International Journal of Digital Crime and Forensics
Volume 12 • Issue 1 • January-March 2020

ENF Based Video Forgery Detection Algorithm

Yufei Wang, South China University of Technology, Guangzhou, China

Yongjian Hu, South China University of Technology, Guangzhou, China & Sino-Singapore International Joint Research Institute, Guangzhou, China

Alan Wee-Chung Liew, Griffith University, Gold Coast Campus, Australia

Chang-Tsun Li, Deakin University, Geelong Campus, Australia

ABSTRACT

The electric network frequency (ENF) is recorded in the videos taken under the lights powered by grid and can be used for digital forensics. However, due to the lack of data caused by the low frame rate of the video, the ENF-based forensics methods always need a reference signal extracted from the grid, which limits the practical application of these methods. In this article, a new ENF-based time domain video forgery detection algorithm is proposed to solve the problem of data lack. The cubic spline interpolation is used to generate suitable data points of the ENF signal, and the detection sequence generated based on the correlation coefficient between data points in adjacent periods is used to catch the phase continuity interruption of the ENF signal and detect the exact position of forgery. The proposed algorithm can be used independently without any reference signals. The experimental results show that the proposed algorithm has good performance in detecting forgery videos with varying degrees of deletion, duplication and insertion of frames.

KEYWORDS

Electric Network Frequency, Forensics, Time Domain, Videos

1. INTRODUCTION

The electric network frequency (ENF) is the frequency of the power distribution networks. The nominal value of the frequency is 50 Hz or 60 Hz, and it usually fluctuates around the nominal value because of the changing load in the grid. A research pointed out that the range of the fluctuation is ±0.6 Hz (Grigoras, 2005). The ENF signal would affect all devices connecting to the grid, and the effect has a high degree of uniformity within the same grid (Sanders, 2008).

The ENF signal has been used in audio forensics in the latest years. The sound recording equipment powered by the grid will record the ENF signal in the audio file, and this signal can be used as evidence for audio forensics. One approach of ENF based audio forensic algorithms is to compare the ENF signal extracted from the audio with the signal extracted from the grid to identify the generation time and location of the sound recording and to detect any tampering in the audio (Brixen, 2008; Cooper, 2008; Hajj-Ahmad et al., 2005; Hajj-Ahmad et al., 2013; Huijbregtse et al., 2009; Kajstura et al., 2005). In order to use these algorithms, the ENF signals reference databases need to be built (Elmesalawy et al., 2014; Liu et al., 2012). There are also some algorithms using the continuity of the ENF signal to detect forgery in the audio (Nicolalde et al., 2009; Rodríguez et al., 2010). These algorithms are independent of the reference signal and more flexible to use.

DOI: 10.4018/IJDCF.2020010107

The ENF signal has also been used in video forensics. In earlier work, the ENF signal for video forensics was extracted from the audio recorded during video shooting (Cooper, 2011; Grigoras, 2007; Grigoras, 2009), while these methods cannot be used when the video does not contain audio track. Later, some video forensic algorithms based on the ENF signal extracted from video had been proposed (Garg et al., 2011; Garg et al., 2013; Su et al., 2014a). These algorithms need to match the ENF signal from a video to a ENF reference database, and recent research effort has focused on finding new methods for rapid and accurate matching (Su et al., 2014b; Hajj-Ahmad et al., 2016). However, the reference ENF signal extracted from the power grid directly is hard to be satisfied in most of the time. As a result, the mentioned methods have serious limitation in practice.

It is natural to consider using the continuity of the ENF signal to detect video forgery. However, it is very difficult to use the continuity of ENF signal extracted from video since the sampling rate of data points is not high enough. The frame rate of the video is always not more than 30fps, which leads to the lack of the data points. In order to solve this problem, we analyze the ENF signal from video and employ a special method for ENF signal interpolation. As a result, we can use the reconstructed ENF signal to detect forgery without relying on a reference ENF database.

The proposed method focuses on detecting the inter-frame video forgery, which tampers the whole frame instead of the region in the frame. Frame deletion, duplication and insertion are three of the most common used inter-frame video forgery methods, so this paper mainly investigates the detection algorithm to these three forgery methods. The main contribution of this paper is an ENF based inter-frame video forgery detection method which does not need the reference ENF signal and can be used to detect the accurate forgery position in the surveillance video with static scene. The proposed method is much more practical than other existing ENF based video forgery detection methods in practice.

The rest of this paper is organized as follows. Section 2 describes the principle and implementation of our algorithm. Section 3 discusses the practical problems in the algorithm. Section 4 analyses the experimental results. Section 5 concludes the paper.

