using the weighted energy method. This is due to the presence of outlier frequencies affects the maximum energy method.

Experiment 2: In our second experiment, a fluorescent light is switched off for the first 17 minutes and then switched on for the next 13 minutes. In this experiment, the power mains signal is started with a delay of 70 seconds from the sensor signal. The room where the experiment is performed also receives some light from a low ambient light source. The spectrogram of the optical sensor recording for this experiment is shown in Fig. 5(c). From this figure, we observe that the ENF signal is weak for the period when the light is off. However, the fluctuations still show high correlation with the ground truth signal (corresponding power mains signal not shown due to space limit), implying that optical sensors can also capture fluctuations from ambient fluorescent lights. In this case, the NCC coefficient is found to be maximum for the lag of nine time-frames or equivalently 72 seconds, authenticating the approximate timing at which the recording started.

In our experiments, we also find a weak signal around 60 Hz in the optical sensor recordings. The energy in this signal is approximately 20 times less than that of the ENF signal at 120 Hz when fluorescent lights are functioning. This 60 Hz signal component is highly correlated with the ground truth ENF signal, indicating that ENF fluctuations are also captured in the optical sensor signal around the nominal ENF value of 60 Hz. We conjecture that the source of this signal may be from electromagnetic interferences of power supply frequency with the light sensing circuitry.

3. ENF MEASUREMENTS IN VIDEO

The ENF signal measurement in video is not as straightforward compared with the ENF measurements in optical sensors due to the low temporal sampling frequency (frame rate) of video cameras. The ENF signal is present at 50 Hz or 60 Hz, which is considerably higher frequency than the sampling rates in commonly available video cameras. Most of these video cameras are available in three different sampling rates: 24fps, 25fps, and 30fps. Cameras using 24fps are used in movie making, while most amateur hand-held cameras are available in 25fps or 30fps. Cameras with 25fps sampling rates originated from analog TV standard of PAL prevalent in Europe and most of Asia, while 30fps camera originated from NTSC standards prevalent across North America and Japan [1].

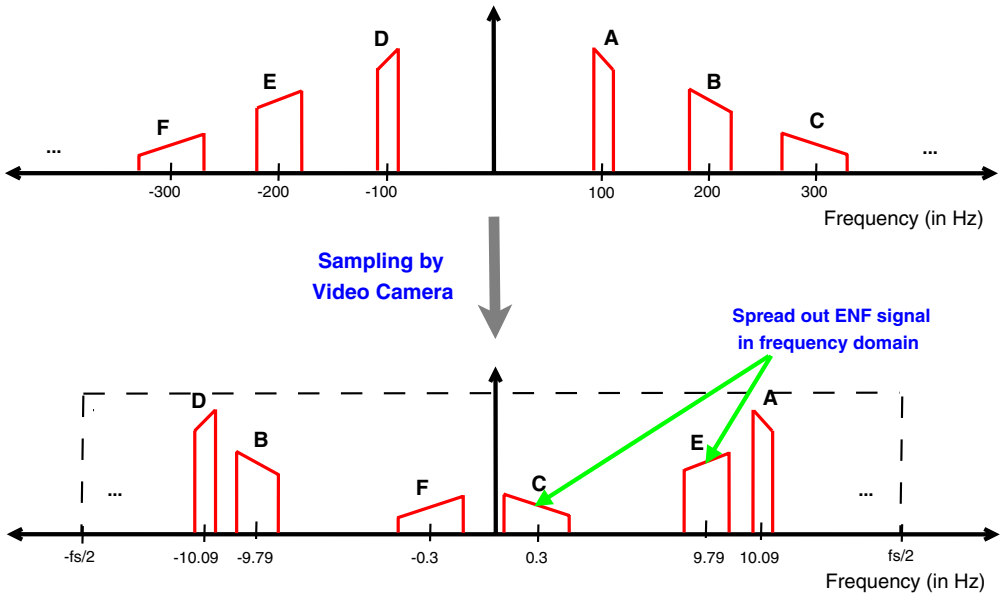


Figure 7: Spectral domain illustration of the aliasing effect in video capturing of ENF signals: (Top) Spectral composition of the original fluorescent light signal powered using 50 Hz supply frequency; (Bottom) Spectral composition of ENF related signals after sampling at 29.97fps by video camera.

