Requirements for extracting ENF data to correctly timestamp a video

Niels den Otter
University of Amsterdam
Amsterdam, The Netherlands
notter@os3.nl

Thomas Ouddeken
University of Amsterdam
Amsterdam, The Netherlands
thomas.ouddeken@os3.nl

Supervisor
Zeno Geradts
Netherlands Forensic Institute
The Hague, The Netherlands

***Abstract*—We look at the extraction of Electrical Network Frequency (ENF) data from video. ENF data can be used in timestamping, classically in audio. Recent research has claimed its presence can be detected in video. We especially focus on the properties of complementary metal-oxide-semiconductor (CMOS) sensors and its resulting videos. The CMOS sensor allows for a sampling rate in video that should be high enough to prevent undersampling, and aliasing as a result, when it comes to ENF data. We discuss the properties of a video that should allow for the extraction of ENF although we were not able to conclusively do so ourselves.**

***Index Terms*—ENF, CMOS, Video, aliasing, undersampling, timestamping**

I. Introduction

The Electrical Network Frequency (ENF) is the frequency of the alternating current of the main power grid. This frequency is around 50Hz for the European and Asian power grid and 60Hz for the American power grid. [1] [2] Small fluctuations in the frequency are caused through a varying amount of power consumption by consumers. When recording this frequency over a longer period in time, these fluctuations in frequency become unique enough to do fingerprinting and time stamping. [3] [4] When recording audio, there are often devices nearby that are connected to the main power grid. These devices may generate a so-called mains hum, which can be picked up by microphones. The frequency of this hum is at 50Hz or 60Hz or multiples thereof, depending on the geographical location. This is the sound one often hears from audio systems as static noise, for example in amplifiers. [5] [6] This is mostly used in audio forensics to determine when the audio was recorded or whether or not an audio file has been tampered with.

Recent studies have shown that the same can be done in relation to video recordings, as described in section II. However, this has not been replicated to the best of our knowledge. Our research will aim to first reproduce the experiments that were successful in obtaining ENF data from video recordings and then discuss how its different properties such as frame rate or length affect obtaining ENF data.

II. Related Work

S. Vatansever et al describe the use of so-called superpixels in ENF fluctuation detection in video. [7] A superpixel is a group of pixels with similar characteristics, such as colour and brightness. The authors describe a video analysis method that tries to group similar neighbouring pixels into larger groups to use them to determine if a video is appropriate for further analysis. They do not extract or analyse the actual ENF fluctuations from the videos, but merely try to differentiate between videos that may or may not require further investigation.

D. Nagothu et al used simultaneous recordings of two audio sources to determine whether or not a so-called false frame injection (FFI) attack is taking place. [8] The authors use two baselines, one being an audio recording and the other directly from the power grid. They then make another audio recording at or near the camera. They overlay these with each other to determine whether or not the attacked audio differs. If the ENF fluctuations from the audio source near the video do not match the two baselines, they ascertain that there is an FFI attack taking place. Although this method may prove useful given their scenario, it still heavily relies on the assumption that the attacked audio and the video are directly correlated. If only the video is attacked, the authors do not provide a way to detect it.

In the paper by R. Garg et al they discuss the use of fluorescent light within a static video to extract ENF data. [9] This is done in two ways. The first way is filming a white wall that is illuminated by a fluorescent light. The second way is taking video surveillance footage that contains a static source of light that is visible in most of the recording. They analyse this light by downsampling the recordings to a frequency that matches a multiplication of the local ENF. The variance in brightness is then sampled from the recordings and compared to the baseline power grid ENF. They conclude that this gives a good enough approximation over a long enough period to be able to accurately match the ENF fluctuation with the recordings. They do a very similar experiment using photo diodes, which resulted in similar conclusions.

H. Su et al further elaborate on the research done by R. Garg et al and propose using ENF as a way to synchronise different video sources. [10] They explain how they use the properties of the complementary metal-oxide-semiconductor (CMOS) sensor to their advantage by extracting data per row instead of per frame. This increases the sampling rate significantly.

