ENF Signal Induced by Power Grid: A New Modality for Video Synchronization

Hui Su
University of Maryland
College Park, USA
hsu@umd.edu

Adi Hajj-Ahmad
University of Maryland
College Park, USA
adiha@umd.edu

Chau-Wai Wong
University of Maryland
College Park, USA
cwwong@umd.edu

Ravi Garg
University of Maryland
College Park, USA
ravigarg.iitm@gmail.com

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

ABSTRACT

Multiple videos capturing the same scene from possibly different viewing angles may be synthesized for novel immersive experience. Synchronization is an important task for such applications involving multiple pieces of audio-visual data. In this work, we exploit the electric network frequency (ENF) signal inherently embedded in the soundtrack and/or image sequence of video to temporally align video recordings. ENF is the supply frequency of power distribution networks in a power grid. Its value fluctuates slightly from its nominal value of 50 Hz or 60 Hz, and the fluctuation trends stay consistent within the same grid. Audio and video recordings that are created in areas of electric activities may capture the ENF signal due to electromagnetic interferences and other physical phenomena. We propose to synchronize video recordings by aligning the embedded ENF signals. Without major constraints on viewing angle and camera calibration as many existing methods impose, the proposed approach emerges as a new synchronization modality.

Categories and Subject Descriptors

I.2.10 [**Computing Methodologies**]: Artificial Intelligence—*vision and scene understanding*; H.5.1 [**Information System**]: Information Interfaces and Presentation—*multimedia information systems*

General Terms

Algorithms, Experimentation.

Keywords

Electric Network Frequency; Audio/Video Synchronization; Multi-modal Processing; Rolling Shutter

1. INTRODUCTION

When an event is recorded simultaneously by multiple independent video cameras and possibly from a variety of an-

gles, fusing the information in these videos may provide a better presentation and novel immersive experience of the event than each recording alone. Using 3D reconstruction techniques, a dynamic scene may be reconstructed from multiple video streams that allows people to choose from different viewing angles of a scene. Several videos of various perspective can be "stitched" together to achieve wider field of view via video panorama [1]. A video sequence of high space-time resolution can be obtained by combining information from multiple low-resolution video sequences [2]. To facilitate these and other tasks involving multiple pieces of video data, the individual video sequences often need to be synchronized before synthesis. Video synchronization therefore becomes an important problem, and the solution to it can enable and enhance existing and potentially new immersive media applications.

In professional video productions such as sports TV broadcasting, the recording cameras may be synchronized based on coordinated hardware and communication protocols to provide synchronized timestamps and ensure accurate temporal alignment. For distributed and ad-hoc settings involving consumer-level devices, different cameras' clocks are not easily synchronized to the frame level. In absence of proactive synchronization mechanisms, the current solutions have to rely primarily on visual content and/or sound content [3, 4, 5, 6, 7], and may not always work well. For example, it is difficult to synchronize video sequences using visual features when they do not share a sufficient amount of common view of the scene, or the viewing angles are significantly different and the cameras are not calibrated beforehand.

In this paper, we propose a new modality for video synchronization by exploiting the electric network frequency (ENF) signal inherently embedded in video recordings. ENF is the supply frequency of power distribution networks in a power grid. The nominal value of the ENF is usually 60Hz (in North America) or 50Hz (in most other parts of the world). The instantaneous value of the ENF typically fluctuates slightly around its nominal value as a result of the interaction between power consumption variations and the control mechanisms of the power grids. The main trends in the fluctuations of ENF are very similar within the same power grid, even for distant locations [8]. The sequence of values of instantaneous ENF over time is regarded as the ENF signal. The ENF signal can be extracted from a power signal measured at a power outlet through a step-down transformer and a voltage divider circuit. Audio recordings created using de-

vices plugged into the power mains or battery-powered near electrical devices can capture the ENF signal due to electromagnetic interferences or acoustic vibrations such as electric humming [8]. More recently, it is found that video cameras are also capable of capturing ENF signals due to the flickering in indoor lightings caused by changes in supply voltage [9]. Several forensic applications have been proposed based on the analysis of ENF signals [10, 11, 12].