2. THE PROPOSED ALGORITHM

2.1 Reconstruction of ENF Signal

The main source of the ENF signal in video is the flicker of the lighting. In one period of the ENF signal, the voltage amplitude will reach its maximum value twice, which makes the frequency of the flicker twice the power grid frequency. As mentioned above, the nominal value of the ENF frequency is commonly 50 or 60 Hz, so the flicker frequency will be 100 Hz or 120 Hz. The flicker cannot be noticed by human because of its frequency, but it could be recorded in the videos taken under lighting.

Each frame in the video is a sample of the flicker signal. The sampling rate (the same as the video frame rate) is usually no more than 30Hz, which is lower than the flicker frequency. Fortunately, both the ENF signal and the flicker signal are narrowband. According to the sampling theorem, the sampling rate fs can be used to capture all the information from a narrowband signal if it satisfies the condition below:

$$f_s \geq 2B\left(1+\frac{\frac{f_H}{B}-\left\lfloor\frac{f_H}{B}\right\rfloor}{\left\lfloor\frac{f_H}{B}\right\rfloor}\right) \tag{1}$$

where B denotes the bandwidth of the narrowband signal, f_H denotes the high frequency boundary of the signal, $\lfloor f_H/B \rfloor$ denotes the integer part of f_H/B, which is a positive integer, and f_H/B-$\lfloor f_H/B \rfloor$ denotes the decimal part of f_H/B, and its range is [0, 1). The bandwidth of the flicker signal is usually very narrow, while the sampling rate of the video is usually not lower than 24.98 Hz, which makes the accurate extraction of flicker signal from the video possible. A sequence $y(n)$, where $n \in [0, N\text{-}1]$, and N denotes the total number of frames in the video, can be obtained by calculating the average luminance of each frame. A suitable bandpass filter can then be used to extract the flicker signal from $y(n)$.

One of the key procedures to design a suitable band pass filter is to determine the center frequency. When the flicker signal with frequency f_{Light} is sampled by the video with frame rate f_s, it will have periodic tiling in the frequency domain. In this situation, the center frequency f_0 is given by:

$$\begin{cases} f_0 = f_{\text{Light}} + kf_s \ ,k = 0,\pm1,\pm2,\cdots \\ \qquad\quad f_0 < 0.5f_s \end{cases} \tag{2}$$

In addition, the second harmonic of the flicker signal also needs to be considered, while the higher harmonics can be ignored due to their low energy. The center frequency of the second harmonic, denoted by $f_0^{'}$, is given by:

$$\begin{cases} f_0^{'} = 2f_{\text{Light}} + kf_s \ ,k = 0,\pm1,\pm2,\cdots \\ \qquad\quad f_0^{'} < 0.5f_s \end{cases} \tag{3}$$

The frequencies of the fundamental tone and the second harmonic in videos with different frame rates affected by the power grid are shown in Table 1.

The frame rate of the video can be extracted from the video header, and the ENF can be confirmed by checking the video shooting location. Moreover, as the ENF only has two possible values, it is fairly easy to design the bandpass filter using a range of frequencies and then select the one with the best performance. The bandwidth of the band pass filter is also an important factor to consider, and we will discuss it in detail in Section 3. Another key point of filter design is that the filter should be zero-phase, because the phase of the signal extracted from video is very important for forensics.

After band pass filtering, the next problem needs to be solved is the lack of data points in the extracted signal. One of the effective solutions is interpolation. There are different interpolation

Table 1. The fundamental tone frequency and second harmonic frequency of the ENF signal in videos with different frame rate and power grid frequency

Power grid frequency (Hz)	Video frame rate (fps)	Fundamental tone frequency (Hz)	Second harmonic frequency (Hz)
50	24.98	0.08	0.16
50	25	0	0
50	29.97	10.09	9.79
50	30	10	10
60	24.98	4.9	9.8
60	25	5	10
60	29.97	0.12	0.24
60	30	0	0

