FACTORS AFFECTING ENF BASED TIME-OF-RECORDING ESTIMATION FOR VIDEO

Saffet Vatansever[1,2] *Ahmet Emir Dirik*[3] *Nasir Memon*[4]

[1] Department of Mechatronics Engineering, Bursa Technical University, Bursa, Turkey
[2] Department of Electrical and Electronic Engineering, Uludag University, Turkey
[3] Department of Computer Engineering, Faculty of Engineering, Uludag University, Bursa, Turkey
[4] Department of Computer Science and Engineering, New York University, Brooklyn, NY, USA

ABSTRACT

ENF (Electric Network Frequency) oscillates around a nominal value (50/60 Hz) due to imbalance between consumed and generated power. The intensity of a light source powered by mains electricity varies depending on the ENF fluctuations. These fluctuations can be extracted from videos recorded in the presence of mains-powered source illumination. This work investigates how the quality of the ENF signal estimated from video is affected by different light source illumination, compression ratios, and by social media encoding. Also explored is the effect of the length of the ENF ground-truth database on time of recording detection and verification.

***Index Terms*—** ENF, electric network frequency, video forensics, time-of-recording

1. INTRODUCTION

In the modern era, images and videos can straightforwardly be created, edited or manipulated. Consequently, developing forensic tools for various tasks, e.g. forgery detection [1], [2], watermarking [3], [4] camera attribution [5], [6] etc., has become a significant field of research. ENF (Electric Network Frequency) based forensic tools are among those emerging.

ENF oscillates about a nominal value (50/60 Hz) due to instantaneous imbalance between consumed and generated power [7]. The oscillation across the whole network at any time instance is expected to be the same, which consequently makes the ENF measured from any power outlet at a particular time instance unique to the network [8], [9]. The luminous intensity of a light source powered by mains electricity alternates in parallel with ENF fluctuations, except for the fact that the frequency of the illumination is double the nominal ENF, i.e., it illuminates in both positive and negative cycles of electric voltage. The alterations in the luminance are intrinsically embedded in video recordings, and ENF can hence be extracted through steady content analysis of these videos [10], [11], [12], [13], [14], [15], [16].

Ambient conditions in the scene, and data properties may have a noticeable impact on the quality of the ENF signal to be estimated from a video. Consequently, the performance of ENF based forensic applications including time-of-recording verification [11], [16], multimedia synchronization [17], [18], camera characterization [14], [16], and ENF presence detection [15] may be affected by these conditions. This work investigates how the performance of ENF based time-of-recording detection and verification is affected by different illumination sources, different compression ratios, and different lengths of the reference ENF data (ground-truth).

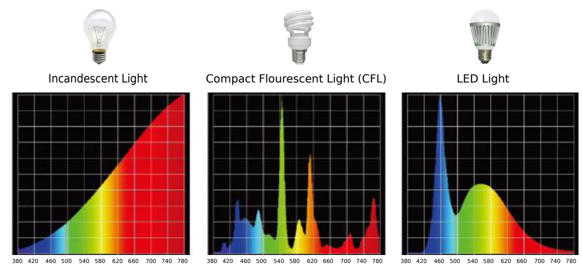

Fig. 1. Spectra of various light sources [19]

Table 1. Different light types used in the recording scene

Source No	Type	Lum. Flux	Color Temperature
S1	Halogen	834 (lm)	2800 (K)
S2	CFL	870 (lm)	6500 (K)
S3	CFL	840 (lm)	2700 (K)
S4	LED	810 (lm)	6500 (K)
S5	LED	810 (lm)	2700 (K)

2. IMPACT OF ILLUMINATION SOURCE ON ENF

This section investigates how the type of mains-powered illumination source affects ENF signal estimation. Fig. 1 illustrates emission spectra of various commonly used light sources, namely Incandescent, white CFL (Compact Fluorescent), and white LED (Light-emitting Diode) [19]. From the figure, it can be seen that Incandescent tungsten has a lower level of power in blue light, though it has the greatest power in red. CFL has a relatively lower level of power across the spectrum except for some distinguishable spikes, including green and red. LED provides the highest power in blue light, though

[3] edirik@uludag.edu.tr (corresponding author)

Table 2. Time-of-recording detection: AUC values for different light sources

Clip Length (min.)	S1	S2	S3	S4	S5
1	0.79	NA	0.32	0.45	0.65
2	0.74	0.48	0.79	0.94	0.91
5	1.00	0.85	0.87	1.00	0.99
10	1.00	1.00	1.00	1.00	1.00

relatively lower power in the red. Given these differences in the power of wavelengths across the visible spectrum for different light sources, in the rest of this section we investigate how the type of light source may affect the quality of the estimated ENF signal, and consequently, the ENF based time-of-recording detection and verification. All experiments in the paper are performed by using wall-scene videos recorded under the different types of illumination by a fixed camera.

