Invisible Geolocation Signature Extraction From a Single Image

This paper was downloaded from TechRxiv (<https://www.techrxiv.org>).

LICENSE

CC BY-SA 4.0

SUBMISSION DATE / POSTED DATE

25-01-2022 / 01-02-2022

CITATION

Choi, Jisoo; Wong, Chau-Wai; Hajj-Ahmad, Adi; Wu, Min; Ren, Yanping (2022): Invisible Geolocation Signature Extraction From a Single Image. TechRxiv. Preprint. <https://doi.org/10.36227/techrxiv.18941414.v1>

DOI

[10.36227/techrxiv.18941414.v1](https://doi.org/10.36227/techrxiv.18941414.v1)

Invisible Geolocation Signature Extraction From a Single Image

Jisoo Choi, *Graduate Student Member, IEEE*, Chau-Wai Wong, *Member, IEEE*, Adi Hajj-Ahmad, *Member, IEEE*,
Min Wu, *Fellow, IEEE*, and Yanpin Ren, *Member, IEEE*

Abstract—Geotagging images of interest are increasingly important to law enforcement, national security, and journalism. Today, many images do not carry location tags that are trustworthy and resilient to tampering; and the landmark-based visual clues may not be readily present in every image, especially in those taken indoors. In this paper, we exploit an environmental signature from the power grid, the electric network frequency (ENF) signal, which can be inherently captured in a sensing stream at the time of recording and carries useful time–location information. Compared to the recent art of extracting ENF traces from audio and video recordings, it is very challenging to extract an ENF trace from a single image. We address this challenge by first mathematically examining the impact of the ENF embedding steps such as electricity to light conversion, scene geometry dilution of radiation, and image sensing. We then incorporate the verified parametric models of the physical embedding process into our proposed entropy minimization method. The optimized results of the entropy minimization are used for creating a two-level ENF presence–classification test. It identifies whether a single image has an ENF trace; if yes, whether it is at 50 or 60 Hz. We quantitatively study the relationship between the ENF strength and its detectability from a single image. This paper is the first comprehensive work to bring out a unique forensic capability of environmental traces that shed light on an image’s capturing location.

Index Terms—Geotagging, Electric Network Frequency (ENF), Frequency Estimation, Entropy Minimization, Rolling Shutter.

I. INTRODUCTION

This digital era has witnessed an unprecedented amount of digital audio, images, and videos being generated daily. Digital media contain many metadata fields, one of which can relate to the location information using the built-in GPS input, which is valuable for law enforcement, national security, and journalism. This piece of metadata, however, is not always available and can be easily tampered with. In this paper, we focus on deriving information from a single image regarding the location at which it was captured, especially in challenging scenarios that existing technologies cannot sufficiently address.

Recent advances, notably through the Finder program by the U.S. agency for the Intelligence Advanced Research Project Activity [2], utilized computer vision techniques and a big data paradigm to exploit visible terrains and landmarks appearing in an image and match them with various geospatial data to identify the image’s geographic location. These technologies would face considerable challenges in the absence of visible landmarks, such as when the background of outdoor images does not have differentiating features or when an image is

captured indoors. The latter scenarios are common in such global efforts as fighting crime on child exploitations [3]–[5], where a significant portion of images of such mistreatment and abuse are taken indoors in private locations by perpetrators. The anonymous nature of the internet makes it difficult to track down the perpetrators. Having new technologies to locate where such images were taken, even at a coarse level, will provide an unprecedented tool to aid global investigations, supporting law enforcement in appropriate jurisdictions.

In this work, we explore an unnoticeable location-related signature that may be inherently captured in an image. This signature comes from the power distribution network whose varying frequency over time is defined as the electric network frequency (ENF) signal [6]–[8]. The ENF signal over time in a multimedia recording reflects the behavior of the power grid at the time and the location where the media recording was captured. Such a relation enables the estimation of the time, location, and integrity of corresponding multimedia recordings [9]–[12].

Due to the temporally varying nature of the ENF that plays the critical role in determining the time and the location of a recording, ENF extraction in the literature thus far requires the hosting signal to have a temporal nature, e.g., an audio or a video. Specifically, to produce one frequency estimate, audio/video recordings need a signal of several seconds long. Images captured by a rolling shutter also have a temporal dimension, but it is very short, i.e., 0.02–0.03 seconds long. Hence, a series of intriguing questions to push the boundary of ENF’s localization capability are: Can the ENF traces be captured in a single image? If yes, can we determine the presence of ENF traces? If a trace is present, can we narrow it down to 50 vs. 60 Hz? This ENF’s localization capability may be a valuable clue to law enforcement investigations and investigative journalism. To the best of our knowledge, this is the first comprehensive work (building on our preliminary conference paper [1]) to develop a unique forensic capability to geotag individual images that do not have a visible landmark or GPS tag, especially those captured indoors.

This paper begins with characterizing the physical embedding process for the ENF by modeling the steps of electricity to illumination conversion, scene geometry dilution of radiation in controlled experiments, and image sensing. Such effort leads to a more precise parametric model than that in our preliminary version [1]. The new model contains time-varying amplitude and trend for embedded ENF traces, which better reflects the realistic scenario. Using the parametric form of the embedded ENF traces and an entropy-related statistical phenomenon of

indoor photos, we approach the research problem of detecting and reconstructing ENF traces in an image as an optimization problem and name it as the *entropy minimization method*. The optimization results will be used for a two-level ENF presence–classification test in which the first level determines whether an embedded ENF trace is present in a test image; if it is, the second-level test determines whether the images come from a 50 or 60 Hz region.

In addition to the proposed entropy minimization method in conjunction with a simple parametric model for the embedded ENF traces presented in our preliminary work [1], we summarize the main contributions of this journal version as follows:

- We model the physical process of ENF embedding that comprises the steps of the light generation, the geometry dilution of radiation, and the image sensing. We also show experimentally and via simulation that ENF traces in camera captured photos may have time-varying amplitude and trend.
- We propose a more precise parametric model than that of [1] for embedded ENF traces based on insights of the physical embedding process, allowing the entropy minimization method to extract more precise ENF traces.
- We design a two-level ENF presence–classification test that produces three decision outcomes: no ENF, ENF at 50 Hz, and ENF at 60 Hz.

The rest of this paper is organized as follows. Section II provides background and preliminaries of the ENF and image acquisition principles of rolling-shutter cameras. Section III investigates the physical embedding process for ENF traces, and Section IV presents our improved parametric model for ENF trace extraction from a single image and the two-level ENF presence–classification test. Section V shows the experimental results. Section VI discusses a few technical points. Finally, Section VII concludes the paper.

II. BACKGROUND AND PRELIMINARIES

A. Electric Network Frequency (ENF)

The electric network frequency (ENF) of power distribution networks has a nominal value of 60 Hz in most of the Americas and 50 Hz in most rest of the world. The instantaneous frequency of the sinusoidal supply signal of a power network does not stay at the nominal value. Instead, it fluctuates around the nominal value due to the mismatch between the power supply and demand across the power grid. The trends in the ENF variations tend to be very similar at different points interconnected in the same grid. The changing instantaneous value of the ENF over time is what we define as the *ENF signal*.

The ENF signal has been a growing subject of multimedia forensics research due to its presence in media sensing signals. If an audio or video recording is made near an electrical power source, the recording may capture the ENF variations at the time of recording. In audio recordings, this has been mainly attributed to the acoustic hum emitted from devices connected to the power mains [13]–[15]. In video recordings, this comes from the invisible flickering of electric lighting [14], [16]–[19].

TABLE I
ESTIMATED PARAMETERS FOR ROLLING-SHUTTER IMAGES.

	iPhone 6s	iPhone 6	iPhone 5
T_{ro} (ms)	19.8	30.9	22.6
number of rows, M	3024	2448	2448
f_{row} (kHz)	152.7	117.1	108.3

This paper will show that it is possible to extract ENF traces captured in a single image taken in electrically lighted settings. In particular, the ENF value can be estimated from an image captured by a camera equipped with widely available complementary metal–oxide–semiconductor (CMOS) image sensors that use the *rolling-shutter* mechanism.

B. Rolling Shutter and Read-Out Time

Unlike cameras with a *global shutter* that simultaneously acquires all pixels of an image frame, a camera with a rolling shutter captures an image frame one row at a time. Although traditionally regarded as the cause of negative artifacts in images/videos, the rolling shutter has been exploited in beneficial ways for computational photography applications [20], [21]. This work shall show that the rolling shutter’s sequential read-out time mechanism allows the resulting image to capture samples of an incident light signal at slightly different time instants and thus obtain a short segment of ENF traces. Studies on ENF from audio [11], [15] and videos [16], [19], [22], [23] show that one frequency estimate of audio and video recordings requires a signal of 5–10 and 12–20 seconds long, respectively. In contrast, the available signal length allowed by a single rolling-shutter image is only about 0.02–0.03 seconds, corresponding to about 2–3.6 cycles that are not long enough to observe ENF fluctuations.

The amount of time for a rolling shutter to acquire the rows of an image frame, which we denote by the read-out time T_{ro} , is specific to the camera’s model line. The read-out time is not typically given in its user manual or technical specifications. To extract ENF information buried in a single image, we may need to know the read-out time T_{ro} value of the camera that had produced the image under question. In our work, we use a protocol inspired by the one proposed in [24], [25] to estimate the T_{ro} value of a camera at hand. TABLE I lists the estimated values we calculated for the cameras on the backsides of different models of iPhone devices.

C. Image Acquisition by Rolling-Shutter Cameras

The rolling-shutter mechanism equivalently turns rows of the CMOS sensor array of a camera into high-frequency samplers. The sampling frequency of a row can be expressed by $f_{row} = M/T_{ro}$, where M is the number of rows of the image. A CMOS camera captures ENF traces by recording a visual scene under an electric light source. The electric light intensity embedded in the resulting image relates to the supply current via the power law, making its nominal frequency twice the nominal ENF value, i.e., 120 Hz in most of the Americas and 100 Hz in most other parts of the world. By considering the acquisition at different row indices i as samples of a fluctuating

Fig. 1. Photos of (a) a piece of printed paper on a desk, (b) an opened magazine on a desk, and (c) a white paper, all with visible ENF traces. (d) Averaged column signals for (a)–(c). The fluctuation of the averaged column signals is mainly contributed by the ENF traces at the time of recording. (e) The sinusoidal parts of the ENF traces for (a)–(c) estimated by the entropy minimization algorithm proposed in Section IV-D.

signal, the ENF trace captured in an image is modeled in our preliminary work [1] as

$$e(i) = A \cos\left(2\pi \frac{f}{f_{\text{row}}} i + \phi\right), \quad (1)$$

where f is set to be a constant of 100 or 120 Hz considering that the frequency changes over the a few cycles is negligible. Note that A is the magnitude, i is the row index of the image in $\{1, \dots, M\}$, and ϕ is the initial phase of ENF trace immediately before acquiring the first row.

Our preliminary work [1] that uses the oversimplified parametric model in (1) showed good performance in images with ENF traces that are synthetically added but showed limited performance in images with real ENF traces. Thus, to improve the model fidelity and, ultimately, the accuracy of the ENF estimation, we begin with investigating the physical embedding process in Section III for a more accurate parametric model of ENF traces embedded in an image.

III. PHYSICAL EMBEDDING PROCESS FOR ENF TRACE

This section will show that the amplitude and bias of embedded ENF traces are not constant but time-varying and investigate what causes the time-varying artifacts. The investigation will be conducted through examining modeling the physical ENF embedding process that includes the light generation, the geometry dilution of radiation, and the image sensing.

Fig. 1(a)–(c) show three scenes corrupted by stripes of ENF traces, in which Fig. 1(a) and Fig. 1(b) have nonuniform foreground and almost uniform background and Fig. 1(c) has a scene for a piece of white paper without foreground object. Fig. 1(d) shows their *averaged column signals*, each of which is obtained by averaging pixel intensities across all columns of an image. Fig. 1(e) shows the sinusoidal part (excluding the trend/biases) of the ENF traces for Fig. 1(a)–(c) estimated by the entropy minimization algorithm proposed in Section IV-D. Each sinusoidal part of the estimated ENF trace exhibits a similar sinusoidal shape as the averaged column signal, evidenced by the well aligned peaks and valleys. One can notice that the overall trend in each averaged column signal is sinusoidal-like, implying an impact from the ENF. Besides, the scene contents of Fig. 1(a) and Fig. 1(b) contribute to biases that alter the sinusoidal shape and result in small teeth of the averaged column signals, making the mixture of the scenes content and the ENF-induced fluctuation nonsinusoidal. Unlike Fig. 1(a) and Fig. 1(b), Fig. 1(c) has no scene content,

but shows a slightly decreasing trend along the row index, which makes its averaged column signal a nonideal sinusoid as well.

A. ENF Embedding into Photos Under Indoor Lighting

In this subsection, we analyze the physical ENF embedding process that primarily consists of light-generation and pixel-sensing processes. We approximate the supply voltage signal to be an ideal sinusoidal wave to obtain analytic expressions and later use simulation from real supply-voltage measurements to demonstrate that ENF traces embedded in an image have a time-varying behavior.