The exact frame rates used in most of video cameras may not exactly meet 25fps or 30fps but vary slightly. The PAL standards specify the frame rate to be 24.98fps, and the NTSC standards specify it to be 29.97fps. These slight changes in the recording rates from the nominal values were introduced to make the transmission of color video compatible with black and white television standards and then newly introduced color televisions in 1960s. As different digital video camera manufacturers may not strictly follow the standards, it is common to have NTSC digital cameras with a frame rate of either 29.97fps or 30fps. Similarly, PAL cameras commonly have frame rates of 24.98fps or 25fps. A video recording in an indoor environment using these readily available cameras under fluorescent lighting may lead to significant aliasing of the captured ENF fluctuations depending on the sampling frequency of the camera.

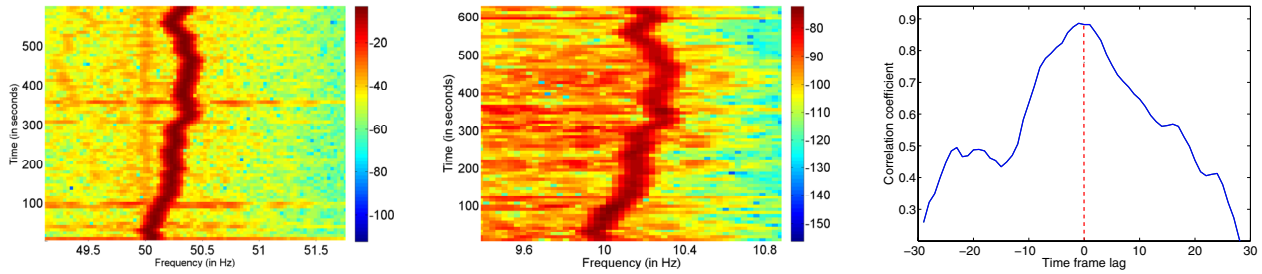
3.1 Aliasing Effect

ENF signals recorded by video cameras under fluorescent lighting undergo significant aliasing due to the lower temporal sampling rate of cameras compared to the 100 Hz/ 120 Hz frequency components present in flickering of fluorescent lights. The ENF signal in the recorded videos appears at different predetermined frequencies, which can be determined from the sampling theorem [7]. As an example, we examine the effect of recording a video under a fluorescent light powered by a 50 Hz source. As current changes polarity at twice the power mains frequency, the fluorescent light flickers at 100 Hz. Additionally, when the power main signal slightly deviates from a perfect sinusoid form, often higher harmonics of decaying energy happen at integer multiples of 100 Hz. The bandwidth of these higher harmonics is also greater than the main component because the practical ENF signal of interest is a narrowband signal and not a perfectly stable sinusoid. So, the bandwidth of the k^{th} harmonic component will be k times the bandwidth of the main ENF component at 100 Hz. Suppose the camera used for capturing video is an NTSC standard camera with a frame rate of 29.97 Hz. Because of the low sampling rate as compared to the required sampling rate to prevent aliasing, the resulting spectra will have periodic tiling of frequency components at $\pm 100 + 29.97k$, $k = 0, \pm 1, \pm 2, \ldots$. Because of the periodic nature of the resulting spectra, it suffices to focus on replicas within one period. Aliasing effect introduced by this video recording example is shown in Fig. 7. From this figure, we observe that multiple copies of the ENF related component will appear in the spectrum of the video signal around different but pre-determined frequencies. These multiple copies arise due to the presence of higher harmonics in the power mains signal, and may be combined strategically to obtain a better estimate of the video ENF signal.

We further recall that the magnitude spectrum of a real valued signal is symmetric about y-axis. Because of this, original spectrum of the ENF signal in fluorescent lights also has symmetric components at -100 Hz and its harmonics, as shown in Fig. 7. After sampling by video camera, the frequency component E, that is at -200 Hz in the spectrum of the original signal, appears at 9.79 Hz in the recorded signal; and the component A present at 100 Hz in the spectrum of the original signal appears at 10.09 Hz. As a result, we obtain replicas of the ENF signal at 9.79 Hz and 10.09 Hz, and they are mirrored versions of each other. Similar analysis can be performed on other combinations of different camera frame rates and power mains frequencies to find the frequencies at which the main component and the first harmonic component of the signal appear. The values of these frequencies are summarized in Table 1.