2.1. Estimation of ENF

The procedure for ENF signal estimation in this work is composed of three consecutive steps. First, the time-series of intensity variations along the video is constructed by averaging all the pixels per frame, leading to one illumination sample per frame [10]. Next, STFT (Short Time Fourier Transform) of these intensity variations is performed, followed by quadratic interpolation [20]. For STFT computation, 20 seconds time windows with 19 seconds overlaps are utilized, consequently resulting in a 1-second temporal ENF resolution [21]. A mains-powered light source peaks at both positive and negative cycles of AC current, hence the main illumination frequency is double the nominal ENF. It is also notable that the sampling frequency of the above-mentioned time-series is the same as the video frame rate. Hence, Nyquist theorem [22] is not satisfied and ENF peak search in the Fourier domain has to be made around the alias frequency.

2.2. Evaluation Metrics

The primary metric used for video time-of-recording detection and verification is the supremum of normalized cross-correlation (NCC). By using this metric, the estimated ENF signal is searched in a reference ENF database (ground-truth), and the lag point of the maximum NCC coefficient is recorded. If the time lag at the peak correlation coefficient corresponds to the time difference from the beginning time of the reference ENF data, the day-and-time of the video recording is verified. Otherwise, it is not verified. These cases correspond to the following binary hypotheses:

H0: Lag point of the maximum correlation coefficient between the estimated video ENF and the reference ENF (ground-truth) does not match with the video recording time.

H1: Lag point of the maximum correlation coefficient between the estimated video ENF and the reference ENF (ground-truth) matches with the video recording time.

Table 3. The rates of correct time-of-recording estimations (%) for different light sources

Clip L. (min.)	S1	S2	S3	S4	S5
1	2.25	0.00	0.50	26.37	20.00
2	100	51.75	47.25	100	97.75
5	100	100	100	100	100
10	100	100	100	100	100

When an estimated ENF vector is searched in a reference ENF data which does not include the correct time period of the video, i.e. if it is the false reference ENF data, the lag point of the peak NCC coefficient intrinsically corresponds to the H0 case. Nevertheless, when searched on a reference ENF data including the correct time period of the video, i.e. if it is true reference ENF data, then the lag point of the peak NCC coefficient may correspond to either H1 case (if the lag point matches with the video time) or H0 case (if the lag point does not match with the video time although the reference ENF database includes the ground truth ENF data of the query video). To measure the performance of the ENF signal detection (localization) in a reference database, ROC (Receiver Operating Characteristic) curves and AUC (Area Under the Curve) are computed with the use of the above hypotheses for various light sources, compression rates, and database search lengths. In order to form the ROC curves, each estimated video ENF signal is first searched on both true reference ENF database (for H1 and H0) and false reference ENF database (just for H0), and the peak correlation coefficient is recorded.

2.3. Experimental Setup and Results

In our experiments, five different types of light sources were used for illumination. Some specifications of these sources are provided in Table 1. Accordingly, S2 and S3 are white and yellow CFL; S4 and S5 are white and yellow LED, respectively. S1 is Halogen, which provides yellow color. Although, not included in Fig. 1, there are almost no spectral differences between the incandescent and the halogen lamps.