Viewed as a continuous random process over time, the ENF signal embedded in audio and video signals can be used as a timing fingerprint that is unique at any specific time instance. We propose to match the ENF signals extracted from video recordings to achieve temporal alignment. ENF signals may be extracted from the soundtracks of the video recordings, as well from the image sequences if the video captures the subtle flickering of lightings. Extracting the weak ENF signal from image sequences is a challenging task. The temporal sampling rate of visual recordings is generally too low to directly estimate the ENF signal that may appear at harmonics of 50 or 60 Hz. The ENF traces in video signals are relatively weak, and may be easily distorted by object and camera motions. Techniques need to be developed to address these challenges.

As the proposed approach does not rely on the perceptual audio and visual information of the recordings, it is fundamentally different from and complementary to conventional methods. One of the main advantages of the proposed approach is that it imposes no major constraints on the viewing angles, camera calibrations and camera motions. This property provides it a strong potential to address such difficult scenarios that are intractable by existing methods. The prerequisite for this approach to work is that the ENF traces in the audio/video recordings are strong enough for reliable estimation of the ENF signal.

2. VIDEO SYNCHRONIZATION USING ENF SIGNALS FROM SOUNDTRACKS

We start out by synchronizing videos based on extracting and aligning the ENF signals from soundtracks.

2.1 Extracting ENF from Audio Recordings

A general and easily implementable approach to estimating ENF signal from a source signal such as audio is the short-time Fourier transform (STFT), which is a popular non-parametric tool for frequency analysis of time-varying signals. It divides a signal into possibly overlapping frames of small durations. Within every frame, the signal can be regarded as wide-sense stationary, and each of the frames undergoes Fourier analysis respectively. For ENF estimation, we apply STFT to a source signal that contains ENF traces, and find the peak frequency within a certain range near the nominal value or the harmonics in each frame.

To facilitate evaluation, the ground-truth ENF signal can be obtained from power outlet measurements using a step-down transformer and a voltage divider circuit. Fig. 1 shows an example of ENF extraction from audio signal. In this example, an audio recording and a power measurement recording were made simultaneously in the US where the nominal value of ENF is 60 Hz. The ENF signal can be extracted from around any harmonics of the nominal value of ENF, as long as the ENF traces are strong enough. Here, we examine the second harmonic for the audio recording and the base frequency for the power recording. As can be seen from

Fig. 1 (c) and (d), the ENF signals estimated from the audio recording exhibit very similar variation trends to the ground-truth ENF signal from the power outlet measurements.
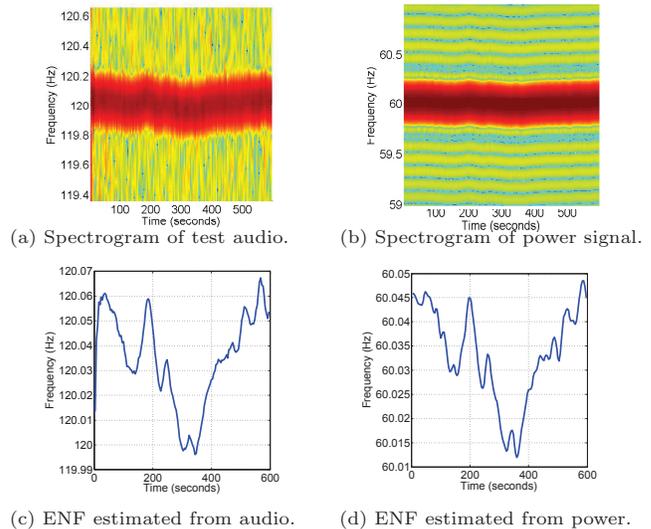

(a) Spectrogram of test audio. (b) Spectrogram of power signal.

(c) ENF estimated from audio. (d) ENF estimated from power.

Figure 1: Spectrograms and ENF estimates from audio and power signals recorded at the same time.

2.2 Synchronizing Videos with ENF From Soundtracks

Given two video clips to be synchronized, the ENF signals are first estimated from both soundtracks. We then calculate the normalized cross-correlation coefficient (NCC) of the ENF signals as a function of the lag between them. The lag corresponding to the largest value of NCC is chosen as the estimated time shift between the two video recordings.