3.2 Experimental Setup for Video Recordings

In this section, we describe the settings under which experiments are conducted to detect the presence of the ENF signal in videos. We collect video data from three geograph-



(a) Power mains ENF signal (b) Video ENF signal (c) NCC coefficient, $\rho(k)$

Figure 8: (a),(b) ENF fluctuations measured for the whitewall video experiment in India; (c) NCC coefficient, $\rho(k)$, as a function of time frame lag k.

Table 1: Aliased frequency for different combinations of power mains frequencies and video camera frame rates

Mains (in Hz)	Video (fps)	aliased (in Hz)	2nd harmonic aliased (in Hz)
60	29.97	0.12	0.24
60	30	0	0
50	29.97	10.09	9.79
50	30	10	10

ical regions – India, China, and USA – to demonstrate the capability of the ENF signal to act as a timestamp and authentication signal. To the best of our knowledge, this is the first work reporting the ENF based forensic signal collection and analysis on the Asian continent. For each of these three geographic areas, we collect video data under two settings:

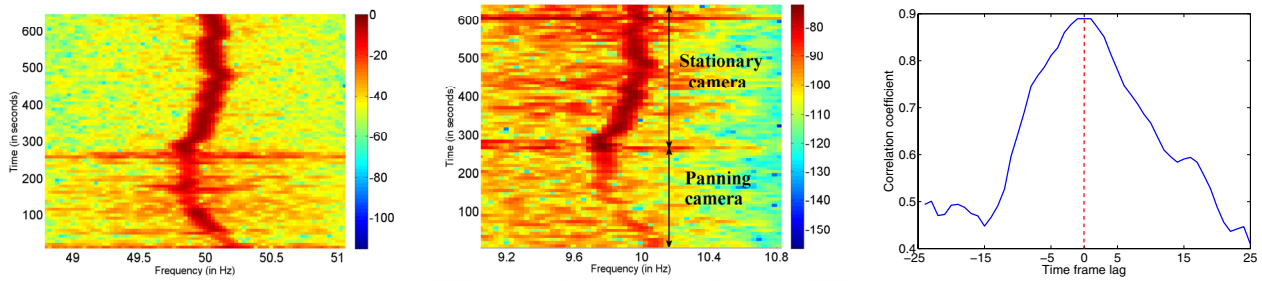
(a) White wall video: In this experiment, we record a constant white wall scene illuminated under fluorescent lights for 10 minutes and examine the video for ENF traces. This experiment can be considered as a first step towards demonstrating the presence of the ENF signal in videos recorded under fluorescent lighting. We find the average intensity of each frame, and pass it through a band pass filter with passband corresponding to the frequencies of interest where aliased components of the ENF signal appear in videos, as listed in Table 1. As the content of the video is mostly the same in each frame, presence of a significant amount of energy in the desired frequency band can be attributed to the ENF signal. A spectrogram is plotted with a window size of 480 video frames ($\approx$ 16 seconds) and an overlap factor of 50%. For all the experiments conducted on video data, a frequency analysis with weighted energy is performed to extract the ENF signal around the aliased frequency bands.

(b) Surveillance video: In this experiment, a camera is placed inside a room for surveillance purposes. A fluorescent light illuminates the room. Occasional movement of people can be seen in the foreground of the video. Recordings performed under this setting can be considered representative of real-life surveillance scenarios in places where few events are expected to occur. In these experiments, because the content in each frame of the video is not constant due to the movement of people, a direct averaging of all frames may not be suitable to perform the frequency analysis. As the ENF signal from fluorescent lights is recorded by each sensor pixel, we can use a part of the video frame to extract the ENF signal. For example, we locate the relatively stationary areas in video frames, where content does not change much. Such areas at the top and bottom areas of each video frame are used to extract the ENF signal based on the spectrogram of average pixel intensity, and the similar settings to those described for white wall video experiments is employed. Alternatively, the process of finding the static regions in frames of the video can be automated using common motion estimation techniques.