For each type of light source in Table 1, 4 still-scene (wall) videos of about 10 minutes lengths were recorded at 30 fps with 640×480 resolution by a Canon PowerShot SX210 IS model CCD camera. Then, each recorded video was divided into 1, 2, 5, and 10-minute clips yielding 10, 5, 2 and 1 video clips of the respective lengths. In all, 40, 20, 8, and 4 video clips of length 1, 2, 5, and 10-minutes were obtained for each type of illumination source. For each clip, the ENF signal was computed. For each estimated ENF, NCC is computed with both a 24-hour true reference ENF data (involving the recording time of the query video) and a 24-hour false reference ENF data (not involving the the recording time of the video). The NCC operation for each video ENF vector was repeated 20 times by taking the start point of the reference database 1 hour back or forward. That is, the start point and end point

Table 4. Time-of-recording detection: AUC values for different compression rates for LED

Bit Rate	Average Size	2 min.	5 min.	10 min.
Original	938 MB	0.94	1.00	1.00
5000 Kbps	415 MB	0.94	1.00	1.00
1000 Kbps	94 MB	0.93	1.00	1.00
500 Kbps	53 MB	0.91	1.00	1.00
100 Kbps	20 MB	0.84	0.62	0.77
Facebook	13 MB	0.57	0.77	0.68

Table 5. The rates of correct time-of-recording estimations (%) for different compression rates for LED

Bit Rate	Average Size	2 min.	5 min.	10 min.
Original	938 MB	100	100	100
5000 Kbps	415 MB	100	100	100
1000 Kbps	94 MB	100	100	100
500 Kbps	53 MB	100	100	100
100 Kbps	20 MB	19.75	87.50	100
Facebook	13 MB	12.25	56.88	100

of the database were changed in each test. Then, ROC were obtained for each light source type based on peak NCC distributions of H1 and H0 cases. Corresponding AUC values of the ROC curves for each light source are given in Table 2.

It can be seen that S4 (white LED) and S5 (Yellow LED) outperform other sources for all lengths of the ENF signal. S2 (white CFL) and S3 (Yellow CFL) show relatively worse performance. The reason for the low value of the AUC for short length ENF signals is false matches due to the similar variations of the successive ENF samples in time. ENF signals of greater length reduce the possibility of false matches, yielding greater AUC values. Table 3 provides the rate (in %) of correct time-of-recording estimation when ENF signal is searched only in true reference database, i.e., the task of time-of-recording verification. As can be seen from the table, the results are parallel to those presented in Table 2. Consequently, different light sources contribute to the quality of the estimated ENF signal differently. ENF is best estimated for videos recorded under LED illumination, though the quality of the ENF signal significantly drops under CFL illumination. For the rest of the paper, all the experiments were conducted using the light sources that lead the best and the worst performances, i.e. white LED and CFL only.

3. THE EFFECT OF COMPRESSION ON ENF

In this section, the effect of video compression on the ENF signal is investigated. Of the videos used in Section 2.3, those captured under white LED (S4 in Table 2) and white CFL (S2 in Table 2) were compressed at various bit rates; specifically

Table 6. Time-of-recording detection: AUC values for different compression rates for CFL

Bit Rate	Average Size	2 min.	5 min.	10 min.
Original	973 MB	0.48	0.85	1.00
5000 Kbps	430 MB	0.55	0.89	1.00
1000 Kbps	98 MB	0.66	0.89	1.00
500 Kbps	55 MB	0.11	0.66	1.00
100 Kbps	20 MB	failed	failed	0.98
Facebook	14 MB	failed	failed	failed

Table 7. The rates of correct time-of-recording estimations (%) for different compression rates for CFL

Bit Rate	Average Size	2 min.	5 min.	10 min.
Original	973 MB	51.75	100	100
5000 Kbps	430 MB	61.00	100	100
1000 Kbps	98 MB	42.50	100	100
500 Kbps	55 MB	6.00	100	100
100 Kbps	20 MB	0	0	51.25
Facebook	14 MB	0	0	0

5000 Kbps, 1000 Kbps, 500 Kbps and 100 Kbps with the use of FFMPEG via the H.264 compression standard. In addition, a Facebook compressed form of each video was created by sequentially uploading to and downloading from its website. For each compression type, the original video resolution was maintained. Next, each of the compressed videos was divided into 2, 5 and 10-minute clips and peak NCC distributions for H0 and H1 cases were obtained in the same way as described in Section 2.3, followed by the computation of the ROC curves. Table 4 provides AUC values for LED case. As can be observed from the Table, up to a rate of 500 Kbps, the performance of time-of-recording detection is quite well for almost all size of video clips. Compression at 100 Kbps and Facebook uploads decreased the detection performance noticeably. Table 5 provides the rate (in %) of correct matches, when the estimated ENF signals were searched in the true reference database only. Accordingly, similar trend in the results are seen, except for the performance for 10-minute clips with 100% success rate for all types of compression.