The voltage to power conversion is illustrated in the first block of Fig. 2. Here, we use a simplified model, and a more realistic scenario is investigated in Appendix A of the supplementary materials and discussed in Section VI-A. When a source light is powered by an AC voltage signal $v(t) = V_l \cos(2\pi f_0 t + \phi)$ with a voltage amplitude $V_l \in \mathbb{R}^+$, a frequency $f_0 \in \{50, 60\}$ Hz, and an arbitrary initial phase ϕ_0 , the instantaneous power has twice the frequency of the supply voltage, namely,

$$p_l(t) = v^2(t)/R \quad (2a)$$

$$= V_l^2 \cos^2(\omega_0 t + \phi)/R \quad (2b)$$

$$= V_l^2 [\cos(2\omega_0 t + 2\phi) + 1]/2R, \quad (2c)$$

where $\omega_0 = 2\pi f_0$ is the angular frequency of power mains. Note that we assume that only resistive loads R are present in the circuit, but the result can be extended to capacitive and inductive loads. It is also noteworthy that the ratio of the AC component to the DC component of the supply power equals one. The model presented in Appendix A of the supplementary materials is more realistic and the model presented here is idealized. Regardless of the models, the frequency component $2\omega_0$ will be dominant in the final camera-captured image.

The circuit load converts the electricity into the illumination with an equivalent lowpass filter, which leads to attenuation in the relative strength of the AC component. The power to illumination conversion is illustrated by the second block of Fig. 2 formulated by a convolution of the power and a single exponential decaying function [26], [27] as follows:

$$\ell(t) = K_l p_l(t) * e^{-\frac{t}{\tau}} u(t), \quad (3)$$

where “ $*$ ” denotes the linear convolution, τ is a time constant accounting for the illumination half-life, and $u(t)$ is a unit step function with value 1 for $t \geq 0$ and 0 for elsewhere. Note that

Fig. 2. An illustration of the ENF embedding process. When a light source is powered by a supply voltage $v(t)$ with frequency ω_0 , the instantaneous power $p_l(t)$ is converted into the illuminance of the light $\ell(t)$, which has a DC component and an AC component flickering at $2\omega_0$ rad/s. The visual light then falls onto an object and interacts with the object's surface. The camera acquires the light reflected or reemitted from the surface and creates a visual image by collecting photons during the exposure time. This photons accumulation process results in a sampled and quantized temporal pixel intensity signal $I(t)$ whose intensity fluctuates at $2\omega_0$ rad/s.

K_l is a scalar with a unit of $\text{lm} \cdot \text{W}^{-1} \cdot \text{m}^{-2}$ that relates the power to the illumination of the light source. The exponential decaying function acts as a lowpass filter, converting the power signal into a more smoothed illumination. Hence, the resulting luminance signal $\ell(t)$ carries a DC component and a relatively smaller flickering AC component.

The third block of Fig. 2 illustrates a geometric setup in which rays from a light source fall onto an object, reflect off, and enter a camera's imaging system. During the light transmission and reflection, its intensity degrades per the *inverse-square law* and surface albedo of the object. The camera prepares to acquire the light reflected from and/or reemitted by the object before forming a visual image of the object. The overall effect due to the geometric setup is a nonuniform spatial scaling to the magnitude of the visual light, yet it does not change the AC-to-DC ratio for the light arriving at the camera lens. We, therefore, defer a more detailed analysis of the geometric setup to Section III-B and focus on the image sensing stage that will further reduce the AC-to-DC ratio. To this end, we introduce a spatial-varying discounting factor $\alpha(\mathbf{p}) < 1$ at location $\mathbf{p}$ of the shooting object to capture light energy loss after being reflected and during transmission caused by the geometric setup.

The incoming light focused by the lens is further processed by the camera's sensing unit to create digital images. Photons of the incoming light emitted from a small region in the scene arrive at an active cell of the sensor array. These photons are converted into electrons and then into a voltage level representing the intensity of a pixel of a digital image. Ignoring the complication of color, the intensity of the pixel is proportional to the number of photons of light landing on the sensor cell. The photons are collected during the camera exposure time t_{exp} , which can be modeled as a convolution of a rectangular window to the perceived light signal as

follows [12]:

$$I(t) = \beta \int_{t-t_{\text{exp}}}^t \alpha(\mathbf{p}) \ell(\xi) d\xi \quad (4a)$$

$$= \beta \alpha(\mathbf{p}) \ell(t) * [u(t) - u(t - t_{\text{exp}})] \quad (4b)$$

$$= \beta \alpha(\mathbf{p}) K_l p_l(t) * e^{-\frac{t}{\tau}} u(t) * [u(t) - u(t - t_{\text{exp}})] \quad (4c)$$

$$= I_e(t) + I_{\text{dc}}, \quad (4d)$$

where β is a scalar that relates the illuminance to the pixel intensity values. The pixel intensity $I(t)$ is decomposed into two components: a time series $I_e(t)$ capturing the flicker of pixel intensity along the time and a constant term I_{dc} capturing the offset of pixel intensity. This camera integration process is a second lowpass filter that further reduces the AC-to-DC ratio.

To derive an analytic form of the relative strength between $I_e(t)$ and I_{dc} , we apply the Fourier transform to the intensity signal in (4c) and obtain:

$$S(\omega) = \mathcal{F}[I(t)] = \beta \alpha(\mathbf{p}) K_l \cdot \mathcal{F}[p_l(t)] \cdot \mathcal{F}\left[e^{-\frac{t}{\tau}} u(t)\right] \cdot \mathcal{F}[u(t) - u(t - t_{\text{exp}})], \quad (5)$$

where

$$\mathcal{F}[p_l(t)] = \frac{\pi V_l^2}{2R} \left\{ 2\delta(\omega) + e^{j\frac{\omega}{\omega_0}\phi} [\delta(\omega - 2\omega_0) + \delta(\omega + 2\omega_0)] \right\}, \quad (6a)$$

$$\mathcal{F}\left[e^{-\frac{t}{\tau}} u(t)\right] = \frac{1}{1/\tau + j\omega}, \quad (6b)$$

$$\mathcal{F}[u(t) - u(t - t_{\text{exp}})] = t_{\text{exp}} \cdot e^{-j\frac{\omega t_{\text{exp}}}{2}} \cdot \text{sinc}\left(\frac{\omega t_{\text{exp}}}{2\pi}\right), \quad (6c)$$

$$\text{sinc}(x) = \begin{cases} 1 & , \quad x = 0, \\ \frac{\sin(\pi x)}{\pi x} & , \quad \text{otherwise.} \end{cases} \quad (6d)$$

Using (6), we calculate the ratio between the AC and DC parts of the pixel intensity:

$$\frac{|S(2\omega_0)|}{|S(0)|} = \frac{|\text{sinc}(\omega_0 t_{\text{exp}}/\pi)|}{2\sqrt{1^2 + (2\tau\omega_0)^2}}, \quad (7)$$

which implies that the relative strength, also known as *modulation index* [28] in the research field for flickering light's

Fig. 3. Illustrations for supply voltage $v(t)$ measured from a power outlet in the USA, simulated instantaneous power $p_l(t)$, simulated light illuminance $\ell(t)$, and simulated pixel intensity $I(t)$ captured by a camera. Note that both illuminance $\ell(t)$ and pixel intensity $I(t)$ resulted from real measurement $v(t)$ have time-varying behaviors, which may be caused by $v(t)$'s the varying frequency or the varying amplitude. [$p_l(t)$ was scaled, and $\ell(t)$ and $I(t)$ were shifted and scaled in the vertical direction to better visualize their variations in the presence of $v(t)$.]

impact on vision, is controlled by the nominal ENF ω_0 , the time constant τ , and the exposure time t_{exp} .

To model more precise ENF traces, we use our analytic pipeline and simulate the pixel intensity signal using an experimentally measured AC voltage source signal as input. We directly measured the supply voltage $v(t)$ from a power outlet in the USA. Fig. 3 reveals that the practical AC voltage is a nonideal sinusoid with varying frequency and varying amplitude. We subsequently simulated the instantaneous power $p_l(t)$, the illuminance of the light $\ell(t)$, and the resulting pixel intensity $I(t)$ using (2a), (3), and (4b) by setting $R = 10 \, \Omega$, $K_l = 1 \, \text{lm} \cdot \text{W}^{-1} \cdot \text{m}^{-2}$, $\alpha(\mathbf{p}) = 1$, $\beta = 1$, $\tau = 1.9 \, \text{msec}$, and $t_{\text{exp}} = 1/165 \, \text{sec}$, respectively, and display them in Fig. 3. It is observed that the upper and lower envelopes of the simulated camera-captured pixel intensity $I(t)$ slightly change over time, which implies that ENF traces embedded in an image can have a time-varying behavior. Such behavior may result from either the varying frequency and/or the varying amplitude of the AC voltage $v(t)$. In Appendix B of the supplementary materials, we further investigate the causes by using simulated voltage signals of different characteristics and reveal that both the varying frequency and the varying amplitude result in a time-varying behavior in the pixel intensity signal, as shown in Fig. 4(a). Also, we shall see that the parameterization of the amplitude and the bias of image-embedded ENF traces over a short duration for temporally sampling the rolling-shutter image as displayed in Fig. 4(b) can be simplified using affine functions.

B. Trends of Amplitude Due to Camera-Scene Geometry

Fig. 5(a) illustrates a geometric setup in which the visual light reaches an object's surface, the incident light is reflected, and the reflected light is to be digitized by a camera imaging system. Given location $\mathbf{p}$, the intensity of the reflected light

Fig. 4. (a) Three simulated pixel intensity signals $I(t)$ using supply voltages $v(t)$ with constant frequency and constant amplitude (ConFreq-ConAmp), varying frequency and varying amplitude (VarFreq-VarAmp), and varying frequency and constant amplitude (VarFreq-ConAmp). (b) Zoomed-in segments of the simulated $I(t)$ of a length allowed by a single image's acquisition duration driven by VarFreq-ConAmp and VarFreq-VarAmp supply voltages. [$I(t)$'s of the VarFreq-VarAmp and the ConFreq-ConAmp were shifted in the vertical direction to better visualize their variations.]

arriving at the camera sensor depends on the distances of the reflected point of interest to the light source d_p and to the camera d'_p as per the inverse-square law. Each pixel in images may have a different pair of (d_p, d'_p) , making it possible to have different amplitudes of the ENF trace at different pixel locations.

The amount of light reflection can be calculated using the diffuse and specular geometric reflection models that have been widely utilized in computer vision and computer graphics [29], [30]. The diffuse reflection corresponds to the scenario that the reflected light is uniformly scattered in all directions, and its strength depends on the angle formed by the directions of the incident light and the surface normal. The specular reflection corresponds to the scenario that the light is reflected as if it were bouncing off a mirrored surface, and its strength depends on the angle formed by the directions of the viewing/sensing angle and reflected light. In our work, we assume perfect diffuse reflection and ignore specular reflection as most reflections are dominantly diffuse.

Considering the light intensity degradation due to the inverse-square law and the diffuse reflection, we model the intensity of the reflected light arriving at the camera as a function of row i of an image as

$$L_i \propto \frac{\rho \cos \theta_i}{d_i^2 d'_i{}^2}, \quad (8)$$

where ρ is the albedo parameter of the surface, d_i is the distance between the light source and the object, d'_i is the distance between the camera and the object, and θ_i is the angle between the surface normal vector and the incident light direction.

The attenuation and reflection factors considered in (8) will allow us to analytically show that the amplitude of the ENF trace is approximately linear in terms of the row index i of an image due to the small-scale assumption of the captured scene. To analytically show the relationship between the light intensity L_i and the row index i , we consider a special case that the camera is in the same plane as that formed by the incident light and the surface normal vector. Further, we use a simplified 2-d scenario in which pixels at the same row

Fig. 5. (a) An illustration for a geometric setup with a camera and its shooting object under a point light source. A point with location p is d_p and d'_p away from the point light source and the camera, respectively, with the incident light angle θ_p . (b) Given the point light source at $(0, h)$, the incident light reaches the object's surface of locations $(b, 0)$ and $(b + u, 0)$ with incident angles θ_0 and θ_i , respectively. The distances from the light source to locations $(b, 0)$ and $(b + u, 0)$ are d_0 and d_i , respectively. After reflected by the object's surface, the light passes through the lens of the camera located at $(b + b', h')$ that is d'_0 and d'_i away from two locations of interest, respectively. (c) Overlaid averaged column signals (top) and detrended averaged column signals (bottom), where the linear trend is removed from the corresponding averaged column signal to visualize the envelope of the amplitude. The overall trend and amplitude of the averaged column signals decrease approximately linearly with the row index.

have the same impact of the inverse-square law and the diffuse reflection, as shown in Fig. 5(b). Given a point light source at $(0, h)$, the incident light reaches the object's surface of locations $(b, 0)$ and $(b + u, 0)$ with incident angles θ_0 and θ_i , respectively. The distances from the light source to locations $(b, 0)$ and $(b + u, 0)$ are defined to be d_0 and d_i . After hitting the object's surface, the reflected light goes through the camera located at $(b + b', h')$ that is d'_0 and d'_i way from each location of interest.