3.3 Results and Discussions

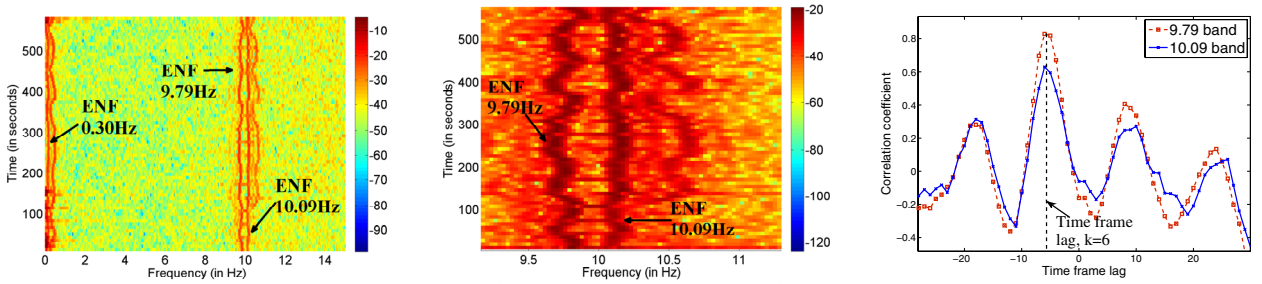
(a) ENF Data collected from India: The nominal value of the ENF in India is 50 Hz, hence, the ENF signal reflected in the flickering of light is expected at a fundamental frequency of 100 Hz. In this experiment, we use a 30fps NTSC Sony HDR SR12 camera to record a video. The video camera recording and the power mains recording start simultaneously. From the aliasing discussion in Sec. 3.1, with this sampling frequency, the ENF signal at 100 Hz will appear at an aliased frequency of 10 Hz in video recordings. Spectrogram of the ENF signal for a white wall video recording is shown in Fig. 8(b). From this figure, we can clearly observe a high correlation between the pattern of the ENF signal at 10 Hz and the ground truth ENF signal at 50 Hz obtained directly from power mains supply, as shown in Fig. 8(a). The maximum value of the NCC coefficient, $\rho(k)$, between the video ENF signal and the power mains signal is found to be 0.91 at $k = 0$, as shown in Fig. 8(c). This indicates that the recording starts simultaneously for the power mains signal and the video signal.

For surveillance video recording experiment, the first five minutes of video is recorded with the camera constantly panning, while the remaining part of the video is recorded with a stationary camera. For this duration of the video, all parts of the frames undergo changes. We show the spectrogram of the ENF signal from this video recording in Fig. 9(b). From this figure, we see that the ENF signal is not strong during the first five minutes of the video recording when the camera pans. Nevertheless, the pattern in the ENF signal around 10 Hz for the first five minutes follows a similar pattern to the ground truth ENF signal shown in Fig. 9(a). The NCC coefficient, $\rho(k)$, attains the highest value at $k = 0$ as shown



(a) Power mains ENF signal (b) Video ENF signal (c) NCC coefficient, $\rho(k)$

Figure 9: (a),(b) ENF fluctuations measured for surveillance video experiment in India; (c) NCC coefficient, $\rho(k)$, as a function of time frame lag k.



(a) Video ENF signal (b) Video ENF signal (amplified view) (c) NCC coefficient, $\rho(k)$

Figure 10: (a),(b) ENF fluctuations measured for the whitewall video experiment in China; (c) NCC coefficient, $\rho(k)$, as a function of time frame lag k.

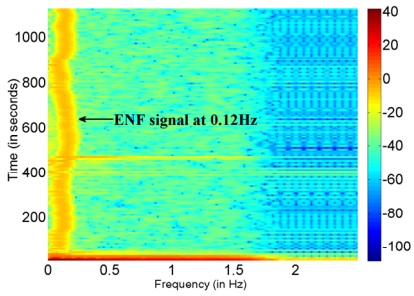
in Fig. 9(c). This indicates that the power mains signal and the video signal recording starts at the same time.