M. Huijbregtse and Z. Geradts describe how the use of a maximum correlation coefficient is a better way to compare a sample of ENF data to a database than the minimum squared

error. [4] They show how the length of an audio file affects the likelihood of being able to determine the time stamp of the recording. The larger the database becomes, the more likely it is to find similarities within the database, which in turn requires a longer recording to accurately timestamp. We will be using the same method as a means to timestamp our recordings. They also show how the length of a recording directly correlates with the amount of correct estimates regarding the time stamp.

III. Research Questions

We have defined the following research question.

What requirements are there to correctly timestamp a video recording using ENF data?

To answer this question we will have to answer the following sub questions.

A. Is it possible to extract ENF data from video?

B. How do the properties of a video affect the presence of ENF data in a recording?

IV. Methodology

The first requirement is a baseline of ENF data to compare recordings against. The baseline should be gathered using at least one of three methods, preferably two. The first option would be to find an online source that provides the data directly from energy providers connected to the main power grid. This would be a reliable baseline to continue off of. [11] The second option is to retrieve ENF values directly from a power outlet by using a step-down transformer and a voltage divider circuit. It is then possible to record the ENF values by connecting the circuit to the sound card of a PC and recording the received signal. [10] However, we do not have the correct hardware to apply this technique and will therefore not attempt to do this. The third option is using the audio of the recording and analysing it for ENF signals. [4] This can then be verified against the database and together serve as reference points. The size of the database that we will use later on for statistics will be highly relevant, since a larger database is more likely to generate, for example, false positives. It is possible that, when comparing short recordings in a large database, the issue of self similarities becomes prevalent. We have used readily available data for the database. [12]

We subsequently set up a testing environment like the one described by R Garg et al. [9] We attempted to reproduce the white wall method, where a white wall is illuminated by a fluorescent light source. The light source is powered by the 50Hz power grid and thus changes polarity at 100Hz. By recording this over a large span of time we can then analyse the resulting footage for ENF fluctuations. We have recorded and then split one large recording into different sized smaller recordings to get a large test set of data to work with. Figure 1 shows a frame recorded while using the white wall method.

The Cameras that were used are the Eken H9R and Canon 1100D with a Canon Zoom Lens EF-S 18-55mm 1:3.5-5.6 lens. They recorded at 60 and 25 frames per second(fps) respectively. We have focused mostly on the Canon EOS 1100D, since this camera provided the best data, with clearly specified hardware. [13]

Fig. 1. Frame taken from recorded video demonstrating white wall method

V. Under sampling

When determining the frequency of a signal it is important to have a sampling rate that is high enough to properly extract the ENF signal from the data. When observing or sampling a sine wave, the sampling rate should be higher than twice the frequency of said sine wave. [14] Sampling at such a high enough sampling rate prevents aliasing in the resulting observed signal. Aliasing is the phenomenon where an incorrect conclusion is drawn regarding the observed signal. [9] [14] An example can be found in Figure 2, where the dots are moments in time where a sample has been taken. We can clearly see how both sine waves fit the sampled data points. This illustrates how, given the same data, different results can be inferred. However, it is not always possible to reach this rate, due to the technical limitations.

VI. Sampling rate in video

Given the fact that we are attempting to observe a signal of 50Hz, we would ideally use a means of sampling that allows for more than 100Hz sampling. Since we want to record video, this would require a camera that records at more than 100 frames per second. Not all cameras currently in use support such a frame rate. Rather, the most common frame rates are 25 and 30 frames per second. Due to old standards these may vary slightly, namely 24.98 and 29.97. [9]

Cameras with more than 100 frames per second do exist and are known as high speed cameras. These are often used in sports and when slow-motion is required. They could prove valuable in a proof of concept. However, they are rarely used in security footage or home-made footage and they are not commonly used, nor practical in the proposed use cases. [15] [16] [17] This, combined with the nature of the research means that we argue that the use of regular cameras has more value, since this is representative of a real world scenario.