Appendix C of the supplementary materials analytically shows that the light intensity L of location $(b + u, 0)$ picked up by the camera has a linear relationship with the row index i , which indicates the approximate linear increase/decrease of the amplitude of the ENF trace across the row. To empirically confirm this theoretical result, we captured a set of images under a fluorescent lamp using an iPhone 6s camera in an image acquisition environment similar to Fig. 5(b). We used a piece of blank paper to ensure a relatively uniform color distribution over the entire scene. To boost ENF traces estimation accuracy, we averaged pixel intensities across all columns, which can increase the SNR in the linear scale by a factor of the number of columns. As a result, the trend of ENF traces at the time of capturing the images can be well revealed. Fig. 5(c) shows overlaid averaged column signals at the top and detrended averaged column signals at the bottom, where the linear trend in the top plot is removed from the corresponding averaged column signal to visualize the envelope of the amplitude. Our controlled experiment confirms that the overall trend and amplitude of the averaged column signals decrease approximately linearly with the row index due to the geometry in Fig. 5(b), where the camera is positioned in the same plane as that formed by the incident light and the surface normal vector. This result, together with the result from the previous subsection, allows us to model the amplitude and bias of the sinusoidal-like ENF trace using affine functions, as is detailed in the next section.

IV. PROPOSED METHOD FOR ENF PRESENCE DETECTION AND CLASSIFICATION

To address the challenging geotagging problem, we ask the following research questions: *Can we determine whether an*

ENF trace is present in an image? If yes, can we reveal if the ENF trace is at 100 or 120 Hz? Lastly, what is the lowest amplitude of the embedded ENF trace for it to be correctly identified? To answer these questions, we propose a method that exploits: *i)* the domain knowledge presented in Section III that the image-embedded ENF trace has a parametric form and *ii)* a largely predictable change in the statistical behavior of the image before and after ENF embedding.

A. Phenomenon of Entropy Increase Due to ENF Embedding

We observe that smooth regions in images tend to become busier or have higher entropy after being corrupted by real-valued sinusoidal signals. In the case of an image column with constant intensity, the sole bin of the histogram before corruption starts to split into multiple bins due to positive and negative additive values contributed by the sinusoidal signal. Such corruption increases the entropy of the constant signal from zero to a positive value. In the case of an image column with linearly increasing intensity values, the similar statistical behavior can be observed even though the increase is not as drastic as the constant intensity case. The histogram of the linearly increasing intensity case before corruption has a rectangular shape. After corruption, the bins at two sides of the rectangular shape will split, which may increase entropy.

We verify the observed statistical behavior claimed above via simulation. First, we simulated the ENF embedding process by adding real-valued sinusoidal signals at 100 or 120 Hz with ten randomly picked parameter configurations to columns of images. Then, we calculated the estimated entropy increase for each column and draw the histogram of entropy increase in Fig. 6(a). It reveals that 99% of all columns have an increase in entropy after being corrupted by sinusoidal-like signals with the understanding that entropy values of texture regions of images may not be sensitive to the additive sinusoidal-like signals.

B. Parameterization of ENF Traces

The theoretical analyses and the controlled experiments in Section III confirm that an ENF trace in images can have time-varying amplitude and bias. This allows us to generalize the simple model (1) proposed in the preliminary version [1]

Fig. 6. (a) Histogram of column's entropy increase due to ENF trace embedding. 99% of all columns have an increase in entropy after being corrupted by sinusoidal signals. (b) 2-d error surfaces for the following three combinations: (a_s, a_t) , (a_s, ϕ) , and (a_t, ϕ) with the rest of the parameters set to be the true values. More bluish or darker color indicates a lower error. Red stars on each error surface represent cost values at the true parameter. Cost functions are generally well-behaved around the true parameters and with slightly biased minimum positions.

of this work beyond constant amplitude A and zero bias. Specifically, in Appendix B of the supplementary materials, we have attributed both the ups and downs of the pixel intensity signal to the varying amplitude and varying frequency of the supply voltage. For rolling-shutter captured images typically with 2–3.6 sinusoidal cycles, a sinusoidal model with linearly changing amplitude and bias can adequately track the change of the pixel intensity. In Section III-B, we have derived and experimentally verified that the amplitude of light intensity arriving at a camera is approximately linearly changing as a function of the row index per the diffuse illumination model and the inverse-square law. Based on these newly obtained evidences, we propose the following sinusoidal function with affine amplitude $A(i)$ and affine bias $B(i)$ to model the ENF trace:

$$e(i) = A(i) \cos\left(2\pi \frac{f}{f_{\text{row}}} i + \phi\right) + B(i), \quad i \in [i_s, i_t], \quad (9)$$

where i is the row index, and i_s and i_t are the indices of the first row and the last row of an available region of an image, respectively.

We parameterize amplitude $A(i)$ and bias $B(i)$ differently, namely,

$$A(i) = \frac{a_t - a_s}{i_t - i_s} (i - i_s) + a_s, \quad (10a)$$

$$B(i) = k^E i + b^E, \quad (10b)$$

where function $A(i)$ is a line passing through the first point (i_s, a_s) and the last point (i_t, a_t) . Parameters k^E and b^E are the slope and the intercept for $B(i)$, respectively. The use of the two-point form for $A(i)$ allows us to constrain the whole line segment to be positive/negative for all i , which is reflected by the constraints of the optimization problem (15) we will later formulate. On the other hand, we use the slope–intercept form for $B(i)$, which can help avoid undesirable local minima in which the two-point form can result. Details of the justification are provided in Appendix D of the supplementary materials.

C. ENF Embedding Using Additive Model

One or more columns from an image can be exploited for estimating an embedded ENF trace. Fig. 7(b) visualizes the

Fig. 7. (a) An image with an ENF trace in which a human-annotated smooth region is boxed in red dashed line and a nonsmooth region is boxed in black solid line. (b) Representative column signals from the smooth region and the nonsmooth region. The general trends of both column signals vary following that of the ENF trace at the time of capture.

image intensities of two columns from the image displayed in Fig. 7(a), where the *column signal* in red comes from the smooth region (boxed in dashed line) and the signal in black comes from the nonsmooth region (boxed in solid line). Since the native image content in the smooth region is nearly uniform, the column signal from that region tends to change with row indices following the ENF trace's fluctuation at the time of capture. Similarly, the column signal from the nonsmooth region roughly follows the ENF variation, but contains outliers compared to the column signal from the smooth region. Based on the observations of Fig. 7(b) and other column signal examples, we make a simplified assumption that the j th column of the native image content (without ENF) comprises an overall linear trend and an error term characterizing the visual components, acquisition noise, and outliers, namely,

$$x^{(j)}(i) = k^{x^{(j)}} i + b^{x^{(j)}} + \epsilon^{(j)}(i), \quad i \in [i_s, i_t], \quad (11)$$

where $k^{x^{(j)}}$ is the slope of the linear trend, $b^{x^{(j)}}$ is the intercept, and $\epsilon^{(j)}(i)$ is the zero-mean error term.

The embedding of ENF traces $e(i)$ into visual content $x^{(j)}(i)$ has been considered as an additive procedure in the literature [14], [16]–[18]. We will contrast the additive and multiplicative models in Section VI-C, but in the rest of the paper, we adopt the additive model. Combining the ENF trace model (9)–(10) and the native image content model (11), the pixel intensity signal at column j can be written as

$$I^{(j)}(i) = x^{(j)}(i) + e(i) \quad (12a)$$

$$= A(i) \cos\left(2\pi \frac{f}{f_{\text{row}}} i + \phi\right) + k_j i + b_j + \epsilon^{(j)}(i), \quad (12b)$$

where the slope term $k_j \stackrel{\text{def}}{=} k^E + k^{x^{(j)}}$ and the bias term $b_j \stackrel{\text{def}}{=} b^E + b^{x^{(j)}}$ are contributed by both models.

D. Entropy Minimization for Frequency Estimation

1) *General Principle*: Below, we outline the proposed idea of estimating the frequency from a single column signal from an ENF-corrupted image, and then extend it to using multiple column signals for improved estimation accuracy. The statistical behavior of ENF-induced entropy increase discussed in Section IV-A guarantees that, in general, if the added sinusoid is removed from the ENF corrupted image, the entropy will

reduce to the level of the uncorrupted image. A linear trend in an image can also increase the entropy by stretching the histogram wider. By examining the ingredients of the pixel intensity model (12b), we note that the first two terms are contributing to the inflated entropy of ENF-corrupted images. We define their sum as the *entropy-affecting signal* $I_e^{(j)}(i)$ to facilitate subsequent description of the proposed algorithm:

$$I_e^{(j)}(i) \stackrel{\text{def}}{=} A(i) \cos\left(2\pi \frac{f}{f_{\text{row}}} i + \phi\right) + k_j i. \quad (13)$$

We remark that the bias term b_j in (12b) does not change the entropy as it merely shifts the histogram. The noise term $\epsilon^{(j)}(i)$ in (12b) is a part of the native image (11) hence it is irrelevant to the entropy changes caused by ENF traces.

When the true parameters of (13) are plugged in and $I_e^{(j)}(i)$ is subtracted from the j th column signal of the ENF-corrupted image, the resulting entropy will be reduced. Formally, we can set up the following cost for the j th column signal as a function of $\theta_j = [a_s, a_t, f, \phi, k_j]^T$ for which the true parameters may lead to the global minimum or a local minimum:

$$J_j(\theta_j) = \text{entropy}\left(\text{hist}\left[\left\{I^{(j)}(i) - I_e^{(j)}(i)\right\}_{i=1}^M\right]\right) \quad (14a)$$

$$= \text{entropy}\left(\text{hist}\left[\mathbf{I}^{(j)} - \mathbf{I}_e^{(j)}\right]\right), \quad (14b)$$

where $\text{hist}(\cdot)$ calculates a histogram of an unordered/ordered set of numbers and $\text{entropy}(\cdot)$ returns the entropy after properly normalizing the input to a probability mass function. We note that $\text{hist}(\cdot)$ can be highly nonlinear as it involves counting numbers in predefined intervals. Among all unknown parameters, the frequency parameter f is our ultimate goal. The remaining parameters a_s , a_t , ϕ , and k_j are called nuisance parameters [31, Sec. 8.2] and need to be jointly solved with f to obtain the correct estimate of the frequency.

2) *The Complete Optimization Problem:* We propose to use multiple columns $\mathbf{I}^{(1)}, \dots, \mathbf{I}^{(K)}$ sparsely sampled from an ENF-corrupted image to improve the accuracy of the frequency estimation. We set up a multi-column optimization problem to solve for $\theta \stackrel{\text{def}}{=} [a_s, a_t, f, \phi, (k_j)_{j=1}^K]^T$ as follows:

$$\min_{\theta} \sum_{j=1}^K J_j(\theta) \quad \text{s.t. } a_s a_t \geq 0, \quad f \in \{100, 120\}. \quad (15)$$

The parameters a_s , a_t , f , and ϕ solely attributed to the embedded ENF trace are shared across K contributing columns, whereas the parameters k_j 's depending on the image content are allowed to vary column-wise.

This multi-column approach significantly improves the single-column approach (14) in two ways. First, the true solution may be merely a local minimum for each of the single-column cost functions $\{J_j(\theta_j)\}_{j=1}^K$. When the cost functions are summed up in (15), the true solution may be boosted to the global minimum and the aggregated cost at another local minimum may increase. A similar phenomenon is observed in frequency estimation algorithm MUSIC [32] in which multiple eigenvectors in the noise subspace contribute to finding the global minimum. Second, even when each of $\{J_j(\theta_j)\}_{j=1}^K$ can return a solution near the ground truth, it can be analytically

Fig. 8. Histograms for the estimates $\hat{a}_s$, $\hat{a}_t$, and $\hat{\phi}$ obtained from the entropy minimization applied on images with synthetically added ENF traces, where the ENF frequency was generated from the uniform distribution with the range $[120 - 0.24, 120 + 0.24]$ Hz and the remaining parameters of the ENF traces were fixed to $a_s = 18$, $a_t = 14$, $k = -1.68 \cdot 10^{-4}$, $b = 0.21$, and $\phi = 3.77$. The red dashed lines indicate each of the true values.

shown that the cost function involving K columns can boost the estimation accuracy by at most $\sqrt{K}$ times measured in the standard deviation of the frequency estimator. The proof can be found in Appendix E of the supplementary materials.

We further comment on the choices of constraints. The constraint $a_s a_t \geq 0$ is to ensure the same sign of the amplitude of the ENF trace at the first row, a_s , and the amplitude at the last row, a_t . If not, the amplitude-varying sinusoid will have a (fractional) row index with zero amplitude. Typical amplitude-varying sinusoids should look similar to those displayed in Fig. 5(c).