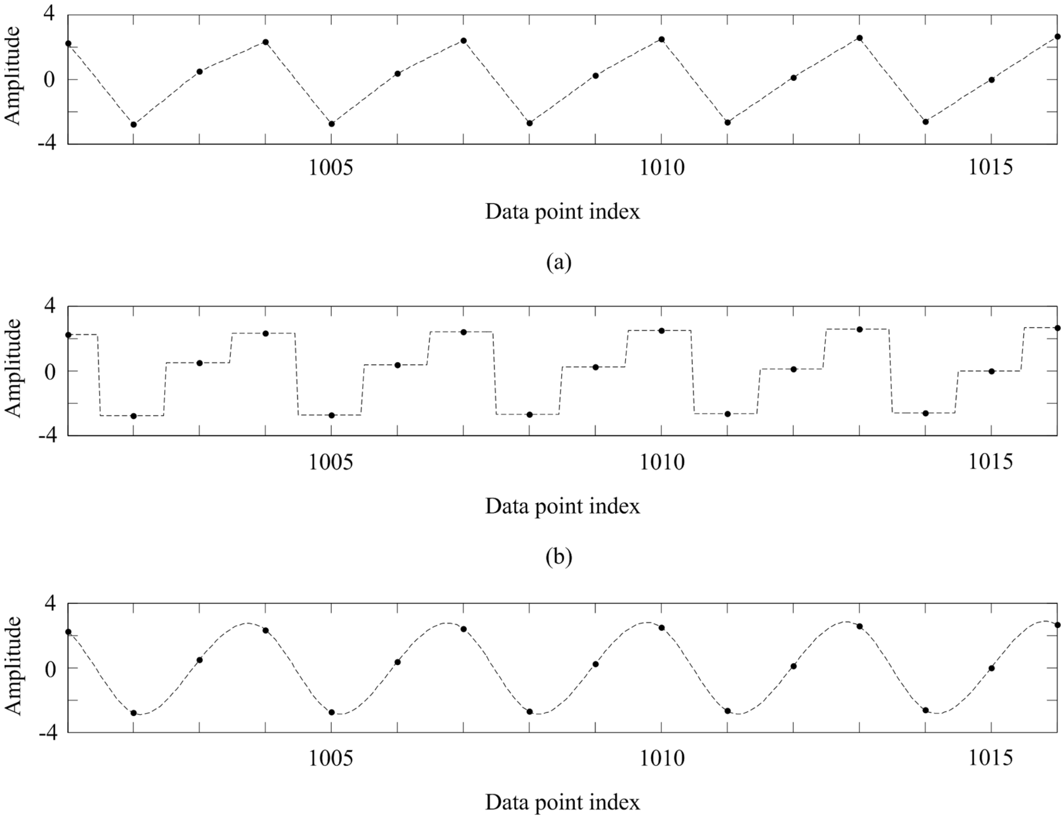
methods, and we experimentally evaluate their effects. We use a Cannon A620 digital camera to shoot a 30 minutes video indoor with fluorescent light. The camera is fixed in location to simulate a surveillance camera and generates a video with static scene. The frame rate of the video is 30fps, the total frames number is 54000. By calculating the average luminance of each frame, we can obtain the average luminance vector $y(n)$, where $n\in[0, 53999]$. The ENF in the shooting location has a frequency of 50Hz, and from Table 1 we can see that the center frequency of the band pass filter should be 10 Hz. In this experiment we select a 0.6 Hz pass band to filter $y(n)$ and then use linear interpolation, nearest-neighbor interpolation and cubic spline interpolation to generate new data. The results are shown in Figure 1. In the figure, the solid dots are the original data points, and the dotted lines are the generated curves.

The ENF signal should be close to a sine wave. Comparing the results in Figure 1, we can find that the curve generated by cubic spline interpolation is smooth and is most similar to a sinusoidal wave. Therefore, using cubic spline interpolation will provide more accurate results and help the subsequent video forensics task. For this reason, we choose cubic spline interpolation in the remaining experiments.

Cubic spline interpolation uses the piecewise third-degree polynomial functions to fit the original curve. For the data points $\{(x_i, y_i) \mid i=0, 1, 2, \dots\}$, each curve between adjacent two points is:

$$S_{i,i+1}\left(x\right) = a_{i,i+1}x^3 + b_{i,i+1}x^2 + c_{i,i+1}x + d_{i,i+1} \tag{4}$$

Figure 1. The curves reconstructed by (a) linear interpolation; (b) nearest- neighbor interpolation; and (c) cubic spline interpolation



(a)

(b)

(c)

where $S_{i,i+1}(x)$ denotes the cubic polynomial function between the data points (x_i, y_i) and (x_{i+1}, y_{i+1}), and $a_{i,i+1}$, $b_{i,i+1}$, $c_{i,i+1}$, $d_{i,i+1}$ are the coefficients of the function that need to be determined.