Table 6 and Table 7 respectively provide AUC values and the rate (in %) of correct matches for CFL case. Although, the time-of-recording detection performance for compression with 5000 Kbps, and 1000 Kbps is almost stable over all sizes of video clips, compression at 100 Kbps and Facebook uploads cause a total failure in the performance of both detection (Table 6) and verification (Table 7), for 2-minute and 5-minute clips. Although Facebook compression causes a failure for 10-minute clips, verification performance for compression with 100 Kbps is acceptable with a 51.25% true de-

Table 8. Time-of-recording detection: AUC values for different lengths of ground-truth ENF data for LED

Database Length	2 min.	5 min.	10 min.
One-day	0.94	1.00	1.00
One-week	0.91	1.00	1.00
One-month	0.82	1.00	1.00

Table 9. The rate of correct time-of-recording estimations (%) for different lengths of ground-truth ENF data for LED

Database Length	2 min.	5 min.	10 min.
One-day	100	100	100
One-week	100	100	100
One-month	100	100	100

tection rate (Table 7). The high AUC value, 0.98, for compression with 100 Kbps in the detection task, Table 6, is resulted because the values of H0 and H1 cases are obtained considerably separate from each other, despite the number of H1 cases, i.e. correct matches are only a half of the H0 cases. Consequently, the verification task clarifies some unexpected AUC values by providing the number of correct matches.

4. EXPERIMENTS WITH GROUND-TRUTH ENF OF DIFFERENT LENGTHS

This section explores how the length of the reference ENF data (ground-truth), in which the estimated ENF signal is searched, affects the time-of-recording estimation depending on the length of the estimated ENF vector. Of the videos in Section 2.3, those captured under white LED and white CFL were divided into 2, 5 and 10-minute clips. ENF signal was extracted from each video clip and searched within one-day, one-week, and one-month length correct and false reference ENF data. Each search was repeated 20 times by taking the start point of the reference database signal 1 hour, 8 hour, and 24 hours back or forward respectively for one-day, one-week and one-month length reference data. That is, start point and end point of the databases was changed in each test as in Section 2.3. Next, peak NCC values for H0 and H1 cases were obtained as described in Section 2.3, followed by the computation of ROC curves. Table 8 provides AUC values for LED cases. Accordingly, detection performance drops as the length of the reference ENF data increases for 2-minute videos. For 5-minute and 10-minute videos, detection performance is quite stable for all lengths of the reference data, yielding AUC values of 1.00. Table 9 provides rates (in %) of correct estimations of time-of-recordings obtained when the estimated ENF signals are only searched in the true reference database. Surprisingly, 100% success rate is obtained for all lengths of video clips and all lengths of reference data.

Table 10 and Table 11 respectively provide AUC and the rate of correct detection (in %) for CFL light. Accordingly,

Table 10. Time-of-recording detection: AUC values for different lengths of ground-truth ENF data for CFL

Database Length	2 min.	5 min.	10 min.
One-day	0.48	0.85	1.00
One-week	0.45	0.80	1.00
One-month	0.38	0.75	1.00

Table 11. The rates of correct time-of-recording estimations (%) for different lengths of ground-truth ENF data for CFL

Database Length	2 min.	5 min.	10 min.
One-day	51.75	100	100
One-week	27.50	100	100
One-month	11.00	98.13	100

overall, the performance rates of both detection and verification decrease as the length of the reference ENF data increases for 2-minute videos. Although the detection performance slightly drops for 5-minute videos, the verification performance is quite stable for all lengths of the ENF reference data. For 10-minute videos, the performance rates are very stable, leading to 100% estimation rate, for all lengths of video (2,5,10 minutes) and reference data.

5. CONCLUSION

In this work, factors affecting ENF based time-of-recording estimation for digital video, i.e. type of light source, video compression rate, length of video, and length of ENF database, is investigated with experimental analysis. It is observed that different type of light sources affect the quality of the estimated ENF signal differently. Although ENF estimation and time-of-recording verification are quite robust for 500 Kbps compression rates, the performance of the verification varies depending on the type of light source. For instance, best verification results are observed under the white LED illumination. On the other hand, CFL yields the worst results according to the experimental analysis. Another observation is that for two minutes Facebook videos, ENF based video time-of-recording verification fails for both CFL (0%) and LED cases (12%) in the experiments.