The constraint $f \in \{100, 120\}$ can significantly reduce the complexity of the optimization by effectively exploring one less dimension of the search space. This is motivated by the fact an instantaneous ENF usually departs by a small fraction, e.g., $\pm 0.2\%$ in Europe [33] and USA [34], from the nominal frequency. This added constraint is also justified by our observation that slight departure from the ground-truth frequency does not significantly affect the optimization results. We have experimentally verified this using images with synthetically added ENF traces, where the ENF frequency was generated from a uniform distribution with the range $[120 - 0.24, 120 + 0.24]$ Hz and the remaining parameters of the ENF traces were fixed to $a_s = 18$, $a_t = 14$, $k = -1.68 \cdot 10^{-4}$, $b = 0.21$, and $\phi = 3.77$. We added each synthetic ENF trace to the same image and applied the entropy minimization by setting the frequency parameter to 100 or 120 Hz, where the parameter search was initiated from the same initial point. Following the way to solve the optimization problem introduced in the next subsection, we obtained the estimates $\hat{a}_s$, $\hat{a}_t$, and $\hat{\phi}$ and drew a histogram for each of them as shown in Fig. 8. One can observe that the obtained estimates have slight bias and small variance, which implies that the optimization results are not significantly affected by slight deviations from the ground-truth frequency.

3) *Solving Optimization Problem via Decomposition:* We propose to solve the multi-column optimization problem (15) by decomposing it into two subproblems. The first subproblem solves for $\{k_j\}_{j=1}^K$ using the RANSAC algorithm [35], which is more robust to outliers than the least-squares method. This step estimates the slope of each selected column without depending on the sinusoidal component since they are relatively uncoupled as evidenced in Fig. 5(c). Plugging in the estimated $\{k_j\}_{j=1}^K$, the second subproblem solves for a_s , a_t , and ϕ for $f = 100$ and 120 Hz separately. The more

restricted 3-d search problem in the space of $[a_s, a_t, \phi]^T$ is easier especially when the cost function is nonconvex. We use a gradient-free optimizer as the numerical solver as it does not require computing gradients and the cost function to be convex or smooth. Specifically, we use the Nelder–Mead simplex method [36] that finds a solution by iteratively narrowing down the search space. In our implementation, we impose search-space constraints $|a_s| \leq 25$, $|a_t| \leq 25$, and $0 \leq \phi \leq 2\pi$, where 25 was empirically determined based on the largest possible amplitude of ENF traces. To understand the landscape of the 3-d error surfaces, we calculated its 2-d projection along the directions of a_s , a_t , ϕ , and many randomly chosen directions in the space of (a_s, a_t, ϕ) . We visualize the evolution of the 2-d projected error surfaces in animations that can be found in the multimedia supplemental files of the paper and in this URL: <https://bit.ly/3IlMilE>. It is observed that the cost function is generally well-behaved as the projected error surface evolves through the true solution, i.e., locally convex around the true solution with slight bias. Each of the 2-d projected error surfaces with the remaining parameters set to the true values is shown in Fig. 6(b).

We have outlined in this subsection a procedure to estimate the frequency when an image is corrupted by an ENF trace. However, this procedure was not designed to answer another important question: Is an ENF trace present in a given image? As we shall see in the next subsection, the presence of the ENF can be inferred from the intermediate quantities that appear in the algorithm for solving the optimization problem (15).

E. Proposed Two-Level ENF Presence–Classification Test

Next, we propose a two-level hypothesis testing framework as summarized in Algorithm 1 to determine the presence of the ENF and conduct a binary classification between 100 or 120 Hz if necessary. A first-level test will determine whether an ENF is present in a given test image. If yes, a second-level test will decide whether the test image was captured in a 100 or 120 Hz environment. Our testing framework relies on the intermediate quantities such as estimated nuisance parameters in Section IV-D and exploits the behaviors of the cost function (15) to construct effective test statistics. We present and explain the two subordinate tests of the framework below.

The first-level hypothesis testing is formulated as follows:

$$\begin{aligned} H_0 &: \text{ENF is not present,} \\ H_1 &: \text{ENF is present.} \end{aligned} \quad (16)$$

We propose a decision rule that rejects null hypothesis H_0 if $\text{entropy_margin} \geq \eta_1$, where $\eta_1 \in \mathbb{R}^+$ is a user-defined decision threshold. Here, the test statistic entropy_margin is defined to be the absolute difference between the values of the cost function (15) evaluated at two candidate frequencies, namely, $\text{entropy_margin} = |br(1) - br(2)|$, where $br(1)$ and $br(2)$ are the costs for $f = 100$ and 120 Hz, respectively.

We justify the use of the entropy margin as the test statistic for the first-level ENF presence test as follows. When the alternative hypothesis H_1 is true, the cost function (15) evaluated at $f = 100$ and 120 Hz will result in significantly different

Algorithm 1: Proposed Two-Level ENF Test.

Input: An RGB image with no ENF or ENF at 100/120 Hz
Output: no ENF or $\hat{f} \in \{100, 120\}$ Hz

Step 1. Initialization

- 1 Convert the image to gray scale
- 2 Randomly select K columns and augment them into a sample matrix $\mathbf{I} = [\mathbf{I}^{(1)}, \dots, \mathbf{I}^{(K)}]$
- 3 Use RANSAC algorithm to obtain $\hat{\mathbf{k}} = [\hat{k}_1, \dots, \hat{k}_K]$

Step 2. Estimation of Nuisance Parameters

- 4 $F(1) = 100, F(2) = 120$
- 5 $br(\ell) = \min_{a_s, a_t, \phi} \text{cost}(\mathbf{I}, F(\ell), a_s, a_t, \hat{\mathbf{k}}, \phi)$ s.t. $a_s a_t \geq 0, \forall \ell \in \{1, 2\}$

Step 3. Two-Level Classification

- 6 $\text{entropy_margin} = |br(1) - br(2)|$
- 7 **if** $\text{entropy_margin} < \eta_1$ **then**
- 8 | no ENF presence
- 9 **else**
- 10 | **if** $br(1) - br(2) < \eta_2$ **then**
- 11 | | $\hat{f} = 100$ Hz
- 12 | **else**
- 13 | | $\hat{f} = 120$ Hz

Subroutine $br = \text{cost}(\mathbf{I}, f, a_s, a_t, \hat{\mathbf{k}}, \phi)$

- 1 $\hat{\mathbf{I}}_e^{(j)}(i) = \left[\frac{a_t - a_s}{i_t - i_s} (i - i_s) + a_s \right] \cos \left(2\pi \frac{f}{f_{\text{row}}} i + \phi \right) + \hat{k}_j i, \quad i \in [i_s, i_t], \forall j$
- 2 $\hat{\mathbf{I}}(i, j) = \mathbf{I}(i, j) - \hat{\mathbf{I}}_e^{(j)}(i), \quad \forall (i, j)$
- 3 $\text{bitrate_array}(j) = \text{entropy}(\text{hist}(\hat{\mathbf{I}}(:, j))), \quad \forall j$
- 4 **return** $\text{mean}(\text{bitrate_array})$

values. Suppose the nominal ENF is 100 Hz. When the cost function is evaluated at $f = 100$ Hz, the residue, $\{\mathbf{I}^{(j)} - \mathbf{I}_e^{(j)}\}_{j=1}^K$ in (15) is close to detrended image content, which should lead to lower cost value per the statistical phenomenon reported in Section IV-A. In contrast, when the cost function is evaluated at $f = 120$ Hz given the true ENF at 100 Hz, the residue contains two extra sinusoidal-like ENF-trace components at 100 and 120 Hz in addition to the detrended image content. The extra components will increase the dynamic range of the residue and therefore lead to a higher entropy or cost function value. Hence, the costs at $f = 100$ and 120 Hz are very different under H_1 , and it is easy to see that the same argument also applies when the nominal ENF is 120 Hz. In contrast, when the null hypothesis H_0 is true, no matter $f = 100$ or 120 Hz, the cost function contains residue that is corrupted by one sinusoidal-like ENF trace contributed by the prediction model (13). Hence, the costs evaluated at the two candidate frequencies may not differ as significantly as in H_1 . Simulated results in Fig. 9 confirms that when no ENF traces are added to images, the entropy margin is close to zero. In contrast, when ENF traces are added to images, the entropy margin are nonzero.

Fig. 9 also shows that the entropy margin increases as a function of the ENF amplitude, which reveals that the proposed first-level ENF presence test can provide higher statistical power with stronger ENF. This is another desirable characteristic of proposed test. In Appendix F of the supplementary materials, we analytically justified that as the ENF

Fig. 9. Box plots of entropy margin for different ENF strengths. The strength is quantified using the amplitude of the synthetically added sinusoidal ENF traces. The pixel intensity ranges from 0 to 255. The entropy margin is almost zero when no ENF traces are present in images, whereas it is nonzero when ENF traces are present. As the ENF strength increases, the entropy margin also increases.

amplitude increases, the entropy margin tends to increase, as revealed in Fig. 9.

If the first-level test decides that an ENF trace is present, the test image will be passed to the second-level hypothesis testing formulated as follows:

$$\begin{aligned} H_0 : f &= 100 \text{ Hz}, \\ H_1 : f &= 120 \text{ Hz}. \end{aligned} \quad (17)$$

Instead of using a simple decision rule such as $br(1) > br(2)$, we propose to reject the null hypothesis H_0 if $br(1) - br(2) \geq \eta_2$, where $\eta_2 \in \mathbb{R}$ is a user-defined decision threshold. This decision rule shares a similar spirit as the likelihood-ratio test [31, Sec. 8.2] and Bayesian decision theory that allows incorporating the prior beliefs and unequal misclassification losses for two candidate classes [37, Sec. II].

V. EXPERIMENTAL RESULTS

A. Results for Images with Synthetic ENF Traces

We created a set of test images to simulate those captured in 50 Hz and 60 Hz regions using six template images that do not contain detectable ENF traces. Based on the parameterization in (9) and (10), we generated synthetic ENF traces parameterized by $(a_s, a_t, \phi, k^E, b^E)$, where the mean sinusoidal amplitude, $L = \frac{a_s + a_t}{2} \in \{2^{-1}, 2^{-0.5}, 2^0, 2^{0.5}, 2^1, 2^2, 2^3, 2^4\}$, has the same unit as the pixel intensity with 256-level shades of gray. For each L , we randomly assigned each of the parameters as follows: $a_s = L - 0.5\delta$ and $a_t = L + 0.5\delta$, where $\delta \sim \mathcal{U}(-L, L)$. $\phi \sim \mathcal{U}(0, 2\pi)$, $k^E \sim \mathcal{U}(-2/M, 2/M)$, and $b^E \sim \mathcal{U}(-1, 1)$, where M is the number of rows of an image. Then, we added the generated synthetic ENF traces to the native images to generate a total of 96 ($= 2 \times 6 \times 8$) images. For each image, the entropy minimization and a baseline method described in Appendix G of the supplementary materials were applied 20 times to ensure that we obtain consistent results. We randomly chose evenly spaced $N = 32$ columns. The performance of the first-level and the second-level classifications was assessed separately.

Fig. 10 shows representative results of the entropy minimization applied on images with synthetic ENF traces. Each row has one example that contains an image with ENF trace,

Fig. 10. Representative results of the entropy minimization applied on images with synthetic ENF traces. Each row shows one example containing an ENF-corrupted image, an ENF-trace-removed image, and an estimated sinusoidal-like ENF trace (along with the true ENF trace) as a function of the row index. The aligned zero crossings of the estimated and true sinusoidal-like ENF traces imply that ENF traces are well estimated by the entropy minimization.

an ENF trace-removed image, and an estimated sinusoidal-like ENF trace (along with the true ENF trace). We only display the sinusoidal-like term $A(i) \cos(2\pi f / f_{\text{row}} i + \phi)$ but ignore the bias $B(i)$ of (9), since the ENF slope term k^E in $B(i)$ is mixed with the content slope term $k^{x^{(j)}}$ as in (12b) and therefore none of them are individually estimable. It is observed that the zero crossings of the estimated and true sinusoidal-like ENF traces are well aligned, which implies that ENF traces are well estimated by the entropy minimization. Fig. 14 in Appendix H of the supplementary materials shows a trend that the stronger an embedded ENF trace is, the more reliable the estimated ENF traces is in general. Fig. 15 in Appendix H also presents receiver operating characteristic (ROC) results of the entropy minimization for images with ENF traces of varying strengths.¹ The ROC results reveal that as the ENF strength increases, both the first-level and the second-level classification performance improves, yielding equal error rates (EER) closer to 0. EERs represent the accuracy at the ROC operating point, where the false positive and false negative rates are equal. Similar results can also be found in the next subsection when we tested images with real ENF traces.

B. Results for Images with Real ENF Traces

We also created a real-world image dataset with iPhone captured indoor photos. By following the image acquisition setup described in Appendix I of the supplementary materials,

¹For images with real ENF traces, we will include the ROC results in the next subsection. Due to the limitation of space, we show most results for images with synthetic ENF traces in Appendix H of the supplementary materials.

we acquired in the United States a dataset with 60 Hz ENF trace using an iPhone 6s; in China a dataset with 50 Hz ENF trace using an iPhone 6. Each dataset contains ten scenes, each of which was obtained with different ENF strength levels set on the camera by changing the exposure time. As shown in (7), the exposure time determines the intensity of the ENF trace embedded in the captured images. We show all captured photos in Fig. 17 of the supplementary materials. For each image, the entropy minimization and the baseline method were applied five times to ensure consistent results.