Using $\{(x_i, y_i) \mid i=0, 1, 2, \ldots, n\}$ to denote the set of data points, the $n+1$ data points will generate n curves. In (4) there are $4n$ unknown coefficients. From the continuity and smoothness conditions of the entire curve, we can obtain the following simultaneous equations:

$$\begin{cases} S_{i,i+1}\left(x_i\right) = y_i & ,i = 0,1,\cdots,n-1 \\ S_{n-1,n}\left(x_n\right) = y_n & \\ S_{i-1,i}\left(x_i\right) = S_{i,i+1}\left(x_i\right) & ,i = 1,2,\cdots,n-1 \\ S^{'}_{i-1,i}\left(x_i\right) = S^{'}_{i,i+1}\left(x_i\right) & ,i = 1,2,\cdots,n-1 \\ S^{''}_{i-1,i}\left(x_i\right) = S^{''}_{i,i+1}\left(x_i\right) & ,i = 1,2,\cdots,n-1 \end{cases} \tag{5}$$

where $S^{'}_{i,i+1}$ and $S^{''}_{i,i+1}$ are the first and second derivative of $S_{i,i+1}$ respectively. The simultaneous equations in (5) have $4n$-2 equations in total, and two more equations are needed to calculate the $4n$ coefficients. We use the not-a-knot end conditions to get the remaining two equations since the third derivative of the function should also be continuous (a sine wave has third derivative continuity), and we have:

$$\begin{cases} S^{'''}_{0,1}\left(x_1\right) = S^{'''}_{1,2}\left(x_1\right) \\ S^{'''}_{n-2,n-1}\left(x_{n-1}\right) = S^{'''}_{n-1,n}\left(x_{n-1}\right) \end{cases} \tag{6}$$

Using (5) and (6), the $4n$ coefficients can be determined.

2.2 Influence of Temporal Video Forgery to ENF

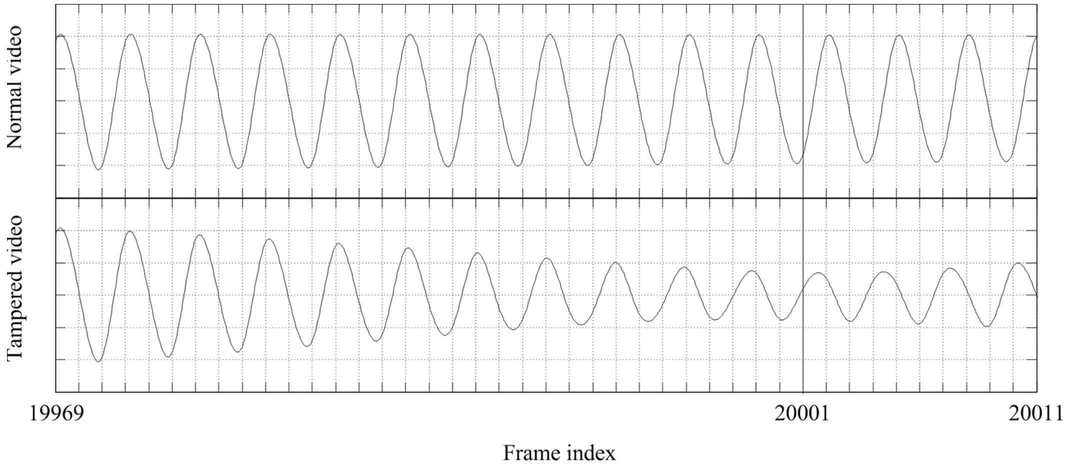
Videos with large part of static content such as surveillance videos have very high similarity between adjacent frames. When this kind of video is altered with inter-frame forgery methods it is very difficult for the naked eyes to discover the abnormity. Moreover, because of the static content, the forgery detection algorithms based on video content cannot have good performance either. However, when the video is tampered, the continuity of the ENF signal will be broken, which indicates the forgery position.

We delete 10 frames (from 20001st frame to 20010th frame) from the video mentioned in Section 2.1 to generate a tampered video, and use the method mentioned in Section 2.1 to reconstruct the ENF signals. The ENF signals in the original video and tampered video are shown in Figure 2, and the forgery position is between the 20000th and 20001st frame.

From Figure 2 we can see that in several periods near the forgery position, the amplitude of the curves is significantly different between the normal and tampered videos, while the phase of the two signals are still closed. However, the difference of phase between the two signals increases suddenly in the period including the forgery position. The phase in the normal signal should be continuous, and the figure indicates that temporal forgery will interrupt the continuity of the phase. Therefore, using the continuity of phase in the extracted ENF signal to detect temporal forgery is a reasonable and effective method.