To demonstrate experimentally the effectiveness of the proposed approach, we made two video recordings of people playing racquetball in a gym with a Canon PowerShot SX230 HS camera and a Canon PowerShot A2300 camera, respectively. The cameras shot the racquetball court from different viewing angles. Both recordings are about 10 minutes long, and one of them starts approximately 20 seconds earlier than the other. The ENF signals are estimated from the soundtracks of the video clips, and their NCC is calculated with different values of lags between them. In Fig. 2 (a), we plot the NCC as a function of the lag and observe a clear peak at 20.52 seconds. We then align the video clips by shifting them relatively by 20.52 seconds. The ENF signals after alignment, along with the reference ENF measured from the power outlet, are shown in Fig. 2 (b). Both the ENF signals extracted from the videos exhibit variation trends that are consistent with those of the reference ENF signal. A few sample pairs of images from the video sequences after alignment are shown in Fig. 2 (c). The images in the same row are from the same video stream, while the images in the same column correspond to the same time instance. By examining the movement of the girl in the images, we can see that the two video sequences are well synchronized.

Accuracy Evaluation Experiments have been conducted to evaluate the synchronization accuracy. We take multiple video recordings simultaneously with two cameras at various locations, including offices, hallways, recreation centers
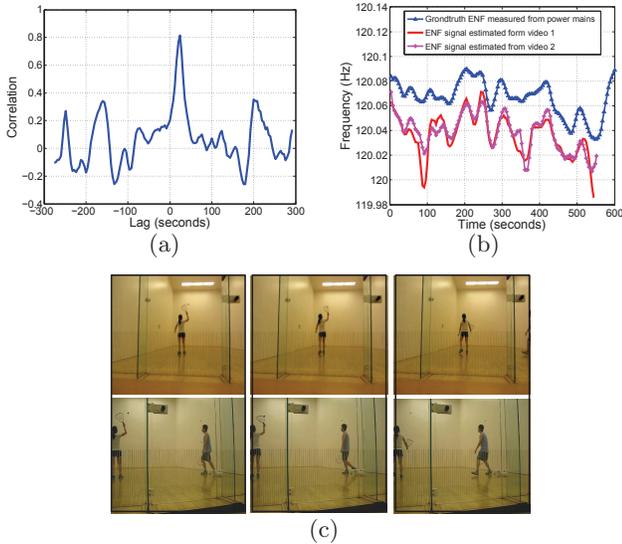
Figure 2: Example of video synchronization by aligning the ENF signals from video soundtracks.

and lobbies. These videos are divided into clips of 10 minutes long and each clip is treated as a test sample. The ground truth of the lag between the recordings are obtained by manually comparing the video frames. Using a total of nearly 7 hours' video organized in 20 pairs of test clips, we carry out synchronization using the proposed method, and the average absolute synchronization error is 0.12 second.

3. EXTRACTING ENF SIGNALS FROM VISUAL RECORDINGS

Visual recordings are also capable of capturing ENF traces. Indoor lightings often vary the light intensity in accordance with the AC voltage supplied, resulting in subtle flickering in the lights. For fluorescent lights and incandescent bulbs, the frequency of the flickering is usually twice that of the ENF, as the light intensity is proportional to the amplitude of the instantaneous input voltage, regardless of its polarity. Although the flickering may be invisible to human eyes, cameras can often capture it in video recordings. In [9], the authors take the mean of the pixel values in every image of a video sequence as source signal, and then use spectrogram analysis to estimate the embedded ENF signal. A major challenge of that scheme is the aliasing effect. By taking one sample from every frame, the ENF signal that appears at harmonics of 50 or 60 Hz is essentially sampled temporally at the frame rate of the video recordings. Current consumer digital cameras usually adopt a frame rate that is around or lower than 30 fps. The ENF signals therefore suffer from severe aliasing effect due to insufficient sampling speed. To overcome this challenge, the rolling shutter has been recently exploited as an attempt to increase the actual sampling rate [13].