We assess performance through the ROC curves shown in Fig. 11. At the first level, the presence of an ENF trace is considered positive; at the second level, the 120 Hz ENF trace is treated as the positive case. The top row of Fig. 11 shows the ROC curve results of images with real ENF traces when the entropy minimization method applied. Performance of the first-level and second-level classifiers is shown in the left and right panels of Fig. 11, respectively. Square, diamond, and triangle markers in the top row's left panel of Fig. 11 are operating points of the first-level classifier with decision thresholds $\eta_1 = 0.006, 0.008, \text{ and } 0.010$, respectively. The top row's right panel of Fig. 11 contains results for the second-level classifier of the entropy minimization applied on true positive cases of the first-level classifier.

The top row's left panel of Fig. 11 shows the first-level decision results of the entropy minimization that as the ENF strength increases, the classification performance generally improves. The second-level decision results of the entropy minimization given in the top row's right panel of Fig. 11 show that when the strength is the lowest, the second-level classifier performs the worst regardless of the operating point of the first-level classifier. Conversely, when the strength increases above 0.23 corresponding to the ENF strength embedded in the images in the first three columns of Fig. 16 in Appendix I of the supplementary materials, the EER begins to approach 0.

The bottom row of Fig. 11 shows the ROC results of images with real ENF traces when the baseline method applied. Square, diamond, and triangle markers in the bottom row's left panel of Fig. 11 are operating points of the first-level classifier with decision thresholds $\gamma_1 = 10, 20, \text{ and } 25$, respectively. The bottom row's right panel of Fig. 11 contains results for the second-level classifier of the baseline method applied on true positive cases of the first-level classifier. One can observe that in most cases, the baseline method underperforms the entropy minimization method. For example, the first-level classifier of the baseline method is little more than a random guess when the strength is below 0.66. Even when the strength is more than equal to 0.66, it is worse than the corresponding performance level of the entropy minimization method. The results for the second-level classifier also show no consistent performance trend as a function of the ENF strength and no perfect classification result with $\text{EER} = 0$ at any strength level.

C. Comparison with Preliminary Parametric Model

To compare the proposed more sophisticated parametric model to the preliminary parametric model (1) proposed

in [1], we applied entropy minimization using the preliminary parametric model on the real-world image dataset of Section V-B. Fig. 12 compares the EER results of the two models for the first-level classifier. The EERs are shown as a function of the ENF strength defined as the numerator of (7), i.e., $|\text{sinc}(\omega_0 t_{\text{exp}}/\pi)|$. The proposed more sophisticated parametric model can provide much better ENF detection performance under challenging scenarios when the ENF is weak. In contrast, when the ENF is strong, the performance is comparable. This observation implies that our proposed model performs better than the preliminary model in more challenging scenarios of dealing with images with weaker ENF traces—The weaker traces are more difficult to extract amid dominating visual content.

VI. DISCUSSION

A. ENF Embedding Process for Fluorescent Lamps

Appendix A of the supplementary materials analyzes the ENF embedding process when the source light is a nonideal fluorescent lamp. A more realistic lamp voltage response is close to a periodic square wave given a sinusoidal-like alternating supply current [38], [39]. The result of (21) of the supplementary materials shows that the steady-state instantaneous power consumed by a fluorescent lamp contains frequency components at $2\omega_0 k$, $k \in \mathbb{N}$. In contrast, the instantaneous power signal of (2c) in the ideal model has the fundamental frequency only at $2\omega_0$. Even though the more realistic case contains extra harmonics, frequency component $2\omega_0$ will end up dominating the final camera-captured image because of the power-to-illumination conversion (3) and the camera integration (4a), which act as lowpass filters. The low-pass filters not only suppress higher harmonics of the power signal but also attenuate the strength of the AC component relative to the DC. Hence, the resulting pixel intensity contains a DC component and a relatively small sinusoidal-like flicker AC component with the frequency component $2\omega_0$, regardless of the characteristic of the voltage response of the lamp.

B. Color Combination

In images captured under the lighting of fluorescent lamps, we may see colored ENF stripes that may be caused by uneven energy distributions in the radiation spectrum. Thus, instead of using the gray color combination being a weighted average of RGB colors, using a weight vector emphasizing the light spectrum with stronger spectral energy may improve the performance. We leave the investigation of colored ENF stripes as future work.

C. Multiplicative Model

In addition to the additive model [14], [16]–[18] that is resilient to the global intensity bias used in this paper, we also considered using the multiplicative model that is resilient to albedo and illumination changes. In a realistic scenario of capturing an image using rolling shutters, the light interacts with physical surfaces within the camera view. This interaction with physical surfaces can be nonlinear, making it more sense

Fig. 11. ROC curves of the proposed entropy minimization method in the top row and the baseline method in the bottom row applied on images containing real ENF traces with different ENF strengths. Performance of the first-level classifier is shown in the left panel and performance of the second-level classifier is shown in the right panel. Square, diamond, and triangle markers in the top row's left panel and in the bottom row's left panel are operating points of the first-level classifier of the entropy minimization with decision thresholds $\gamma_1 = 0.006, 0.008, \text{ and } 0.010$ and the first-level classifier of the baseline method with decision thresholds $\gamma_1 = 10, 20, \text{ and } 25$, respectively. The top row's right panel contains results for the second-level classifier of the entropy minimization applied on true positive cases of the first-level classifier. The bottom row's right panel contains results for the second-level classifier of the baseline method applied on true positive cases of the first-level classifier. The entropy minimization outperforms the baseline method in most cases and the classification performance of both methods generally improves as the ENF strength increases.

Fig. 12. EER comparison of the first-level classifiers of the proposed parametric model and the preliminary model [1] as a function of the ENF strength. The proposed, more sophisticated model has much better ENF detection performance under challenging scenarios when ENF is weak.

to use the multiplicative model as shown in (18a). A large part can, however, be considered additive to the native image content when the strength of the AC is small relative to that of the DC component of ENF traces, namely, $e(i) \ll 1$. We mathematically demonstrate the argument as follows:

$$I^{(j)}(i) = x^{(j)}(i) \cdot [1 + e(i)] \quad (18a)$$

$$\approx x^{(j)}(i) \cdot \left[1 + \frac{C}{x^{(j)}(i)} e(i)\right] \quad (18b)$$

$$= x^{(j)}(i) + C \cdot e(i), \quad (18c)$$

where C is the average intensity of the image. Also, the additive model is more mathematically tractable. Our experimental results showed that an algorithm developed based on

the additive model has a better performance on real-world data than the multiplicative model.

VII. CONCLUSION AND FUTURE WORK

In this paper, we have explored an almost unnoticeable power signature, the ENF trace that can be embedded in images acquired by CMOS cameras for geotagging purposes. Specifically, we have addressed a very challenging research question of whether the basic attributes of an embedded ENF trace can be correctly identified. To this end, we have investigated physical ENF embedding processes considering experimentally measured power supply signals and a light reflection-transmission model. We have discovered that an image-embedded ENF trace is a sinusoidal wave with varying amplitude and varying bias.

We have proposed the parametric model of image-embedded ENF traces with varying amplitude and varying bias, which allows the entropy minimization method to extract more precise ENF traces. We have subsequently designed the two-level ENF presence-classification test to identify images among no ENF, 50 Hz ENF, and 60 Hz ENF. The experimental results have shown that our proposed method is able to make higher decision accuracy compared to the baseline method.

In this proof-of-concept work, we have demonstrated a unique forensic capability of extracting invisible traces to help narrow down the capturing geographic region of an image. It is important to recall the significant challenge in extracting

ENF traces over a short duration buried in single images, as it is extremely difficult to overcome the dominating visual content and reveal the weak ENF traces. In contrast, the ENF extraction from video [16], [18] can take advantage of multiple frames to estimate the native visual content and separate the ENF from the visual content. Despite the difficulty, the results have shown promising capability by our proposed algorithms as the first attempt to tackle the challenge.

In future work, we plan on examining the performance on a larger scale investigation with more cameras and images. We also plan to explore additional location information that may be inferred from the ENF traces using a set of image frames captured in a continuous shooting mode. It may facilitate wider adoption of the ENF analysis in the image forensics field.

ACKNOWLEDGMENT

We thank Mr. Yuan Frank Fang for providing the references on LED drivers and for discussing the lighting mechanisms.

REFERENCES

- [1] C.-W. Wong, A. Hajj-Ahmad, and M. Wu, "Invisible geo-location signature in a single image," in *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, May 2018, pp. 1987–1991.
- [2] "Finder (IARPA Research Program)," [Online]. Available: <https://www.iarpa.gov/index.php/research-programs/finder>, Accessed November 2016.
- [3] "Operation Predator – Wikipedia," [Online]. Available: https://en.wikipedia.org/wiki/Operation_Predator, Accessed November 2016.
- [4] "Child Exploitation Investigations Unit, U.S. Department of Homeland Security," [Online]. Available: <https://www.ice.gov/predator>, Accessed November 2016.
- [5] "Child Protection from Violence, Exploitation and Abuse, UNICEF at United Nation," [Online]. Available: <https://www.unicef.org/protection/>, Accessed November 2016.
- [6] C. Grigoros, "Digital audio recordings analysis: The electric network frequency (ENF) criterion," *Int. J. Speech Lang. Law*, vol. 12, no. 1, pp. 63–76, 2005.
- [7] —, "Applications of ENF analysis in forensic authentication of digital audio and video recordings," *Journal of the Audio Engineering Society*, vol. 57, no. 9, pp. 643–661, 2009.
- [8] Y. Liu, L. Zhan, Y. Zhang, P. N. Markham, D. Zhou, J. Guo, Y. Lei, G. Kou, W. Yao, J. Chai *et al.*, "Wide-area-measurement system development at the distribution level: An FNET/GridEye example," *IEEE Trans. Power Deliv.*, vol. 31, no. 2, pp. 721–731, Apr. 2015.
- [9] D. P. N. Rodríguez, J. A. Apolinário, and L. W. P. Biscainho, "Audio authenticity: Detecting ENF discontinuity with high precision phase analysis," *IEEE Trans. Inf. Forensics Security*, vol. 5, no. 3, pp. 534–543, Sep. 2010.
- [10] P. A. Andrade Esquef, J. A. Apolinário, and L. W. P. Biscainho, "Edit detection in speech recordings via instantaneous electric network frequency variations," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 12, pp. 2314–2326, Dec. 2014.
- [11] 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.
- [12] 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.
- [13] N. Fechner and M. Kirchner, "The humming hum: Background noise as a carrier of ENF artifacts in mobile device audio recordings," in *Int. Conf. IT Secur. Incid. Manag. IT Forensics*, May 2014, pp. 3–13.
- [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 *ACM Int. Workshop on Immersive Media Experiences*, Nov. 2014, pp. 13–18.
- [15] A. Hajj-Ahmad, C.-W. Wong, S. Gambino, Q. Zhu, M. Yu, and M. Wu, "Factors affecting ENF capture in audio," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 2, pp. 277–288, Feb. 2018.
- [16] R. Garg, A. L. Varna, and M. Wu, "'Seeing' ENF: Natural time stamp for digital video via optical sensing and signal processing," in *ACM Int. Conf. Multimed.*, 2011, pp. 23–32.
- [17] 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.
- [18] H. Su, A. Hajj-Ahmad, R. Garg, and M. Wu, "Exploiting rolling shutter for ENF signal extraction from video," in *IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 5367–5371.
- [19] S. Vatansever, A. E. Dirik, and N. Memon, "Factors affecting ENF based time-of-recording estimation for video," in *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, May 2019, pp. 2497–2501.
- [20] O. Ait-Aider, A. Bartoli, and N. Andreff, "Kinematics from lines in a single rolling shutter image," in *IEEE/CVF Conf. Comput. Vision Pattern Recog. (CVPR)*, Jun. 2007, pp. 1–6.
- [21] J. Gu, Y. Hitomi, T. Mitsunaga, and S. Nayar, "Coded rolling shutter photography: Flexible space-time sampling," in *IEEE Int. Conf. Comput. Photogr. (ICCP)*, Mar. 2010, pp. 1–8.
- [22] S. Vatansever, A. E. Dirik, and N. Memon, "Analysis of rolling shutter effect on ENF-based video forensics," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 9, pp. 2262–2275, Sep. 2019.
- [23] J. Choi and C.-W. Wong, "ENF signal extraction for rolling-shutter videos using periodic zero-padding," in *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2019, pp. 2667–2671.
- [24] A. Hajj-Ahmad, S. Baudry, B. Chupeau, and G. Doërr, "Flicker forensics for pirate device identification," in *ACM Workshop Inf. Hiding Multimed. Secur.*, ser. IH&MMSec, Portland, Oregon, USA, 2015, pp. 75–84.
- [25] A. Hajj-Ahmad, S. Baudry, B. Chupeau, G. Doërr, and M. Wu, "Flicker forensics for camcorder piracy," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 1, pp. 89–100, Jan. 2017.
- [26] *A practical guide to time-resolved luminescence lifetime determination using dedicated Time-Correlated Single-Photon Counting systems*, Horiba, Accessed July 2021. [Online]. Available: <https://www.horiba.com/fileadmin/uploads/Scientific/Downloads/UserArea/Fluorescence/Manuals/IBH-TCS-SPC-Guide.pdf>
- [27] M. Wahl, "Time-correlated single-photon counting," Pico-Quant, Berlin, Germany, Tech. Rep., 2014, Accessed July 2021. [Online]. Available: <https://drive.google.com/drive/u/1/folders/1gJFCJUYw6xGWpTWlh3pvVgAqyOMQESaW>
- [28] G. D. Boreman and E. R. Raudenbush, "Modulation depth characteristics of a liquid crystal television spatial light modulator," *Applied Optics*, vol. 27, no. 14, pp. 2940–2943, Jul. 1988.
- [29] S. K. Nayar, K. Ikeuchi, and T. Kanade, "Surface reflection: Physical and geometrical perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 7, pp. 611–634, Jul. 1991.
- [30] R. Szeliski, *Computer Vision: Algorithms and Applications*. Springer Science, 2010.
- [31] G. Casella and R. L. Berger, *Statistical Inference*. Cengage Learning, 2001.
- [32] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, Mar. 1986.
- [33] C. Grigoros, "Applications of ENF criterion in forensic audio, video, computer and telecommunication analysis," *Forensic Sci. Int.*, vol. 167, pp. 136–145, Apr. 2007.
- [34] R. W. Sanders, "Digital audio authenticity using the electric network frequency," in *Audio Eng. Society Int. Conf. on Audio Forensics—Theory and Practice*, Jun. 2008.
- [35] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. of the ACM*, vol. 24, pp. 381–395, Jun. 1981.
- [36] S. Singer and J. Nelder, "Nelder-Mead algorithm," *Scholarpedia*, vol. 4, no. 7, p. 2928, 2009.
- [37] H. V. Poor, *An Introduction to Signal Detection and Estimation*. Springer Science & Business Media, 2013.
- [38] E. Gluskin, "The fluorescent lamp circuit," *IEEE Trans. Circuits Syst.*, vol. 46, no. 5, pp. 529–544, May 1999.
- [39] E. Gluskin, F. V. Topalis, I. Kateri, and N. Bisketzis, "The instantaneous light-intensity function of a fluorescent lamp," *Physics Lett. A*, vol. 353, no. 5, pp. 355–363, Mar. 2006.
- [40] W. Elenbaas, *Fluorescent Lamps*. Macmillan, 1971.
- [41] *Curve Fitting Toolbox*, MathWorks, Inc., Natick, MA, USA, Accessed Dec 2021. [Online]. Available: <https://www.mathworks.com/products/curvefitting.html>
- [42] P. D. Hiscocks and P. Eng, "Measuring luminance with a digital camera," Syscomp Electronic Design Limited, Tech. Rep., 2011, Accessed July 2021.