For the above mentioned Sony HDR SR12 camera, videos are stored in MPEG-2 format, which involves lossy compression. The bitrate of recorded videos has a range of 700-750 kbps. Our above experiments also suggest that the ENF signal can be extracted from the lossy compressed videos and can be used to find time of recordings of moderately compressed videos.

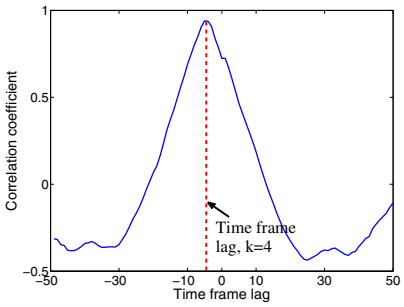
(b) ENF Data collected from China: In China, 50 Hz supply frequency is used across the power distribution network. As a result, the ENF signal in the fluorescent light flickering can be expected at 100 Hz. We use a NTSC standard Panasonic camera with the frame rate of 29.97fps to record videos in China. For this experiment, video recording is started at a delay of 50 seconds with respect to the power mains signal. As discussed previously in Sec. 3.1 and shown in Fig. 7, we expect the main component of the ENF signal in video recordings to be present in the band around the aliased frequency of 10.09 Hz. Higher harmonics of the ENF signal in video recording are present in the band around 9.79 Hz (from the second harmonic) and 0.30 Hz (from the third harmonic).

The spectrogram of the ENF signal for the white wall video recording is shown in Fig. 10(a). In this figure, we observe that multiple copies of the ENF signal are present near the DC frequencies and 10 Hz frequencies. A magnified version of the figure around 10 Hz band is shown in Fig. 10(b), in which we can see multiple copies of the ENF signal around different frequencies in accordance with our analysis in Sec. 3.1. We also observe from this figure that the bandwidth of the video ENF signal present around 9.79 Hz is double that of the bandwidth of the video ENF signal around the frequency band at 10.09 Hz. This can also be explained from Fig. 7 as the copy of the video ENF signal at 9.79 Hz originates from the second harmonic of the ground truth ENF signal, which is present at 200 Hz from the flickering of fluorescent lights.

As explained in Section 3.1, the 9.79 Hz component of the recorded signal is a mirrored version of the 10.09 Hz component. Therefore, we flip the ENF signal extracted from 9.79 Hz component in NCC coefficient calculation to make the overall correlation positive. The plot of NCC coefficient between the video ENF signal and power mains ENF signal evaluated using the video ENF signal present around 10.09 Hz and 9.79 Hz frequency band is shown in Fig. 10(c). The plot obtained using the video ENF signal from 10.09 Hz band is similar to the plot obtained using the ENF signal from 9.79 Hz frequency band. The maximum value of the NCC coefficient is obtained at a lag of $k = 6$ time frames. This indicates that the video recording was started with a delay of approximately 48 seconds from the power mains signal, which is close to the actual delay of 50 seconds. The small difference is due to the window based signal processing that leads to a limit in the temporal resolution for the ENF based timestamping. It is possible to extract other harmonic copies of ENF traces in the video signal by passing the video ENF signal through bandpass filters with passband frequencies around which each harmonic copy is located. These



(a) Video ENF signal



(b) NCC coefficient, $\rho(k)$

Figure 11: (a) ENF fluctuations measured for the whitewall video experiments in US; (b) NCC coefficient, $\rho(k)$, as a function of time frame lag k.

copies of the signal can then be combined to perform correlation analysis to find the time of video recording.

(c) ENF Data collected from USA: The nominal value of the power supply frequency in the United States is 60 Hz. The ENF signal from the flickering of fluorescent lights can be expected at 120 Hz. We use the same 29.97fps Panasonic camera from experiments done in China to conduct video recording in the United States. From Table 1, the ENF signal in the video recordings is expected at aliased frequency of 0.12 Hz. Fig. 11(a) shows the spectrogram of the ENF signal for the white wall video. The presence of a significant energy band at 0.12 Hz confirms the presence of the ENF signal in video recordings. The NCC coefficient between the video ENF signal and the power mains ENF signal is plotted in Fig. 11(b). The maximum value of the NCC coefficient is found to be 0.95 corresponding to a lag of $k = 4$, indicating the video recording was started with a delay of approximately 32 seconds from the power signal. The actual delay in recording time is approximately 38 seconds. This again shows the timestamping capability on the order of approximately 8 seconds.