3.1 Rolling Shutter of CMOS Sensors

Rolling shutters are commonly adopted for complementary metal-oxide semiconductor (CMOS) camera sensors. Unlike global shutters often employed in charge-coupled device (CCD) sensors that record the entire frame from a snap-shot of a single point in time, a camera with a rolling shutter scans the vertical or horizontal lines of each frame in a sequential manner. As a result, different lines in the same frame are exposed at slightly different times. In addition, some rolling shutter may adopt a possible idle period between finishing the scan of one frame and proceeding to the next frame. Since the pixels in different rows or columns are exposed at different times but are displayed simultaneously during playback, the rolling shutter may cause such spatial distortions as skew, smear, and other visual artifacts.

The sequential read-out mechanism of rolling shutter has been conventionally considered detrimental to image/video quality due to its accompanying artifacts. However, recent works have shown that the rolling shutter can be exploited with computer vision and computational photography techniques [14, 15]. The authors in [13] propose to take advantage of the rolling shutter to solve the problem of insufficient sampling rate for estimating the ENF signal from the image sequence of video recordings. By treating each line of the frame as a sample point, the sampling rate can be much higher than the frame rate. The work in [13] on rolling shutters is relatively preliminary as it was limited to videos of static scenes. In this paper, we carry out a further study along this direction, and develop techniques to handle videos with motions.

Without loss of generality, we assume the rolling shutter scans the frame row-by-row. Consider a video signal $s(r, c, n)$, where $1 \leq r \leq R, 1 \leq c \leq C$ and $1 \leq n \leq N$ denote the row index, column index and frame index, respectively. The video signal contains mainly two components: one is the visual component v corresponding to the visual scene; and the other is the ENF component e:

$$s(r, c, n) = v(r, c, n) + e(r, c, n). \quad (1)$$

From Eq. (1), we see that the signal-to-noise-ratio (SNR) of e in s may be low in the presence of the visual component v. For fixed spatial indices r and c, the visual component $v(r, c, n)$ as a function of n is in general a low-pass signal. In order to suppress the effect of v and extract the ENF component e, we apply an appropriate high-pass filtering to the video signal s.

3.2 Static Videos

We first consider the case where the scene in the video is static so that the visual signals of every frame in the video are identical, i.e., $v(r, c, n) = v(r, c)$. Under this assumption, Eq. (1) is reduced to

$$s(r, c, n) = v(r, c) + e(r, c, n). \quad (2)$$

We can apply a high-pass filter to s by subtracting from it its mean value across all frames:

$$\hat{s}(r, c, n) = s(r, c, n) - \bar{s}_n(r, c)$$
$$= s(r, c, n) - \frac{1}{N} \sum_{m=1}^{N} s(r, c, m)$$
$$= e(r, c, n) - \frac{1}{N} \sum_{m=1}^{N} e(r, c, m). \quad (3)$$

For any given r and c, $e(r, c, n)$ as a function of $n = 1, 2, ..., N$ is essentially the sinusoidal ENF signal sampled at the frame rate of the video recording. Since the frequency of the ENF signal is changing over time, $e(r, c, n)$ for $n = 1, 2, ..., N$

tends to have different phases and cancel out. So for a sufficiently large N, the average of these samples is close to 0, i.e.

$$\bar{e}_n(r,c) = \frac{1}{N}\sum_{m=1}^{N} e(r,c,m) \simeq 0. \quad (4)$$

This leads to

$$\hat{s}(r,c,n) \simeq e(r,c,n). \quad (5)$$

After the high-pass filtering, the SNR of the ENF signal in $\hat{s}$ is much higher than that in the original video signal s. We then use the spatial average of each row in $\hat{s}(r,c,n)$ as the source signal to estimate the ENF signal:

$$R(r,n) = \frac{1}{C}\sum_{c=1}^{C} \hat{s}(r,c,n). \quad (6)$$

$R(r,n)$ is referred to as the *row signal* hereafter.