SUPPLEMENTARY MATERIALS

APPENDIX A

PHYSICAL EMBEDDING PROCESS FOR FLUORESCENT LAMPS

A study in [40] shows that gas-discharge lamps such as fluorescent lights have a negative voltage-current characteristic, where an increase in current across the lamp terminals results in a voltage drop. Consequently, when an alternating current of sinusoidal form is applied to a fluorescent lighting system, the lamp voltage is significantly distorted. Fig. 7.1 of [40] illustrates the lamp voltage formation against one half cycle of sinusoidal alternating current. During the following half cycle, the same lamp voltage pattern with the opposite sign will be observed.

To facilitate analysis, the lamp voltage has often been represented as a square wave [?], [?]: $v_l(t) = V_l \text{sign}[i_l(t)]$, where $V_l \in \mathbb{R}^+$ is the lamp voltage amplitude, $i_l(t)$ is the alternating supply current, and the sign is defined as follows:

$$\text{sign}(x) = \begin{cases} -1, & \text{if } x < 0, \\ 0, & \text{if } x = 0, \\ 1, & \text{if } x > 0. \end{cases} \quad (19)$$

Using the power law, we define the steady-state instantaneous power consumed by the lamp at time t as

$$p_l(t) = v_l(t) i_l(t) = V_l I_l |\sin(\omega_0 t + \phi)|, \quad (20)$$

where $I_l \in \mathbb{R}^+$ is the current amplitude, $\omega_0 = 2\pi f_0$ is the angular frequency relating to the mains frequency $f_0 \in \{50, 60\}$ Hz, and $\phi \in [0, 2\pi)$ is the initial phase. According to the Fourier series of $|\sin(\omega_0 t)|$, the power $p_l(t)$ can be expanded in terms of cosine functions:

$$p_l(t) = V_l I_l \left[\frac{2}{\pi} - \frac{4}{\pi} \left(\frac{\cos(2\omega_0 t + 2\phi)}{3} + \frac{\cos(4\omega_0 t + 4\phi)}{15} + \dots \right) \right], \quad (21)$$

in which the frequency component at $2\omega_0$ dominates. The time-dependent part of $p_l(t)$ determines changes in light intensity.

Fluorescent lamps generate light by exciting mercury vapor that produces ultraviolet radiation through the discharge process. The ultraviolet generated in the gas discharge causes the phosphor on the inner wall of the tube to radiate visible light. To detect photons of visible light and measure fluorescence lifetime, a technique named *time-correlated single photon counting* (TCSPC) [26], [27] is commonly used. Using the fact that the likelihood of observing a photon at a certain time after a pulsed light excitation is proportional to fluorescence intensity at that time, TCSPC helps to draw histograms of fluorescence photon arrival times from a series of excitation pulses. Studies [26], [27] show that the fluorescence intensity is modeled by a single exponential decaying function or a sum of multiple exponential functions. Following the literature, we assume that the impulse response of the power-to-light conversion process takes the form of single exponential decay. Thus, the resulting illuminance of fluorescent lamps is represented as

$$\ell(t) = K_l p_l(t) * e^{-\frac{t}{\tau}} u(t), \quad (22)$$

where “ $*$ ” is the linear convolution operator, τ is a time constant accounting for the illumination half-life, and $u(t)$ is a unit step function with value 1 for $t \geq 0$ and 0 for elsewhere. Note K_l is a power factor considering that a portion of the consumed power will contribute to the illumination of the light.

APPENDIX B

ATTRIBUTION OF ARTIFACTS OF TIME-VARYING BEHAVIOR

In Section III-A, we notice that the time-varying behavior of the camera-captured pixel intensity signal is present when the nonideal sinusoidal-like supply voltage is applied. To investigate whether the varying frequency and/or the varying amplitude of the supply voltage are the cause(s), we design an experiment, in which simulated voltage signals of different characteristics are fed into the physical embedding process for a factor analysis. We adopt the following three parametric models of supply voltage, namely, a constant frequency and constant amplitude (ConFreq-ConAmp) sinusoidal model, a varying frequency and constant amplitude (VarFreq-ConAmp) model, and a varying frequency and varying amplitude (VarFreq-VarAmp) model, denoted by

$$\text{ConFreq-ConAmp: } v(t) = V_l \cos(\omega_0 t + \phi), \quad (23a)$$

$$\text{VarFreq-ConAmp: } v(t) = V_l \cos \left(\int_0^t \omega_0(t) dt + \phi \right), \quad (23b)$$

$$\text{VarFreq-VarAmp: } v(t) = V_l(t) \cos \left(\int_0^t \omega_0(t) dt + \phi \right), \quad (23c)$$

where V_l , $V_l(t)$, ω_0 , $\omega_0(t)$, and ϕ represent the constant amplitude, the varying amplitude, the constant angular frequency, the varying angular frequency, and the initial phase, respectively. The ConFreq-ConAmp model is the standard sinusoid with fixed parameters, whereas the VarFreq-ConAmp and VarFreq-VarAmp models have time-varying amplitude and/or frequency parameters.

We estimate the parameters of each model as follows. For the ConFreq-ConAmp model (23a), we estimate the constant parameters V_l and ω_0 by applying a trust-region-based nonlinear least-squares algorithm [41] to the whole voltage signal. We calculated a sequence of 6.8 msec segments with 50% overlap and averaged them based on triangular-windowed moving average to obtain the simulated signals of the VarFreq-ConAmp and VarFreq-VarAmp models. For the VarFreq-VarAmp model (23c), we estimated a voltage amplitude–frequency pair for each signal segment. By creating a sequence of overlapped windows, we will obtain a sequence of time-varying voltage amplitude and frequency. The estimation procedure for the VarFreq-ConAmp model (23b) contains an extra step in addition to that for the VarFreq-VarAmp model: the constant voltage V_l is estimated by averaging the sequence of the time-varying voltage amplitude.

Fig. 4(a) shows three simulated pixel intensity signals $I(t)$ driven by the different voltage models in (23). We used the same supply voltage measurement in Section III-A to generate the three simulated voltage signals. Then, we simulated the pixel intensity signals $I(t)$ by setting the parameters to be

the same as those in the experimental setup of Fig. 3 except with $t_{\text{exp}} = 40$ msec. Note that we used an exposure time of 40 msec, so that the amplitude and the frequency can vary dynamically within the time window when considering the length of simulated voltage segments of the VarFreq-ConAmp and VarFreq-VarAmp models. One can observe that the VarFreq-ConAmp and VarFreq-VarAmp cases can result in time-varying trends and the VarFreq-VarAmp leads to even more variations, which means that both the varying amplitude and the varying frequency in the AC source are the cause of the time-varying behavior for the embedded ENF traces in images. Fig. 4(b) shows a few zoomed-in segments of the simulated $I(t)$ for the VarFreq-ConAmp and VarFreq-VarAmp models, in which the displayed signal length roughly corresponds to 2, 2.5, and 3 cycles of a single image's acquisition duration. The segments' envelopes and biases roughly follow linear trends, which allows us to simplify the parameterization of the amplitude and the bias of the ENF trace using affine functions in Section IV-B.

APPENDIX C

AMPLITUDE OF AN ENF TRACE AS A FUNCTION OF ROW INDEX

Fig. 5(b) shows that given a point light source at $(0, h)$, the incident light reaches the object's surface of locations $(b, 0)$ and $(b+u, 0)$ with incident angles θ_0 and θ_i , respectively. The distances from the light source to locations $(b, 0)$ and $(b+u, 0)$ are defined to be d_0 and d_i . After hitting the object's surface, the reflected light goes through the camera located at $(b+b', h')$ that is d'_0 and d'_i away from each location of interest.

The light intensity L_i of the reflected light at location $(b+u, 0)$ arriving at the camera can be expressed as a function of row i of an image, namely,

$$L_i = \frac{c \cdot \cos \theta_i}{d_i^2 d'_i{}^2} \quad (24a)$$

$$= ch [h^2 + (b+u)^2]^{-\frac{3}{2}} [h'^2 + (b'-u)^2]^{-1}, \quad (24b)$$

where c is a constant per (8) and we assume the scale of the captured scene u is much smaller than the vertical distance h and horizontal distance b to the light source and the horizontal distance b' to the camera, i.e.,

$$0 \ll u \ll \min(h, b, b'), \quad (25)$$

which we refer to as the small-scale assumption. To verify that L_i is approximately linear in u , we take the partial derivative with respect to u and examine if it is constant.

$$\begin{aligned} \frac{\partial L_i}{\partial u} &= -3ch \left\{ b^2 \cdot \left[\left(\frac{h}{b} \right)^2 + \left(1 + \frac{u}{b} \right)^2 \right] \right\}^{-\frac{5}{2}} \left\{ b'^2 \cdot \left[\left(\frac{h'}{b'} \right)^2 + \left(1 - \frac{u}{b'} \right)^2 \right] \right\}^{-1} (b+u) \\ &\quad + 2ch \left\{ b^2 \cdot \left[\left(\frac{h}{b} \right)^2 + \left(1 + \frac{u}{b} \right)^2 \right] \right\}^{-\frac{3}{2}} \left\{ b'^2 \cdot \left[\left(\frac{h'}{b'} \right)^2 + \left(1 - \frac{u}{b'} \right)^2 \right] \right\}^{-2} (b'-u) \end{aligned} \quad (26a)$$

$$\begin{aligned} &= -3chb^{-5}b'^{-2} \left[\left(\cot^2 \theta_0 + \left(1 + \frac{u}{b} \right)^2 \right)^{-\frac{5}{2}} \left[\cot^2 \theta'_0 + \left(1 - \frac{u}{b'} \right)^2 \right]^{-1} (b+u) \right. \\ &\quad \left. + 2chb^{-3}b'^{-4} \left[\left(\cot^2 \theta_0 + \left(1 + \frac{u}{b} \right)^2 \right)^{-\frac{3}{2}} \left[\cot^2 \theta'_0 + \left(1 - \frac{u}{b'} \right)^2 \right]^{-2} (b'-u) \right] \right] \end{aligned} \quad (26b)$$

Due to the small-scale assumption of the captured scene, we have $\cot^2 \theta_0 + (1 + u/b)^2 \approx \cot^2 \theta_0 + 1 = \sin^{-2} \theta_0$ and $\cot^2 \theta'_0 + (1 - u/b')^2 \approx \cot^2 \theta'_0 + 1 = \sin^{-2} \theta'_0$. We further reduce (26) to

$$\frac{\partial L_i}{\partial u} \approx -3chb^{-5}b'^{-2} \sin^5 \theta_0 \sin^2 \theta'_0 (b+u) \quad (27a)$$

$$+ 2chb^{-3}b'^{-4} \sin^3 \theta_0 \sin^4 \theta'_0 (b'-u) \quad (27b)$$

$$= -3chd_0^{-5}d_0'^{-2}(b+u) + 2chd_0^{-3}d_0'^{-4}(b'-u) \quad (27c)$$

$$= -3chd_0^{-4}d_0'^{-2} \left(\frac{b}{d_0} + \frac{u}{d_0} \right) + 2chd_0^{-3}d_0'^{-3} \left(\frac{b'}{d'_0} - \frac{u}{d'_0} \right). \quad (27d)$$

Applying the small-scale assumption $u \ll b < d_0$ and $u \ll b' < d'_0$ again, we finally have

$$\frac{\partial L_i}{\partial u} \approx -3chd_0^{-4}d_0'^{-2} \left(\frac{b}{d_0} \right) + 2chd_0^{-3}d_0'^{-3} \left(\frac{b'}{d'_0} \right) \quad (28a)$$

$$= -3chbd_0^{-5}d_0'^{-2} + 2chb'd_0^{-3}d_0'^{-4}, \quad (28b)$$

which is a constant not depending on u .