In order to detect the change in continuity of the phase and prevent the false alarm caused by the amplitude of the signal at the same time, correlation coefficient is used as a measure. Assume that two adjacent periods in the reconstructed ENF signal are $S_1(n)$ and $S_2(n)$ respectively, each period contains N_0 data points, the correlation $R(S_1,S_2)$ between the two periods can be calculated as follows:

Figure 2. Comparison of ENF signals at frame deletion position



$$R\left(S_1,S_2\right)=\sum_{n=1}^{N_0}S_1\left(n\right)S_2\left(n\right) \tag{7}$$

As the sine wave is a good approximation of the ENF signal, we can assume that the amplitude of $S_1(n)$ and $S_2(n)$ are A_1 and A_2, the angular frequencies are ω_1 and ω_2, and the phases are φ_1 and φ_2 respectively. In this case, (7) can be rewritten as follows:

$$\begin{aligned}R\left(S_1,S_2\right)=\sum_{n=1}^{N_0}A_1\sin\left(\omega_1 n+\varphi_1\right)A_2\sin\left(\omega_2 n+\varphi_2\right)=-\frac{A_1A_2}{2}\sum_{n=1}^{N_0}\cos\left[\left(\omega_1+\omega_2\right)n+\varphi_1+\varphi_2\right]\\+\frac{A_1A_2}{2}\sum_{n=1}^{N_0}\cos\left[\left(\omega_1-\omega_2\right)n+\varphi_1-\varphi_2\right]\end{aligned} \tag{8}$$

Because $S_1(n)$ and $S_2(n)$ are contiguous, their angular frequencies are approximately equal. Substitute $\omega_1 \approx \omega_2$ into (8) we have:

$$R\left(S_1,S_2\right)=\frac{A_1A_2}{2}\sum_{n=1}^{N_0}cos\left(\varphi_1-\varphi_2\right) \tag{9}$$

Equation (9) shows that the correlations between adjacent periods in ENF signal are related to the difference of phase between them. The smaller the phase difference is, the higher the correlation will be, and vice versa. Therefore, the correlation between adjacent periods in ENF signal can be used to detect the interruption of the phase continuity.

In (9), the amplitude A_1 and A_2 still affect the value of correlation, while Figure 2 shows that accurate detection needs to avoid the influence of amplitude. For this reason, we use Pearson correlation coefficient to measure the correlation. The Pearson correlation coefficient is calculated as follows:

$$R\left(S_1,S_2\right)=\frac{\sum_{n=1}^{N_0}\left[S_1\left(n\right)-E_1\right]\left[S_2\left(n\right)-E_2\right]}{\sqrt{\sum_{n=1}^{N_0}\left[S_1\left(n\right)-E_1\right]^2}\sqrt{\sum_{n=1}^{N_0}\left[S_2\left(n\right)-E_2\right]^2}} \tag{10}$$

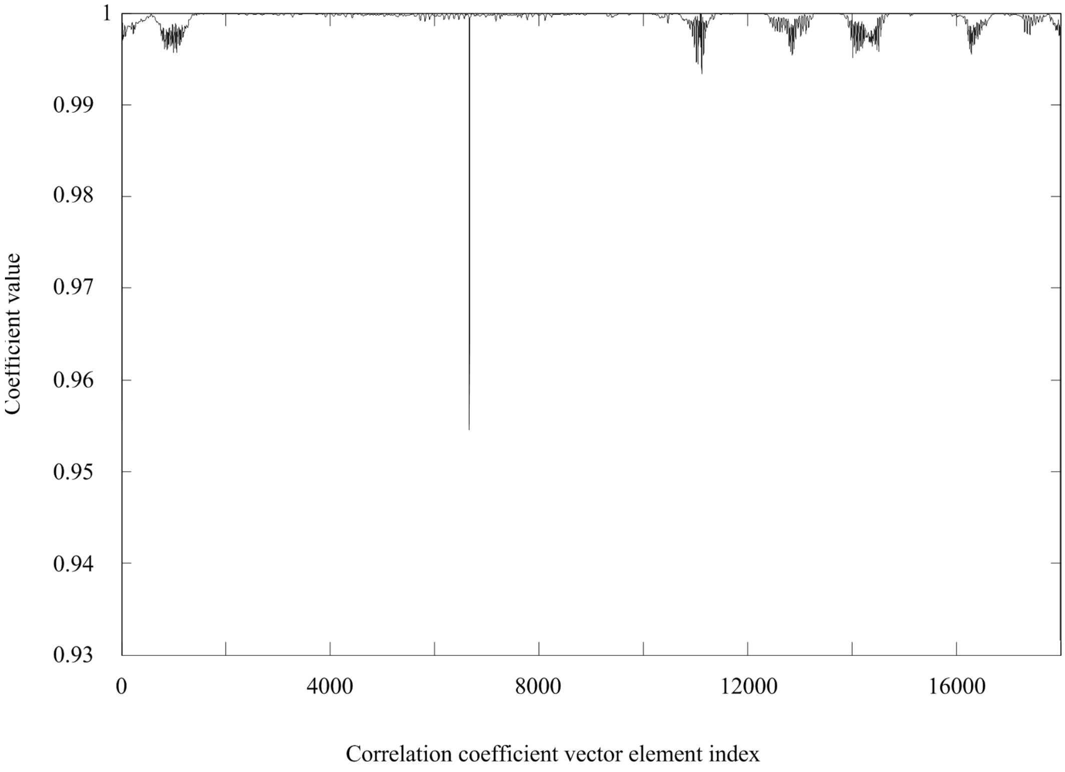
where E_1 and E_2 denote the mean of the data points in $S_1(n)$ and $S_2(n)$, respectively, that is:

$$\begin{cases} E_1=\dfrac{\sum_{n=1}^{N_0}S_1\left(n\right)}{N_0} \\ E_2=\dfrac{\sum_{n=1}^{N_0}S_2\left(n\right)}{N_0} \end{cases} \tag{11}$$

By using the Pearson correlation coefficient, the phase continuity interruption can be detected effectively without being affected by the amplitude of the signal. Using the tampered video sample mentioned above, we interpolate 4 data points between each two original data points. That means each period in the ENF signal has 15 data points. We calculate the Pearson correlation coefficient between each pair of adjacent periods, the result is shown in Figure 3.

From Figure 3 we can see that most of the correlation coefficients are above 0.99, while at the forgery position the coefficient decreases to below 0.96 suddenly, which is significantly different

Figure 3. Correlation coefficients between adjacent periods of data in the ENF signal extracted from a forgery video



from nearby coefficients. This example shows that Pearson correlation coefficient can be used to detect forgery.

In Figure 3 the decrease of the correlation coefficients at the beginning and the end of the signal are also significant. We investigate the ENF signal at the beginning and the end of the video. The reconstructed signal is shown in Figure 4.

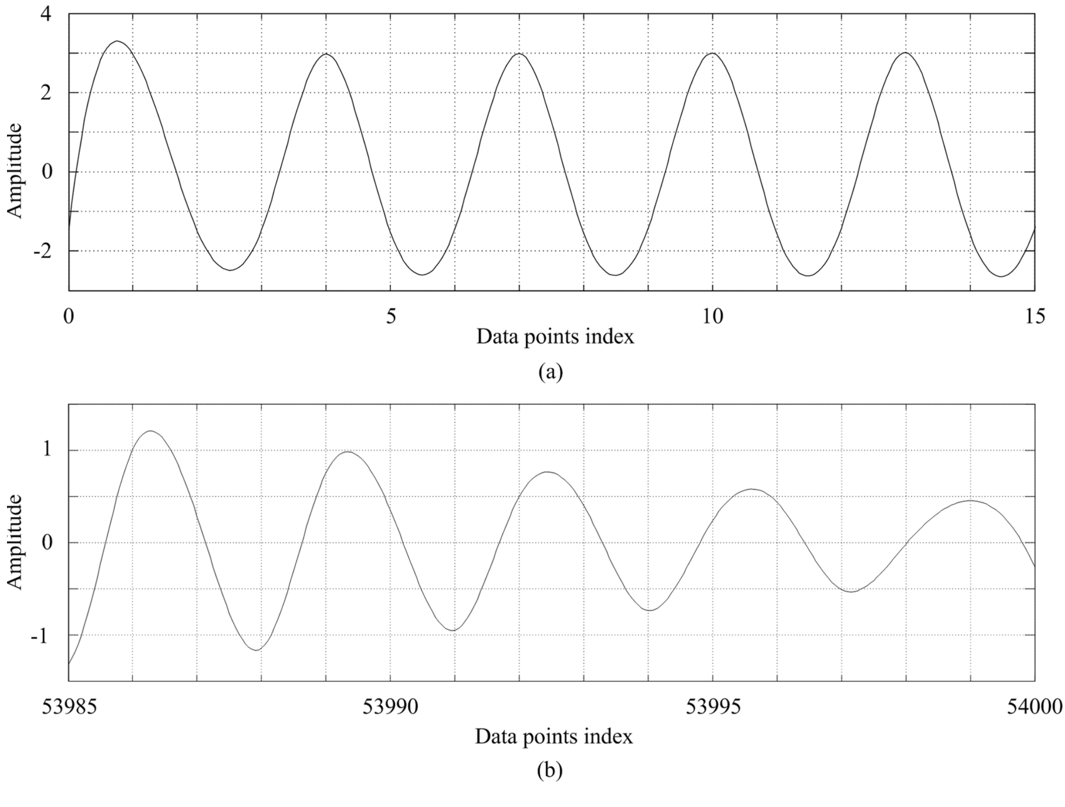
From Figure 4 we can see that the phase of the ENF signal also changes at the beginning and the end of the video. The reason for this phenomenon is that the filtering at the boundary of the finite signal is always inaccurate due to windowing effect. In actual computation, the finite signal needs to be padded at the beginning and end when being filtered, and the padding data points are always different from the actual ones, which are not available in practice. In this situation, the ENF signal extracted by the filtering method will inevitably also have several incorrect periods at the beginning and the end. Fortunately, the surveillance video always has long duration, and the suspicious forgery position is limited. We can divide the surveillance video into subsequences with 50% overlap to cover the suspicious forgery position, which ensures the boundary errors in one subsequence will be handled by its adjacent subsequences. For this reason, we do not need to deal with the boundary errors in practice.