We have conducted experiments using a Canon PowerShot SX230 HS camera that is equipped with a rolling shutter. Fig. 3 shows an example of ENF estimation from a static video recording. The test video here is a recording of a white wall under fluorescent lightings, and the camera was fixed on a tripod during the recording. Fig. 3 (a) shows a snapshot of the test video. We calculated the row signal according to Eq. (6), and then vectorized it by concatenating its entries frame after frame to form the source signal for ENF estimation. Fig. 3 (b) shows a segment of the source signal. We can see that the source signal exhibits sinusoidal waveforms except for some periodic phase shifts. These phase shifts exist because of the idle period of the rolling shutter between exposing the last row of one frame and starting the first row of the next frame. During the idle period, no recording is conducted, and a phase jump of the source signal may thus occur on every frame border (every 240 samples in this experiment). In the spectrogram of the source signal in Fig. 3 (c), we observe that due to the phase discontinuities, the ENF signal is shifted by multiples of the frame rate (29.97 Hz in this experiment). We estimate the ENF signal from around 60 Hz as we see from the spectrogram that the SNR of the ENF signal is the highest in this frequency range. The ENF signal estimated from the video signal together with its simultaneous reference ENF signal extracted from the power measurements are shown in Fig. 3 (d). The signals are properly shifted to facilitate comparison. The variation patterns in the ENF signal from the test video match well with those in the reference ENF signal.

3.3 Videos with Object Motion

It is more challenging to extract ENF signals from video recordings of scenes with moving objects. In such a scenario, Eq. (2) does not hold anymore, and the method for high-pass filtering in the previous subsection would no longer work.

If the scene in the video contains a background that is static, we can use these static regions to estimate the ENF signal. Following the notations of last section, given two image frames $s(r,c,n)$ and $s(r,c,m)$, we are interested in finding the regions that are not affected by object motion in either of the frames. The mutually motion-free regions between $s(r,c,n)$ and $s(r,c,m)$ are represented by a binary matrix $M^{n,m}(r,c)$, defined as

$$M^{n,m}(r,c) = \begin{cases} 1 & \text{if frame } n \text{ and frame } m \text{ are both static} \\ & \text{at pixel } (r,c) \\ 0 & \text{otherwise} \end{cases}$$
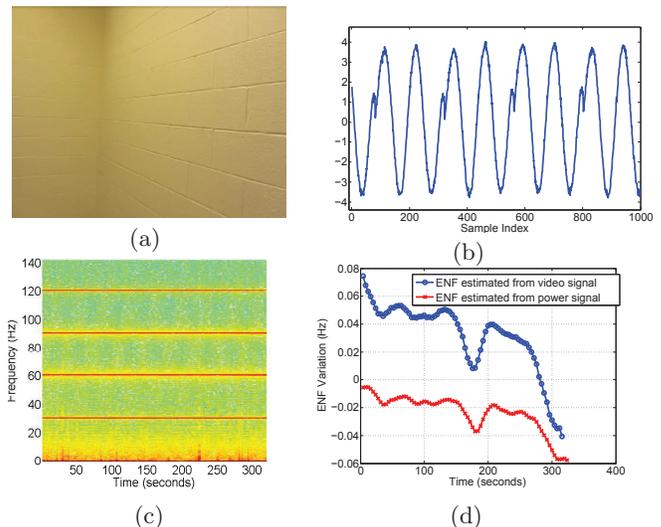

(a)   (b)

(c)   (d)

Figure 3: Example of ENF estimation from a static video recording.

A simple way to identify the motion-free regions is thresholding on the pixel-wise differences of the pixel intensity between the two images.

With a similar strategy to what was presented in Sec. 3.2, we apply a high-pass filter to the video signal by subtracting from it a smoothened version of the original signal. For an image frame s from the video sequence, we search for its mutual motion-free regions against all the other frames. The pixel values of the frames in their respective motion-free regions can be averaged to form a smoothened version of s, which is then subtracted from s:

$$\hat{s}(r,c,n) = s(r,c,n) -$$
$$\frac{1}{\sum_{m\neq n} M^{n,m}(r,c)}\sum_{m\neq n} s(r,c,m)\cdot M^{n,m}(r,c) \quad (7)$$

The row signal is obtained by taking the row average of $\hat{s}$, from which the ENF signal can be estimated. We have conducted an experiment with a video that records people walk in the hallway in an office building. The video was made with similar settings to the experiments in Sec. 3.2. We used the proposed scheme to extract the ENF signal from this test video. The reference ENF signal was also estimated from a simultaneously recorded power signal. We can see from Fig. 4 that the variation trends of the ENF signal estimated from the test video are consistent with those of the reference ENF signal.