The result of (28) shows that the light intensity L_i has a linear relationship in u that can be related to the row index i of the image. When the light reflecting off the object's surface travels through the camera lens, in the case of DSLR cameras, it is turned upside-down and then flipped over again by a pentaprism. In the case of mobile cameras, as there is no reflecting mechanism, the light directly hits the camera imaging sensor. Basically, regardless of the camera type, a linear relationship between u and the row index i holds, meaning that as u increases, it contributes to forming the higher row of the image. Therefore, it is concluded that the light intensity L_i of location $(b+u, 0)$ picked up by the camera has a linear relationship with the row index i of an image.

APPENDIX D

COMPARISON OF ERROR SURFACES FORMED BY SLOPE-INTERCEPT AND TWO-POINT FORMS

There are two common ways of parameterizing $B(i)$ of (9): the slope-intercept form and the two-point form. The slope-intercept form is written as

$$B(i) = k^E i + b^E, \quad (29)$$

where k^E and b^E are the slope and the intercept for $B(i)$, respectively. The two-point form can be written as

$$B(i) = \frac{b_t - b_s}{i_t - i_s} (i - i_s) + b_s, \quad (30)$$

where function $B(i)$ is a line passing through the first point (i_s, b_s) and the last point (i_t, b_t) .

Given an image corrupted by a synthetic ENF trace, we visualize the landscapes of the cost function to determine a better parameterization for $B(i)$ between the two possibilities. Fig. 13(a) and Fig. 13(b) correspond to error surfaces of the slope–intercept form and the two-point form, respectively. Each of them is shown with respect to two search variables when the remaining search variables, including the frequency parameter, are set to the true values. Regardless of the parameterization, the cost functions for the three combinations (a_s, a_t) , (a_s, ϕ) , and (a_t, ϕ) are generally well-behaved, i.e., locally convex around the true solution with slight bias. On the other hand, as long as one of the parameters relates to the bias-related parameter, i.e., b for the slope–intercept form and b_s and b_t for the two-point form, there is little change in the cost value as a function of the bias-related parameter.

An interesting observation is made from the error surfaces of (k, b) of Fig. 13(a) and (b_s, b_t) of Fig. 13(b), which suggests that the slope–intercept form is preferred over the two-point form. The variable b of the slope–intercept form merely shifts the histogram without changing the histogram shape. This does not change the cost function, making it impossible to estimate b . Meanwhile, k changes the histogram shape leading to the change in the cost function, which makes it possible to determine k . These behaviors are seen in Fig. 13(a). Yet, in practice, we choose to combine each of them with other parameters for the native image content signal and set them to some fixed values using the RANSAC algorithm, as described in Section IV-C of the paper.

The variables b_s and b_t of the two-point form can be related to b of the slope–intercept form, since changing b_s and b_t simultaneously corresponds to changing b of the slope–intercept form. The error surface of (b_s, b_t) in Fig. 13(b) shows that the cost values are the same or very similar when both b_s and b_t increase in the top-right direction, which implies that a unique solution cannot be determined if b_s and b_t are involved in a two-variable landscape. Furthermore, multiple local minima around the true parameters are observed. Thus, optimization algorithms are likely to get stuck in local minima, making it even more difficult to correctly estimate the correct ENF frequency.

APPENDIX E

ESTIMATION ACCURACY IMPROVEMENT DUE TO USING MULTIPLE COLUMNS

A. Using Single Column to Estimate Parameters

Let us define a column of smooth native image content as $\mathbf{x}^{(j)} \in \mathbb{R}^{(i_t - i_s + 1) \times 1}$, and an embedding ENF signal as $\mathbf{e}$, where

$$e(i) = \left\{ \frac{a_{t,g} - a_{s,g}}{i_t - i_s} (i - i_s) + a_{s,g} \right\} \cos \left(2\pi \frac{f_g}{f_{\text{row}}} i + \phi_g \right) + k_g^E i + b_g^E. \quad (31)$$

We define $\boldsymbol{\theta}_g = [a_{s,g}, a_{t,g}, f_g, \phi_g, k_g^E, b_g^E]^T$ and $\boldsymbol{\theta} = [a_s, a_t, f, \phi, k^E, b^E]^T$ to be a vector of the ground-truth parameters and search parameters, respectively. The corrupted sensing

signal is therefore $\mathbf{I} = \mathbf{x}^{(j)} + \mathbf{e}$. We set up a cost function to estimate $\boldsymbol{\theta}$ as follows:

$$J_j(\boldsymbol{\theta}) = \text{entropy} \left(\text{hist} \left[\left\{ \mathbf{x}^{(j)}(i) + r(i; \boldsymbol{\theta}_g, \boldsymbol{\theta}) \right\}_{i=1}^M \right] \right) \quad (32a)$$

$$= \text{entropy} \left(\text{hist} \left[\mathbf{x}^{(j)} + \mathbf{r}(\boldsymbol{\theta}_g, \boldsymbol{\theta}) \right] \right), \quad (32b)$$

where $r(i; \boldsymbol{\theta}_g, \boldsymbol{\theta}) = e(i) - \left\{ \frac{a_{t,g} - a_{s,g}}{i_t - i_s} (i - i_s) + a_{s,g} \right\} \cos \left(2\pi \frac{f_g}{f_{\text{row}}} i + \phi_g \right) + k_g^E i + b_g^E$. We denote $\hat{\boldsymbol{\theta}}(\mathbf{x}^{(j)})$ as a numerical suboptimal solution to the optimization problem

$$\min_{\boldsymbol{\theta}} J_j(\boldsymbol{\theta}) \quad \text{s.t. } a_s a_t \geq 0, \quad f \in \{100, 120\}. \quad (33)$$

By taking into consideration the randomness of the image, the suboptimality of the numerical algorithm for finding the optimal solution, we model $\hat{\boldsymbol{\theta}}(\mathbf{x}^{(j)})$ as a random vector with a small bias $\mathbf{b}$ and a positive definite variance-covariance matrix Σ , namely,

$$\mathbb{E} \left[\hat{\boldsymbol{\theta}}(\mathbf{x}^{(j)}) \right] = \boldsymbol{\theta}_g + \mathbf{b}, \quad \text{VarCov} \left(\hat{\boldsymbol{\theta}}(\mathbf{x}^{(j)}) \right) = \Sigma. \quad (34)$$

B. Using Multiple Columns to Improve the Accuracy of the Estimation

We use a random sample $(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(K)})$ to improve the accuracy of the estimation. We use $\hat{\boldsymbol{\theta}}(\{\mathbf{x}^{(j)}\}_{j=1}^K)$ to denote a numerical suboptimal solution to the multi-column optimization problem

$$\min_{\boldsymbol{\theta}} \sum_{j=1}^K J_j(\boldsymbol{\theta}) \quad \text{s.t. } a_s a_t \geq 0, \quad f \in \{100, 120\}. \quad (35)$$

We conduct Taylor-series expansion around $\hat{\boldsymbol{\theta}}(\mathbf{x}^{(j)})$ for $J_j(\boldsymbol{\theta})$, and obtain

$$J_j(\boldsymbol{\theta}) = J_j(\hat{\boldsymbol{\theta}}(\mathbf{x}^{(j)})) + [\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}(\mathbf{x}^{(j)})]^T \nabla J_j(\hat{\boldsymbol{\theta}}(\mathbf{x}^{(j)})) + \frac{1}{2} [\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}(\mathbf{x}^{(j)})]^T \nabla^2 J_j[\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}(\mathbf{x}^{(j)})], \quad (36)$$

where $\nabla^2 J_j$ is a Hessian matrix whose individual components are 2nd-order partial derivatives evaluated near $\hat{\boldsymbol{\theta}}(\mathbf{x}^{(j)})$. We further assume $J_j(\cdot)$ is continuous hence $\nabla^2 J_j$ is a symmetric matrix.

Substituting the Taylor expanded cost function for each column in (36) into the multi-column cost function in (35), and by setting the gradient with respect to $\boldsymbol{\theta}$ to zero, we obtain the optimal analytical solution to (35) as follows:

$$\begin{aligned} & \hat{\boldsymbol{\theta}}(\{\mathbf{x}^{(j)}\}_{j=1}^K) \\ &= \left(\sum_{j=1}^K \nabla^2 J_j \right)^{-1} \left\{ \sum_{j=1}^K [\nabla^2 J_j \hat{\boldsymbol{\theta}}(\mathbf{x}^{(j)}) - \nabla J_j(\hat{\boldsymbol{\theta}}(\mathbf{x}^{(j)}))] \right\} \end{aligned} \quad (37a)$$

$$= \sum_{j=1}^K \mathbf{w}_j \hat{\boldsymbol{\theta}}(\mathbf{x}^{(j)}) - \left(\sum_{j=1}^K \nabla^2 J_j \right)^{-1} \left(\sum_{j=1}^K \nabla J_j \right) \quad (37b)$$

$$\approx \sum_{j=1}^K \mathbf{w}_j \hat{\boldsymbol{\theta}}(\mathbf{x}^{(j)}), \quad (37c)$$

Fig. 13. Error surfaces of a pair of search variables when $B(i)$ is parameterized in (a) the slope-intercept form and (b) the two-point form. For all cases, two search variables vary with the rest search variables set to be the true values. More bluish or darker color indicates lower error. Red star on each error surface represent cost values at the true solution. The error surface of (b_s, b_t) in (b) shows that the cost values are the same or very similar when both b_s and b_t increase in the top-right direction, which implies that a unique solution cannot be determined if b_s and b_t are involved in a two-variable landscape.

where $\mathbf{W}_j = \left(\sum_{j=1}^K \nabla^2 J_j \right)^{-1} \nabla^2 J_j$. The sample average of gradients in (37b) converges to 0 in distribution as K increases. Hence,

$$\mathbb{E} \left[\hat{\boldsymbol{\theta}} \left(\{\mathbf{x}^{(j)}\}_{j=1}^K \right) \right] \approx \boldsymbol{\theta}_g + \mathbf{b}, \quad (38)$$

$$\text{VarCov} \left(\hat{\boldsymbol{\theta}} \left(\{\mathbf{x}^{(j)}\}_{j=1}^K \right) \right) \approx \sum_{j=1}^K \mathbf{W}_j \boldsymbol{\Sigma} \mathbf{W}_j^T. \quad (39)$$

The above results reveal that by using K columns, the estimator has a significantly reduced variance and a similar bias as the single column case.

In the 1-d case, the variance formula is degenerated to $\text{Var}(\hat{\theta}) \approx \left(\sum_{j=1}^K w_j^2 \right) \sigma^2$. When the weight vector is uniform, i.e., $w_i = \frac{1}{K}$, the best variance reduction is achieved, i.e., $\text{Var}(\hat{\theta}) \approx \sigma^2/K$.

APPENDIX F ENTROPY MARGIN BEHAVIOR

For an arbitrary column of an ENF-containing image, $I(i)$, we measure entropy of a random variable Y with

$$H(Y) = - \sum_y p_I(y) \log p_I(y), \quad (40)$$

where y represents pixel intensity with 256 shades of gray, i.e., $y \in [0, 255]$ and $p_I(y)$ is the probability mass function obtained by normalizing the histogram of the column of the ENF-containing image.