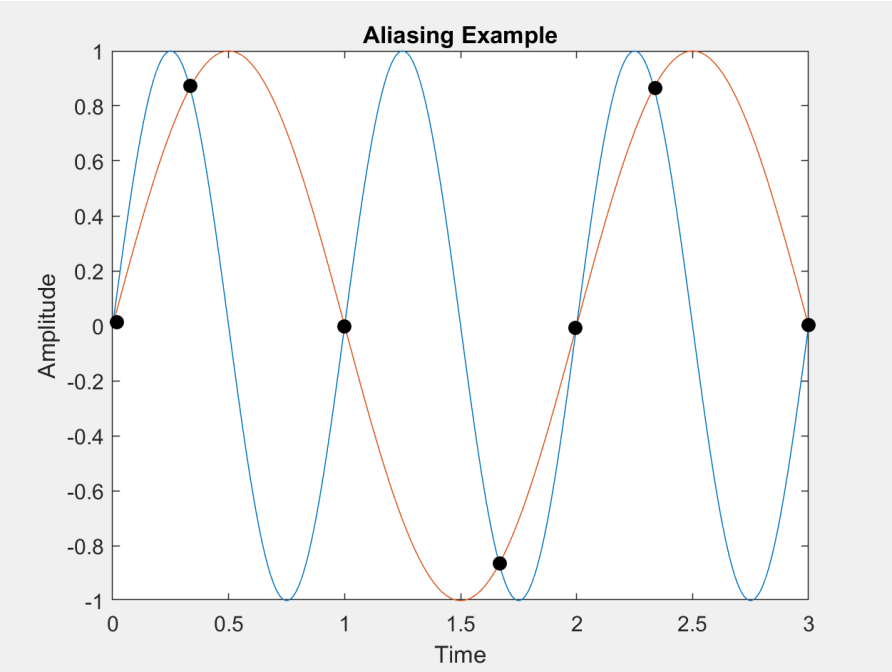

Fig. 2. Example of how under sampling can result in aliasing.

VII. Anti aliasing

When undersampling a signal aliasing occurs, as described in Section V. Because we will be using a camera that operates on a relatively low frame rate we can expect a large amount of aliasing to occur in our signal when looking at the frames as a whole. [9] The observed light source, as described in section IV, operates on the 50Hz power grid and thus flickers at 100Hz. The cameras that were used, operate at 25 and 60 fps. Rarg et al describe how they solve this issue by sampling at specific frequencies. [9] However, they do not mention how they sample at these frequencies, which are different from the frame rate.

VIII. CMOS Sensor

A CMOS sensor is the sensor that registers the properties of pixels. It works unlike a traditional analog camera that uses light sensitive film to capture an image. The traditional image is taken by exposing the light sensitive film to light for a short period of time. This film is exposed in its entirety for the whole duration. [18]

A CMOS sensor, however, registers the light intensity from the top to the bottom, per row. [19] This means that for every frame that is being registered, every row is registered slightly later than the one before. This means that the number of rows registered per second is far greater than the number of frames per second. This behaviour is the cause of the so-called rolling shutter effect, where a fast moving object may seem warped in the image. It increases our sampling rate significantly, theoretically multiplying it by the number of rows of pixels in a frame, which in our case increases the sampling rate from $25fps$ to $720 * 25 = 18000fps$. This is given that the sequential reading happens at a linear rate. There may be a slight delay between the last row of one frame and the first row of the following frame, depending on the sensor. This may cause phase shifting, but should not be too much of an issue when attempting to gather ENF data. [10]

The Canon EOS 1100D uses a 12.2 megapixel 22.2mm x 14.7mm APS-C CMOS sensor. [13] CMOS sensors can have different sizes. However, all CMOS sensors should be sufficient to prevent undersampling. The combination of the resolution and the size of the sensor decide the eventual sampling rate that can be achieved, along with the frame rate. Taking the mean value of intensity for every pixel in a row, for every row in every frame of a video recording could produce ENF data. An example of gathered data is shown in Figure 3. We suspect more filtering of the original video data needs to be done in order to extract the correct ENF data, if it is possible.

IX. Type of recording

The type of recording is important for our suggested method of obtaining ENF data from video that was captured using a CMOS sensor. The footage must contain either a lamp or a surface area illuminated by a lamp that flickers at a multiplication of the mains power grid frequency. One such example would be a fluorescent lamp. This lamp or surface area must be clearly visible for a prolonged period of time, preferably over 240 seconds. Preferably the camera is not being moved during this period. This allows for all other variables to be as constant as possible, which makes it more likely to extract the correct ENF data from the footage.