This work may pave the way for the application of light source detection and verification. Such an application may also inspire research in ENF based tamper detection by doing ENF analysis for different contents in video.[1]

[1]Acknowledgement: This material is based on research sponsored by DARPA and the Air Force Research Laboratory (AFRL) under agreement number FA8750-16-2-0173. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of DARPA and the Air Force Research Laboratory (AFRL) or the U.S. Government.

6. REFERENCES

[1] G. Naumovich and N. Memon, “Preventing piracy, reverse engineering, and tampering,” *Computer*, vol. 36, no. 7, pp. 64–71, July 2003.

[2] S. Bayram, H. T. Sencar, and N. Memon, “An efficient and robust method for detecting copy-move forgery,” in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, April 2009, pp. 1053–1056.

[3] N. Mettripun, T. Amornraksa, and E. J. Delp, “Robust image watermarking based on luminance modification,” *Journal of Electronic Imaging*, vol. 22, pp. 22 – 22 – 16, 2013.

[4] S. Tachaphetpiboon, K. Thongkor, T. Amornraksa, and E. J. Delp, “Digital watermarking for color images in hue-saturation-value color space,” *Journal of Electronic Imaging*, vol. 23, pp. 23 – 23 – 14, 2014.

[5] J. Fridrich, “Digital image forensics,” *IEEE Signal Processing Magazine*, vol. 26, no. 2, 2009.

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

[7] M. H. Bollen and I. Y. Gu, *Signal processing of power quality disturbances*, Wiley-Interscience, 2006.

[8] C. Grigoras, “Digital audio recording analysis–the electric network frequency criterion,” *International Journal of Speech Language and the Law*, vol. 12, no. 1, pp. 63–76, 2005.

[9] S. Vatansever and A. E. Dirik, “Forensic analysis of digital audio recordings based on acoustic mains hum,” in *2016 24th Signal Processing and Communication Application Conference (SIU)*, May 2016, pp. 1285–1288.

[10] R. Garg, A. L. Varna, and M. Wu, “Seeing ENF: natural time stamp for digital video via optical sensing and signal processing,” *Proceedings of the 19th ACM international conference on Multimedia*, pp. 23–32, 2011.

[11] 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 Transactions on Information Forensics and Security*, vol. 8, no. 9, pp. 1417–1432, 2013.

[12] H. Su, A. Hajj-Ahmad, R. Garg, and M. Wu, “Exploiting rolling shutter for ENF signal extraction from video,” in *2014 IEEE International Conference on Image Processing (ICIP)*, 2014, pp. 5367–5371.

[13] M. Wu, A. Hajj-Ahmad, and H. Su, “Techniques to extract enf signals from video image sequences exploiting the rolling shutter mechanism; and a new video synchronization approach by matching the enf signals extracted from soundtracks and image sequences,” December 2015, US Patent US9916857B2.

[14] A. Hajj-Ahmad, A. Berkovich, and M. Wu, “Exploiting power signatures for camera forensics,” *IEEE Signal Processing Letters*, vol. 23, no. 5, pp. 713–717, 2016.

[15] 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, no. 10, pp. 1463–1467, Oct 2017.

[16] S. Vatansever, A. E. Dirik, and N. Memon, “Analysis of rolling shutter effect on enf based video forensics,” *IEEE Transactions on Information Forensics and Security*, pp. 1–1, 2019.

[17] 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*, 2014, pp. 13–18.

[18] H. Su, A. Hajj-Ahmad, M. Wu, , and D. W. Oard, “Exploring the use of ENF for multimedia synchronization,” *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, pp. 4613–4617, 2014.

[19] UPRtek, “Spectrum measuring,” [Online]. Available: https://www.uprtek.com/en/application/Spectrum-Measuring/Case-Study-Spectrum-Measuring.html. [Accessed: 2019-02-17].

[20] C. Grigoras, “Applications of ENF criterion in forensic audio, video, computer and telecommunication analysis,” *Forensic Science International*, vol. 167, no. 2-3, pp. 136–145, 2007.

[21] A. Cooper, “The electric network frequency (ENF) as an aid to authenticating forensic digital audio recordings–an automated approach,” *AES 33rd International Conference*, pp. 1–10, 2008.

[22] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing: Principles, Algorithms, And Applications, 4/E*, Pearson Education, 2007.