We formulate our research problem of identifying whether an image contains 100 or 120 Hz as an optimization problem to solve for the search parameters, $\boldsymbol{\theta} = [a_s, a_t, f, \phi, k^E, b^E]^T$ as follows:

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\text{argmin}} \underbrace{\text{entropy} \{ \text{norm-hist} \{ I(i) - e(i; \boldsymbol{\theta}) \} \}}_{\stackrel{\text{def}}{=} p(y; \boldsymbol{\theta})}, \quad (41)$$

where $p(y; \boldsymbol{\theta})$ is the normalized distribution of the column of the ENF-trace-removed image.

We define the entropy margin as the absolute difference between the two entropy values at the ground-truth ENF f_g and the non-ground-truth ENF f_g^- as follows:

$$\begin{aligned} & |H(p(y; \boldsymbol{\theta}, f_g)) - H(p(y; \boldsymbol{\theta}, f_g^-))| \\ &= \left| \sum_y p(y; \boldsymbol{\theta}, f_g) \log \frac{1}{p(y; \boldsymbol{\theta}, f_g)} - \sum_y p(y; \boldsymbol{\theta}, f_g^-) \log \frac{1}{p(y; \boldsymbol{\theta}, f_g^-)} \right|. \end{aligned} \quad (42)$$

We assume that the image column of our native image has a constant intensity, such that it contains the sole bin of the histogram. Thus, when the estimated ENF trace $\hat{e}(y; \boldsymbol{\theta})$ obtained by the optimization problem (41) is very close to the ground-truth ENF trace, we have

$$\begin{aligned} & \sum_y p(y; \boldsymbol{\theta}, f_g) \log \frac{1}{p(y; \boldsymbol{\theta}, f_g)} \\ &= p(y_i; \boldsymbol{\theta}, f_g) \log \frac{1}{p(y_i; \boldsymbol{\theta}, f_g)} + \sum_{y \neq y_i} p(y; \boldsymbol{\theta}, f_g) \log \frac{1}{p(y; \boldsymbol{\theta}, f_g)} = 0, \end{aligned} \quad (43)$$

where y_i is the pixel location of the sole bin. The first term of (43) equals to 0 by $\log(1)$ and the second term equals 0 by $p(y; \boldsymbol{\theta}, f_g)$ because $p(y; \boldsymbol{\theta}, f_g)$ except when $y = y_i$ is zero. This leads the entropy margin to only depend on $H(p(y; \boldsymbol{\theta}, f_g^-))$ as below:

$$|H(p(y; \boldsymbol{\theta}, f_g)) - H(p(y; \boldsymbol{\theta}, f_g^-))| = \left| \sum_y p(y; \boldsymbol{\theta}, f_g^-) \log \frac{1}{p(y; \boldsymbol{\theta}, f_g^-)} \right| \quad (44)$$

which is the entropy of the ENF-trace-removed column at the non-ground-truth ENF f_g^- , $I(i) - e(i; \boldsymbol{\theta}, f_g^-)$. When the estimated ENF is not equal to the ground-truth ENF, we can view the ENF-trace-removed column as the image column containing two ENF traces characterized by two different ENF frequencies. Thus, we may observe bins spread over some range around the sole bin of the native image column, which contributes to the entropy increase. Also, as the ENF strength gets larger, the bins are spread over a wider range of the pixel intensity, which makes the entropy even higher. This trend has been confirmed by our simulation results in Fig. 9.

For an image that does not contain ENF, the entropy margin can be expressed as

$$\begin{aligned} & |H(p(y; \theta, f = 100)) - H(p(y; \theta, f = 120))| \\ &= \left| \sum_y p(y; \theta, f = 100) \log \frac{1}{p(y; \theta, f = 100)} \right. \\ &\quad \left. - \sum_y p(y; \theta, f = 120) \log \frac{1}{p(y; \theta, f = 120)} \right| \end{aligned} \quad (45a)$$

$$\approx \left| \sum_y p_I(y) \log \frac{1}{p_I(y)} - \sum_y p_I(y) \log \frac{1}{p_I(y)} \right| = 0, \quad (45b)$$

where the conversion from (45a) to (45b) is due to $\hat{e}(y; \theta, f = 100) \approx 0$ and $\hat{e}(y; \theta, f = 120) \approx 0$.

APPENDIX G

CURVE FITTING AS A BASELINE METHOD

We propose a baseline method based on curve fitting, where the pixel intensity signal in (12b) is used as the fitting function. Using the nonlinear least squares fitting in the MATLAB Curve Fitting Toolbox [41] where the trust-region optimization and the bisquare weights fitting methods are used, we fit the function to the averaged pixel values of selected K image columns: $\bar{I}(i) = \frac{1}{K} \sum_{j=1}^K I^{(j)}(i)$, to obtain parameter estimates for f , a_s , a_t , k , ϕ , and b . As in the entropy minimization method, we also consider the constraints on $f \in \{100, 120\}$ Hz and $a_s a_t \geq 0$, thereby leading to generating a total of 4 ($= 2 \times 2$) fitting results given one single image. Among the four fitting results, we choose the frequency showing the best fit RMSE as the final frequency estimate $\hat{f}$.

Similar to the entropy minimization, we use the two-level ENF presence-classification test, where *mse margin* defined as the absolute difference between the MSE values of $f = 100$ and $f = 120$ and the difference between the MSE values of $f = 100$ and $f = 120$ are exploited as test statistics for the first-level and second-level tests, respectively.

APPENDIX H

DETAILED RESULTS FOR IMAGES WITH SYNTHETIC ENF TRACES

To examine the accuracy of the estimators, we calculated Pearson's correlation coefficients between the sinusoidal components of each estimated ENF trace and a true ENF trace. The performance evaluation using the correlation coefficient is shown in Fig. 14 using box plots.² It is revealed from the box plots that as the ENF amplitude level increases, overall, the median of the correlation coefficients is closer to 1 with smaller variance. Under all ENF amplitude levels except the strongest amplitude, the entropy minimization method generally outperforms the baseline method in terms of both the median and the variance of the correlation coefficients.

Fig. 15 illustrates the ROC and the EER performances on images with synthetic ENF traces as a function of the ENF

Fig. 14. Comparison of the performance of the entropy minimization and the baseline methods at different levels of ENF amplitude. The Pearson's correlation coefficients of sinusoidal parts are summarized using box plots. Under all ENF amplitude levels except 16, the entropy minimization method outperforms the curve fitting method in terms of both the median and the variance of the correlation coefficients.

strength. The top row of Fig. 15 shows the ROC curves results when the entropy minimization method applied. Performance of the first-level and second-level classifiers is shown in the left and right panels of Fig. 15. Square, diamond, and triangle markers in the top row's left panel of Fig. 15 are operating points of the first-level classifier with decision thresholds $\eta_1 = 0.002, 0.004$, and 0.006 , respectively. The top row's right panel of Fig. 15 contains results of the second-level classifier of the entropy minimization applied on true positive cases of the first-level classifier.

The top row's left panel of Fig. 15 shows the first-level decision results of the entropy minimization that as the ENF strength increases, the classification performance improves. When the strength exceeds 2.0, the first-level detector starts to perform perfectly. The second-level decision results of the entropy minimization is presented in the top row's right panel of Fig. 15 showing that no matter what the threshold of the first-level classifier is, as long as the strength is more than 0.7, the second-level classifier can almost perfectly distinguish images between 100 Hz and 120 Hz.

The bottom row of Fig. 15 shows the ROC curve results of images with synthetic ENF traces when the baseline method applied. Square, diamond, and triangle markers in the bottom row's left panel of Fig. 15 are operating points of the first-level classifier with decision thresholds $\gamma_1 = 4, 5$, and 6 , respectively. The bottom row's right panel of Fig. 15 contains the ROC results for the second-level classifier of the baseline method applied on true positive cases of the first-level classifier. It is observed that when the ENF strength is less than 8, the first-level detector of the baseline method performs almost as the same as a random guess. When the strength is greater than equal to 8, the EER approaches 0. When the strength is 16, it performs the best, but not entirely perfect. The results for the second-level classifier also demonstrate that the performance is roughly proportional to the ENF strength regardless of the threshold of the first-level classifier. It is noted that only when the strength is 16, the second-level classifier performs perfectly.

²Each box shows the median, and the upper and lower quantiles, and the most extreme data points excluding outliers. The outliers are plotted individually using the "+" symbol.

Fig. 15. ROC curves of the proposed entropy minimization method in the top row and the baseline method in the bottom row applied on images containing synthetic ENF traces with different ENF strengths. Performance of the first-level classifier is shown in the left panel and the performance of the second-level classifier is shown in the right panel. Square, diamond, and triangle markers in the top row's left panel and the in the bottom row's left panel are operating points of the first-level classifier of the entropy minimization with decision thresholds $\eta_1 = 0.002, 0.004, \text{ and } 0.006$ and the first-level classifier of the baseline method with decision thresholds $\gamma_1 = 4, 5, \text{ and } 6$, respectively. The top row's right panel contains results of the second-level classifier of the entropy minimization applied on true positive cases of the first-level classifier. The bottom row's right panel contains results for the second-level classifier of the baseline method applied on true positive cases of the first-level classifier. The entropy minimization significantly outperforms the baseline method and the classification performance of both methods generally improves as the ENF strength increases.

APPENDIX I

DETAILED RESULTS FOR IMAGES WITH REAL ENF TRACES

Under the same image acquisition environment, $\tau\omega_0$ in the denominator of (7) is constant. This allows us to define the ENF strength using the numerator of (7), i.e., $|\text{sinc}(\omega_0 t_{\text{exp}}/\pi)|$. Each scene of the 60 Hz dataset was taken with the exposure time of $[1/30, 1/55, 1/100, 1/150, 1/203, 1/250, 1/311]$ sec corresponding to the ENF strength of $[0, 0.079, 0.156, 0.234, 0.517, 0.662, 0.773]$; each scene of the 50 Hz dataset was taken with the exposure time of $[1/25, 1/46, 1/83, 1/125, 1/170, 1/208, 1/258]$ sec corresponding to the ENF strength of $[0, 0.077, 0.159, 0.234, 0.521, 0.661, 0.771]$. Even though τ and ω_0 are different in each dataset, for fair comparisons, we intentionally chose the above sets of exposure times, such that each set has almost the same ENF strength in terms of the sinc function as follows: $[0.00, 0.08, 0.16, 0.23, 0.52, 0.66, 0.77]$. Each sequence of chosen values leads to monotonically increasing sinc values. Note that setting the exposure time as $1/30$ sec in a 60 Hz country and as $1/25$ sec in a 50 Hz country gives us images including no flickering AC component of ENF traces, i.e., $|\text{sinc}(4)| = 0$. Detailed image capturing steps at the operational level are summarized in the Photos Acquisition Guide available at https://people.engr.ncsu.edu/cwong9/temp/image_acquisition_manual.pdf.

To obtain a similar average brightness on images even with different shutter speeds, ISO sensitivity was varied according

to the shutter speed change. The authors in [42] provided the following formula relating a pixel value in an image, N_d to the camera settings and scene luminance:

$$N_d = K_c \left(\frac{t_{\text{exp}} S}{f_s^2} \right) L_s \quad (46)$$

where K_c , f_s , S , and L_s are the calibration constant, the aperture number (f-stop), the ISO sensitivity, and the scene luminance, respectively. A pixel value represents the light intensity or the brightness at that pixel location of scene. Thus, assuming a uniform color intensity over the entire pixel area and the same scene luminance over different scenery images, it is possible to calculate the ISO sensitivity required to obtain a similar average brightness on images captured with different exposure time lengths. For example, if we use 1800 as the ISO sensitivity for the $1/311$ sec shutter speed, the ISO value for the $1/30$ sec exposure time should be an integer around $30 \times 1800 / 311$. The left half of Fig. 16 shows ten scenes taken from the United States, where the brightness of the electric lighting varies at 120 Hz by the exposure time; for each scene, images with the seven ENF strength levels controlled by the exposure time are shown. As we can see, a higher ENF strength can capture clearer fluctuations of the ENF trace.

Fig. 17 shows representative results of the entropy minimization applied on images with real ENF traces. The left half contains good examples and the right half contains bad examples. Each example contains an ENF-corrupted image,

Fig. 16. Ten scenes from the image dataset taken (a) in the USA with nominal ENF at 60 Hz using iPhone 6s and (b) in China with nominal ENF at 50 Hz using iPhone 6. For each scene, seven ENF strength levels are shown. The strongest ENF traces are observed on the leftmost column, whereas no embedded trace is observed in the rightmost column.

an ENF-trace-removed image, and an estimated sinusoidal-like ENF trace as a function of the row index. Good and bad results are differentiated by whether the ENF trace in an image is completely removed. In bad examples, horizontal stripes caused by the ENF are still largely observable.

Fig. 17. Representative results of the entropy minimization method applied on images with real ENF traces. The left half contains good examples and the right half contains bad examples. Each example contains an ENF-corrupted image, an ENF-trace-removed image, and an estimated sinusoidal-like ENF trace as a function of the row index. Good and bad results are differentiated by whether the ENF trace in an image is completely removed. In bad examples, horizontal stripes caused by the ENF are still largely observable.