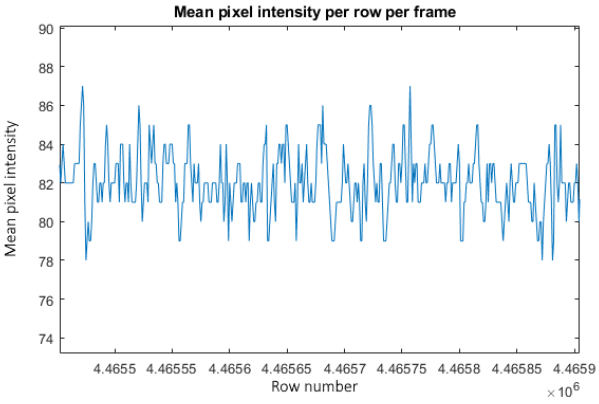

Fig. 3. Pixel intensity retrieved by taking mean values per row per frame

Fig. 4. Frame taken from a post pre-processing sample

X. Filtering

According to literature, it is important to preprocess the video before trying to extract ENF data. [10] The authors made use of a blur and high-pass filter. We have done so as well, which resulted in the still of the pre-processed video shown in Figure 4. This should amplify the presence of ENF data and reduce any noise that is present. The resulting video can then be analysed.

XI. Randomness of ENF

ENF is sufficiently random that, given a long enough sample of audio, we can uniquely timestamp this using an ENF database. [20] [4] Generally speaking, the shorter the sample, the more likely it is that we find more than one candidate that may correspond to the sample, which holds for both audio and video. [4]

XII. Conclusion

Because the properties of a CMOS sensor allow for a high enough sampling rate to prevent undersampling, we argue that the following properties affect the collection of ENF data from video. Because of the way in which a CMOS sensor sequentially reads the data per row, we can say that the sampling rate is directly affected by the size of the sensor and the resolution. We argue that as long as the frames per second multiplied by the number of rows in a frame allow for a sufficient sampling rate, it should be possible to extract ENF data from video. Logically, we theorise that the larger the sensor and the larger the resolution, the more data there is and thus the more likely it is that the data is correct. The length of a recording directly correlates with the likelihood of correctly timestamping it, where the longer the recording the higher the chance of a correct timestamp. [20] [4] We theorise that there should be no significant difference between ENF data extracted from audio and that from video, when successful. Thus the properties that hold for ENF data from audio should also hold for video.

XIII. Discussion

We were not able to entirely reproduce the research by R. Garg et al. [9] We are of the opinion that the paper is not clear enough in describing the process they use to collect the ENF data from video. However, when combining the research that was done by H Su et al, we were able to extract data that may resemble ENF data. We were not able to conclusively do so and timestamp this correctly. This has mainly been due to time constraints. If one were to be aware of and try to avoid the possibility of gathering ENF data from video, circumventing it would be trivial. Using light that does not flicker at a multiplication of the main power grid frequency, using a camera that does not use a CMOS sensor or even not filming the light for a long enough period of time would be good ways to prevent time stamping video using ENF data. Combining footage from multiple cameras and using ENF to show correlation between the footage and their timestamps may prove useful as well, without needing a reference database for example. We have not taken into account how other factors such as the ISO value or possible white balancing and the way our device handles the video. These may also influence the data and results in a meaningful way.

XIV. Future work

For future work we would be interested to see if a tool can be made, possibly in combination with the superpixel based approach as proposed by S Vatansever et al. A video could be automatically scanned to see if it would contain ENF data and subsequently extract this data. The superpixel based approach would allow for better analysis of moving footage, where the regions of interest can be isolated and analysed. It would be interesting to look at the difference and influence that filtering and pre-processing has on the footage that is being used. Since we theorize that it should be possible to extract ENF data from video footage recorded using a CMOS sensor, it would be interesting to look into how, when altering aspects like the sensor size or resolution, this affects the data and at what point it is no longer possible to extract useful data. Using the maximum correlation coefficient as shown by M. Huijbregtse and Z. Geradts it would be interesting to do a large scale statistical research where one could accurately determine the minimal length at which a sample can no longer accurately be timestamped. Another interesting aspect would be forging ENF data into footage, this might lead to incorrect timestamping if no other information was available.

REFERENCES

[1] P. T. de Boer, "Accuracy and stability of the 50 hz mains frequency." https://wwwhome.ewi.utwente.nl/ ptdeboer/misc/mains.html, 2005. Accessed on 03.06.2020.