3.4 Compensating Brightness Changes

Many cameras are equipped with an automatic brightness control mechanism that would adjust camera's sensitivity to light in response to the illumination conditions so that the overall brightness of the acquired image remains visually pleasing. As an example of such a phenomenon, two images from a video sequence are shown in Fig. 5. As the person in the second image is closer to the camera, the background wall appears brighter than in the first image. Such brightness changes introduce challenges to the estimation of the ENF signal using the techniques described in previous subsections.
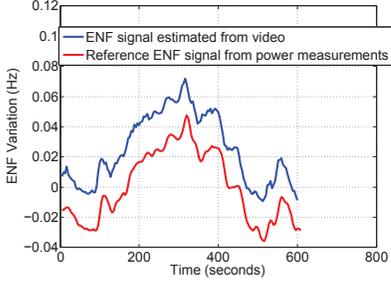
16

Figure 4: The ENF signal estimated from the test video matches well with the reference ENF signal. The signals are properly shifted to facilitate comparison.

To investigate the mitigation of the negative effect due to brightness change, we have created the following recording: during the first 4 minutes, a person walked around in a hallway relatively far from the camera so that the camera's automatic brightness adjustment was not triggered; after 4 minutes, the person moved closer to the camera, and such brightness changes occurred as shown in Fig. 5. The ENF signal is extracted from this test video using the techniques discussed in previous subsections without addressing the brightness changes. In Fig. 6, we see that the estimated ENF signal from the test video becomes distorted after 4 minutes into the recording as a result of the brightness changes in the image sequence.

Figure 5: Two image frames from a test video recording illustrating camera's automatic brightness control.

We have examined the relationship of the pixel values in different images of the same scene. For two images, we examine the regions in which both of the images are static. We find that the brightness change can be well modeled by a linear transform. Given two frames $s(r, c, n)$ and $s(r, c, m)$, we have

$$s(r, c, n) = a^{n,m} \cdot s(r, c, m) + b^{n,m}. \qquad (8)$$

For a frame $s(r, c, n)$, the pixel values in the static background regions are used to estimate the parameters $a^{n,m}$ and $b^{n,m}$. For brightness change compensation, we apply Eq. (8) to each frame $s(r, c, m)$. Eq. (3) then becomes

$$\hat{s}(r, c, n) = s(r, c, n) - \frac{1}{\sum_{m \neq n} M^{n,m}(r, c)} \cdot$$
$$\sum_{m \neq n} \left(a^{n,m} \cdot s\left(r, c, m\right) + b^{n,m} \right) \cdot M^{n,m}(r, c) \quad (9)$$

This compensation scheme was applied to the test video, and the result of ENF estimation is shown in Fig. 6. With our proposed method, the ENF signal estimated from the test video now exhibits consistent variations with the reference ENF signal.
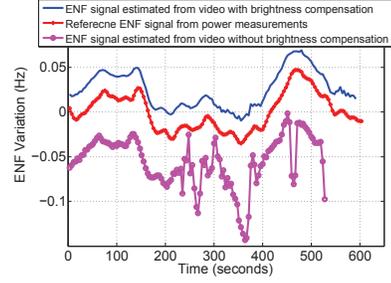

Figure 6: The effectiveness of the brightness change compensation technique. The signals are properly shifted to facilitate comparison.

3.5 Compensating Camera Motion

In the previous discussions, we have assumed that the camera is fixed during recording so that the pixels in different image frames are spatially aligned. In practice, people may hold the camera by hand to make a video recording, and camera motion compensation is needed.