(d) Tampering detection: We design an experiment to demonstrate an application of the ENF analysis for tampering detection of surveillance videos. For this purpose, we record a video of 980 seconds' long in an indoor environment in India illuminated with fluorescent lights. We perform a simple video clip insertion by cutting the last 160 seconds of video and inserting it in between the remaining clip of the

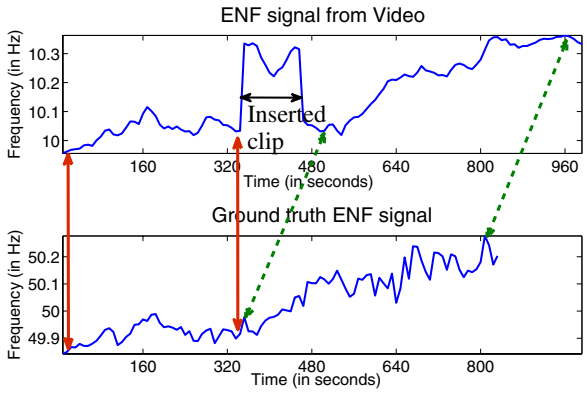
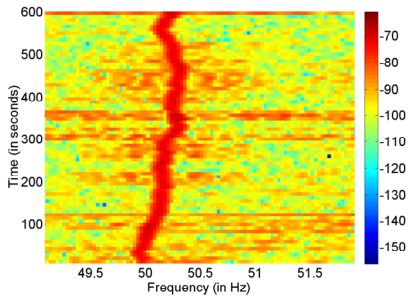


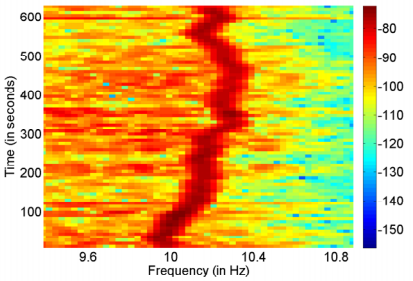
Figure 12: ENF matching result demonstrating the detection of video tampering based on the ENF traces.

video close to the 340^{th} second. The ENF signal extracted from this tampered video and the ground truth ENF signal extracted from the power mains signal database corresponding to the time of recording is shown in Fig. 12. From this figure, we can locate the regions in the ENF signal from the video, which has similar patterns as the ground truth ENF signal for the given time of recording. We observe that the ENF signal from video recording shows the same pattern as the ground truth ENF signal for 340 seconds in the beginning of the video as shown by the red arrows in the figure. Similarly, video ENF signal from time index 500 second to 980 seconds have the same pattern as the ground truth ENF signal between time index 340^{th} and 820^{th} seconds. The video ENF signal between time index 340^{th} seconds and 500^{th} seconds does not match with the ground truth ENF signal, suggesting that the video recording is likely to be modified in the corresponding region and a video clip insertion operation was performed. More rigorous verification can be performed over visual examination by quantitatively comparing the segments of ENF sequence from video with the reference ENF database.

(e) Audio-visual authentication and synchronization: We now explore an important application of the ENF analysis for *forensic binding* of audio-visual recording. This determines whether the audio and visual tracks in a video recording are captured together, or if the sound track from a different recording is manually placed with the visual track. ENF signals extracted from both visual and audio tracks can provide natural alignment and binding of audio-visual recordings. As noted above, the ENF signal is captured in the video track due to the flickering of fluorescent lights and in the audio track through electromagnetic interference. When audio and video are recorded simultaneously, we expect that the ENF traces in these two streams should exhibit similar patterns, and the synchronization between the audio and video tracks can be verified by quantitatively comparing these two signals. Discrepancies between the ENF traces in the audio and video would provide a strong indicator of tampering of either the audio or the video. Thus, the synchronization can be authenticated even if the ground truth ENF signal from power mains is not available at the time of recording. Fig.13 shows an example of the ENF traces obtained from the audio and visual tracks of a video record-



(a) ENF signal from the audio track



(b) ENF signal from the video track

Figure 13: ENF matching between audio and video tracks of a video recording.

ing. From this figure, we see that the ENF signal captured in audio, as shown in Fig.13(a) exhibit similar patterns and have high correlation with the ENF signal captured in visual track, as shown in Fig.13(b). Further quantitative cross correlation analysis can confirm with confidence that the audio-visual tracks in the video are temporally synchronized and authentic.