[2] M. Stolworthy, "Power grid frequency." https://gridwatch.co.uk/frequency, 2020. Accessed on 04.06.2020.

[3] D. Chowdhury and M. Sarkar, "Location forensics analysis using enf sequences extracted from power and audio recordings," December 2019.

[4] M. Huijbregtse and Z. Geradts, "Using the enf criterion for determining the time of recording of short digital audio recordings," in *Computational Forensics*, pp. 116–124, Springer Berlin Heidelberg, 2009.

[5] K. Hashemi, "Mains hum." http://www.opensourceinstruments.com/Electronics/A3013/HTML/Hum.html, 2018. Accessed on 16.06.2020.

[6] M. Andrei, "Why does electricity hum — and why is it a b flat in the us, and a g in europe?." https://www.zmescience.com/other/feature-post/why-electricity-hum-07112017/, February 2020. Accessed on 16.06.2020.

[7] S. Vatansever, A. E. Dirik, and N. Memon, "Detecting the presence of enf signal in digital videos: A superpixel-based approach," *IEEE Signal Processing Letters*, vol. 24, pp. 1463–1467, August 2017.

[8] D. Nagothu, Y. Chen, E. Blasch, A. Aved, and S. Zhu, "Detecting malicious false frame injection attacks on surveillance systems at the edge using electrical network frequency signals," *Sensors*, vol. 19, p. 2424, 05 2019.

[9] R. Garg, A. L. Varna, and M. Wu, ""seeing" enf: Natural time stamp for digital video via optical sensing and signal processing," in *Proceedings of the 19th ACM International Conference on Multimedia*, MM '11, (New York, NY, USA), p. 23–32, Association for Computing Machinery, November 2011.

[10] 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 *Proceedings of the 2nd ACM International Workshop on Immersive Media Experiences*, ImmersiveMe '14, (New York, NY, USA), p. 13–18, Association for Computing Machinery, November 2014.

[11] Mains Frequency, "Measurement of the mains frequency." https://www.mainsfrequency.com, April 2019. Accessed on 04.06.2020.

[12] National Grid ESO, "Historic frequency data." https://www.nationalgrideso.com/balancing-services/frequency-response-services/historic-frequency-data, February 2020. Accessed on 10.06.2020.

[13] Canon, ""canon eos 1100d specifications"." https://www.canon-europe.com/support/consumer_products/products/cameras/digital_slr/eos_1100d.html?type=specifications", 2011. Accessed on 04.06.2020.

[14] S. Ruzin, "Capturing images." http://microscopy.berkeley.edu/courses/dib/sections/02Images/sampling.html, 2009. Accessed on 10.06.2020.

[15] Dicsan, "Security camera features - frame rate." https://dicsan.com/Security_Cameras/security_cameras_frame_rate/, 2020. Accessed on 16.06.2020.

[16] IPVM, "Frame rate guide for video surveillance." https://ipvm.com/reports/frame-rate-surveillance-guide, August 2014. Accessed on 16.06.2020.

[17] CCTV Camera World, "Frame rate vs. bandwidth." https://www.cctvcameraworld.com/ip-cameras-frame-rate-bandwidth/, October 2020. Accessed on 16.06.2020.

[18] K. Keller, H. Kampfer, R. Matejec, O. Lapp, W. Krafft, H. Frenken, H. Lührig, R. Scheerer, M. Heilmann, H. Meckl, P. Bergthaller, D. Hübner, E. Wolff, B. Morcher, W. Zahn, H. Buschmann, R. Blank, R. Tromnau, J. Plamper, A. Seiler, K. Nieswandt, I. Boie, E. Moisar, R. Winiker, M. Schellenberg, and L. Ketellapper, *Photography*. American Cancer Society, June 2000.

[19] M. W. Davidson, "Introduction to cmos image sensors." https://micro.magnet.fsu.edu/primer/digitalimaging/cmosimagesensors.html, November 2015. Accessed on 24.06.2020.

[20] C. Grigoras, "Applications of enf criterion in forensic audio, video, computer and telecommunication analysis," *Forensic science international*, vol. 167, pp. 136–45, 05 2007.