For two image frames $s(r, c, n)$ and $s(r, c, m)$, we denote by $(\delta_r^{n,m}, \delta_c^{n,m})$ the pixel-wise shift between them due to the camera motion:

$$s(r, c, n) = s(r + \delta_r^{n,m}, c + \delta_c^{n,m}, m). \qquad (10)$$

To compensate for the camera motion, we need to shift the pixels in two frames relatively by $(\delta_r^{n,m}, \delta_c^{n,m})$ so that they are spatially aligned. The registered frames can be processed as described in the previous subsections. Considering the camera motion compensation, Eq. (3) becomes

$$\hat{s}(r, c, n) = s(r, c, n) - \frac{1}{N} \sum_{m=1}^{N} s(r + \delta_r^{n,m}, c + \delta_c^{n,m}, m), \quad (11)$$

and the ENF signal can then be estimated from $\hat{s}(r, c, n)$.

Optical flow methods can be used to estimate the pixel-wise displacement between image frames. These methods calculate the motion field (V_r, V_c) between two frames $s(r, c, n)$ and $s(r, c, n + \delta_n)$ based on the optical flow equation $\frac{\partial s}{\partial r} V_r + \frac{\partial s}{\partial c} V_c + \frac{\partial s}{\partial n} = 0$, and certain additional conditions and constraints for regularization. In this work, we have used the implementation of the optical flow estimation developed by [16].

An experiment was conducted to verify the proposed camera motion compensation scheme. We used the Canon PowerShot SX230 HS camera to make a video recording of a hallway. The camera was held by hand during the recording, and we deliberately shook the camera to create noticeable motion in the video recording. The ENF signal estimated from the test video with camera motion compensation matches with the groundtruth ENF signal, as shown in Fig. 7.

4. VIDEO SYNCHRONIZATION USING ENF SIGNALS FROM IMAGE SEQUENCES

In Sec. 2, we have demonstrated video synchronization by aligning the ENF signals extracted from the soundtracks of video clips. In certain scenarios such as some surveillance recordings, video recordings may have been muted or the soundtrack may have been edited, and thus have no reliable audio available. As an alternative, we may extract the ENF signal from the image sequence of the visual track using the
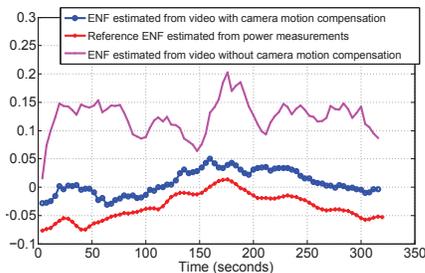
17

Figure 7: The ENF signals estimated from a test video with camera motion. The signals are properly shifted to facilitate comparison.

techniques described in Sec. 3. In this section, we present experimental results of this approach.

We used two Canon PowerShot SX230 HS cameras that are equipped with CMOS sensor and rolling shutter to video tape a hallway illuminated by an indoor light in an office building. The cameras were placed to capture the hallway from different view angles. A person walked through the hallway back and forth, and his movements were captured by both cameras.

We apply the methods discussed in Sec. 3 to estimate the ENF signals from the image sequences of both the video recordings. The NCC of the estimated ENF signals as a function of the lags between them is plotted in Fig. 8 (a), from which we find a peak NCC value of 0.96 at 60.72 seconds. The ENF signals after alignment are shown in Fig. 8 (b), and we see that the variation patterns of the ENF signals match well with each other. In Fig. 8 (c) we show several image frames from the synchronized video recordings. For comparison, we manually aligned the two videos by comparing the image frames and the soundtracks in both video clips, and found the lag to be 60.80 seconds, which is very close to the value obtained by the proposed approach.
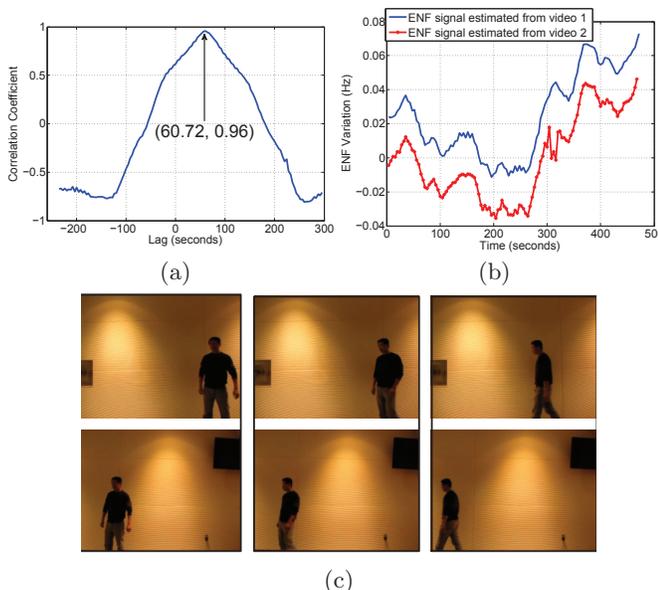