On the other hand, some fluctuations of the coefficient value can also be observed in Figure 3, though the coefficients are still above 0.99. The fluctuations are caused by frequency fluctuation of the ENF signal, and we will analyze it in detail in Section 3.

2.3 The Detection Algorithm

The procedure of the temporal forgery detection algorithm for indoor surveillance video is as follows:

Figure 4. The reconstructed curves of the ENF signal at the (a) beginning; and (b) end of the video



(a)

(b)

First, divide the long duration surveillance video into subsequences with 50% overlap, and divide each frame into non-overlapping, same size blocks $b_{n,i,j}$, which denotes the block located at position (i, j) of frame n. To a video with N frames, $n \in [0,N-1]$. Calculate the average luminance of the block $y^b_{n,i,j}$. For all N blocks in the same position, find the minimal value $y^{i,j}_{\min}$ and maximum value $y^{i,j}_{\max}$, i.e.

$$\begin{cases} y^{i,j}_{max} = \max\left\{y^{b}_{n,i,j} \mid n \in \left[0, N-1\right]\right\} \\ y^{i,j}_{min} = \min\left\{y^{b}_{n,i,j} \mid n \in \left[0, N-1\right]\right\} \end{cases} \tag{12}$$

In digital video, the luminance of a pixel is between 0 and 255. When the original luminance is close to 0 or 255, the fluctuation of luminance may be limited. In order to ensure the accuracy, these blocks should not be used in forgery detection. On the other hand, the movements of objects will also cause luminance change and reduce the reliability of the detection algorithm, so we will only use the static blocks to extract ENF signal. We set a high luminance threshold Th_H, a low luminance threshold Th_L, and a difference threshold Th_D. The blocks that are used in ENF signal reconstruction should meet the following conditions:

$$\begin{cases} y^{i,j}_{max} < Th_{H} \\ y^{i,j}_{min} > Th_{L} \\ y^{i,j}_{max} - y^{i,j}_{\min} < Th_{D} \end{cases} \tag{13}$$

All blocks used to extract ENF signal in one frame constitute a set B, and the number of elements in B is N_B. The average luminance of the nth frame $y(n)$ is calculated as follows:

$$y\left(n\right) = \frac{\sum_{\left(i,j\right)\in B} y^{b}_{n,i,j}}{N_{B}} \tag{14}$$

After obtaining the average luminance of all frames, a zero-phase FIR band pass filter is designed according to the frame rate and power grid frequency as mentioned in Section 2.1, and the filter will be used to extract the ENF signal.

Then, cubic spline interpolation is used to interpolate n_0 new data points between each two adjacent original data points. After interpolation, each period of signal has N_0 data points, and

$$N_{0} = \left\lfloor \frac{f_{s}}{f_{0}} \right\rfloor \left(n_{0} + 1\right) \tag{15}$$

where f_0 denotes the center frequency of the filter and f_s denotes the frame rate of the video.

Assume that the interpolated ENF signal is $S(n)$, where $n \in [0, N_{\text{total}}-1]$, and N_{total} is the total number of data points in the interpolated signal. The Pearson correlation coefficient of each two adjacent periods is computed to get a correlation coefficient sequence, that is:

$$C\left(i\right)=\frac{\sum_{n=Th_{\text{front}}+\left(i-1\right)\times N_0}^{Th_{\text{front}}+i\times N_0-1}\left[S\left(n\right)-E_{i}\right]\left[S\left(n+N_{0}\right)-E_{i+1}\right]}{\sqrt{\sum_{n=Th_{\text{front}}+\left(i-1\right)\times N_0}^{Th_{\text{front}}+i\times N_0-1}\left[S\left(n\right)-E_{i}\right]^2}\sqrt{\sum_{n=Th_{\text{front}}+\left(i-1\right)\times N_0}^{Th_{\text{front}}+i\times N_0-1}\left[S\left(n+N_{0}\right)-E_{i+1}\right]^2}}C\left(i\right) \tag{16}$$

where $C(i)$ denotes the Pearson correlation coefficient between the data points of the i^{th} and the $i+1^{\text{th}}$ period in the ENF signal and E_i denotes the mean of the data points in the i^{th} period which can be calculated as follows:

$$E_{i}=\frac{\sum_{n=Th_{\text{front}}+\left(i-1\right)\times N_0}^{Th_{\text{front}}+i\times N_0-1}S\left(n\right)}{N_{0}} \tag{17}$$