4. CONCLUSIONS & FUTURE WORKS

In this paper, we have described mechanisms to capture the ENF fluctuations in fluorescent lighting using optical sensors and video camera, and demonstrated the presence of the ENF in indoor video surveillance recordings. Recorded optical and video signals have been shown to have high correlation with the ENF signals captured directly from the power mains supplies. The estimated time of recordings for sensor signals have been shown to be accurate up to 8 seconds. Results from our investigations suggest that the ENF signal can be used as a natural timestamp for optical sensors and video surveillance recordings in an indoor environment. The ENF signal can also be used to facilitate the authentication of sensor and video data, as data modifications by adversaries are likely to cause changes in the ENF signal at the corresponding part of the recordings. Modifications in video recordings can be identified by the discontinuity or alteration in the ENF signal and the ENF traces provides a natural binding of audio and visual tracks to verify their temporal synchronization and integrity. To the best of our knowledge, this is the first work investigating the presence and forensic analysis of the ENF signal in optical and visual recordings, and reporting the ENF based forensic signal collection and analysis in the Asian continent.

In future work, we plan to devise methods to adapt the ENF analysis technique to non-static videos, where the camera itself moves. In such videos, it may be difficult to find any part in the frame that is stationary throughout the video, calling for the need to adapt ENF authentication techniques to non-static video. Furthermore, we are collecting more experimental data to prepare for a larger scale statistical study on tampering detection in video and audio-video synchronization.

5. ACKNOWLEDGEMENTS

The work was supported in part by research grants from U.S. AFOSR #FA9550-09-1-0179 and NSF #0824081, and a University of Maryland Litton Graduate Fellowship. The authors would like to thank Prof. Xianli Wu and Mr. Hui Su for their help in collecting the ENF data in China, and Mr. Daniel Garcia-Romero for recommending references [3] [10]. The authors would also like to thank the anonymous reviewers and the area co-chairs, Dr. Qibin Sun and Dr. Tian-Tsong Ng, for their constructive comments.

6. REFERENCES

[1] A. Bovik. *Handbook of Image and Video Processing*. Academic Press, 2000.

[2] S. Chen, C. Lo, M. Foo, and K. How. Testing of fluorescent lamps for its flickering susceptibility towards interharmonic voltages. In *International Power Engineering Conference*, December 2007.

[3] C. Grigoras. Applications of ENF criterion in forensics: Audio, video, computer and telecommunication analysis. *Forensic Science International*, 167(2-3):136 – 145, April 2007.

[4] S. Haykin. *Advances in Spectrum Analysis and Array Processing*, volume 1. Prentice-Hall, 1991.

[5] W. Hernandez. Input-output transfer function analysis of a photometer circuit based on an operational amplifier. *Sensors*, 8(1):35–50, September 2008.

[6] M. Huijbregtse and Z. Geradts. Using the ENF criterion for determining the time of recording of short digital audio recordings. In *Proceedings of the 3rd International Workshop on Computational Forensics*, IWCF '09, pages 116–124. Springer-Verlag, August 2009.

[7] A. V. Oppenheim, R. W. Schafer, and J. R. Buck. *Discrete-time Signal Processing (2nd ed.)*. Prentice-Hall, Inc., 1999.

[8] L. Rabiner and B.-H. Juang. *Fundamentals of Speech Recognition*. Prentice-Hall, Inc., 1993.

[9] D. Rodriguez, J. Apolinario, and L. Biscainho. Audio authenticity: Detecting ENF discontinuity with high precision phase analysis. *IEEE Transactions on Information Forensics and Security*, 5(3):534 –543, September 2010.

[10] R. W. Sanders. Digital authenticity using the electric network frequency. In *33rd AES International Conference on Audio Forensics, Theory and Practice*, June 2008.