(a)　　　　　　(b)

(c)

Figure 8: Example of video synchronization by aligning the ENF signals from image sequences.

5. CONCLUSIONS

We have exploited the ENF signal inherently embedded in the soundtrack and image sequence of videos as a timing fingerprint to temporally align multiple video recordings. A critical step of the proposed approach is the estimation of the ENF signal from audio-visual data. Extraction of the ENF signal from an image sequence is particularly challenging, and to the best of our knowledge, few research attempts have been made to adequately address it. We have proposed several techniques to effectively overcome the difficulties that one may face when extracting the ENF signal from image sequences. Through our experiments, we have demonstrated that video recordings can be accurately synchronized by aligning the inherently embedded ENF signals.

6. REFERENCES

[1] A. Agarwala, K. C. Zheng, C. Pal, M. Agrawala, M. Cohen, B. Curless, D. Salesin, and R. Szeliski, "Panoramic video textures," in *SIGGRAPH*, 2005.

[2] E. Shechtman, Y. Caspi, and M. Irani, "Space-time super-resolution," *IEEE Trans. PAMI*, Apr. 2005.

[3] Y. Caspi and M. Irani, "A step towards sequence -to-sequence alignment," in *IEEE CVPR*, 2000.

[4] T. Tuytelaars and L. V. Gool, "Synchronizing video sequences," in *IEEE CVPR*, 2004.

[5] Y. Caspi, D. Simakov, and M. Irani, "Feature-based sequence-to-sequence matching," *Int'l J. Comput. Vision*, vol. 68(1), Jun. 2006.

[6] P. Shrestha, H. Weda, M. Barbieri, and D. Sekulovski, "Synchronization of multiple video recordings based on still camera flashes," in *ACM Multimedia*, 2006.

[7] F. Padua, R. Carceroni, G. Santos, and K. Kutulakos, "Linear sequence-to-sequence alignment," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32(2), Feb. 2010.

[8] C. Grigoras, "Applications of ENF criterion in forensics: Audio, video, computer and telecommunication analysis," *Foresnsic Science International*, April 2007.

[9] R. Garg, A. Varna, and M. Wu, "Seeing ENF: Natural time stamp for digital video via optical sensing and signal processing," in *ACM Multimedia*, 2011.

[10] C. Grigoras, "Applications of ENF analysis in forensic authentication of digital and video recordings," *J. Audio Eng. Soc.*, 2009.

[11] R. W. Sanders, "Digital authenticity using the electric network frequency," in *33rd AES Int'l Conf. Audio Forensics, Theory & Practice*, Jun. 2008.

[12] D. Rodriguez, J. Apolinario, and L. Biscainho, "Audio authenticity: Detecting ENF discontinuity with high precision phase analysis," *IEEE TIFS*, vol. 5(3), pp. 534–543, Sep. 2010.

[13] 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 TIFS*, vol. 8(9), 2013.

[14] O. Ait-Aider, A. Bartoli, and N. Andreff, "Kinematics from lines in a single rolling shutter image," in *CVPR*, 2007.

[15] J. Gu, Y. Hitomi, T. Mitsunaga, and S. Nayar, "Coded rolling shutter photography: Flexible space-time sampling," in *IEEE ICCP*, 2010.

[16] C. Liu, "Beyond pixels: Exploring new representations and applications for motion analysis," *Doctoral Thesis, Massachusetts Institute of Technology*, May 2009.