When the total number of data points in the ENF signal is not divisible by N_0, considering the boundary errors mentioned in Section 2.2, we can just discard the excess data points at the end of the signal, and the range of i should be:

$$0\le i\le\left\lfloor\frac{N_{\text{total}}}{N_{0}}\right\rfloor-1 \tag{18}$$

Meanwhile, because of the boundary errors, we set two discarding threshold Th_{front} and Th_{back}. The $C(i)$ will be used for forgery detection only when i satisfies the following condition:

$$Th_{\text{front}}<i<Th_{\text{back}} \tag{19}$$

The $C(i)$ obtained from the procedure above is between -1 and 1, and the normal value should be close to 1. For convenience, we introduce a detection sequence $C_t(i)$, and let

$$C_{t}\left(i\right)=1-C\left(i\right) \tag{20}$$

The $C_t(i)$ is in the range of [0,2]. It will be close to 0 in normal position and will increase significantly at the position of sudden phase change.

Next, we set a threshold Th_1 according to the mean of $C_t(i)$ as follows:

$$Th_{1}=\frac{a_{1}}{Th_{\text{back}}-Th_{\text{front}}-1}\sum_{i=Th_{\text{front}}+1}^{Th_{\text{back}}-1}C_{t}\left(i\right) \tag{21}$$

where a_1 is a positive number greater than 1 selected as needed. When the value of $C_t(i)$ is greater than Th_1, it can be considered as a candidate of forgery position and will be investigated further.

The fluctuations in Figure 3 must be processed in further detection. Assume that the forgery candidate is $C_t(I_{\text{check}})$, in order to prevent false alarm caused by fluctuations, we select N_{check} elements before and after the $I_{\text{check}}{}^{\text{th}}$ element, and set the other threshold Th_2 based on the mean of the selected elements as follows:

$$Th_2 = \frac{a_2}{2N_{\text{check}} + 1} \sum_{i=I_{\text{check}} - N_{\text{check}}}^{I_{\text{check}} + N_{\text{check}}} C_t\left(i\right) \tag{22}$$

where a_2 is also a positive number greater than 1. When $C_t(I_{\text{check}})$ is greater than Th_2 at the same time, it can be asserted that the abnormal value of $C_t(I_{\text{check}})$ is caused by temporal forgery.

After that, we need to determine which period the forgery position is located in. Considering that two forgery positions should not be very close in practice, we can assume that among the selected $2N_{\text{check}}+1$ elements there is only one forgery position. The forgery position period can be obtained by comparing the value of the two elements beside the peak $C_t(I_{\text{check}})$. Assume that the forgery position is in the $I_{\text{forgery}}^{\text{th}}$ period, the relation between I_{forgery} and I_{check} is:

$$I_{\text{forgery}} = \begin{cases} I_{\text{check}} & ,C_t\left(I_{\text{check}} - 1\right) \geq C_t\left(I_{\text{check}} + 1\right) \\ I_{\text{check}} + 1 & ,C_t\left(I_{\text{check}} - 1\right) < C_t\left(I_{\text{check}} + 1\right) \end{cases} \tag{23}$$

In practice, there may be more than one forgery in one video, so we discard the $2N_{\text{check}}+1$ elements which have been used to determine a forgery position, and then use the remaining ements to repeat the procedure above, until all the forgery periods have been found.

After confirming the forgery position is in period I_{forgery}, we can verify that the forgery position is in the range of $[N_0 \times I_{\text{forgery}}, N_0 \times (I_{\text{forgery}}+1)-1]$ to the data points of the interpolated ENF signal. According to the interpolation method mentioned above, considering that the precise location of the forgery should be between two adjacent frames, the forgery position n in the original video will be in the range of

$$\frac{N_0 \times I_{\text{forgery}}}{n_0 + 1} - 1 < n < \frac{N_0 \times \left(I_{\text{forgery}} + 1\right)}{n_0 + 1} \tag{24}$$

By (24) we can obtain the range of the final forgery position in original video. The precision of the detection result is based on the frame rate of the video and the power grid frequency. When the power grid frequency is 50Hz and the frame rate is 30fps, the precision of the detection result will be 3 frames, or 0.1 seconds.

3. PRACTICAL PROBLEMS AND ANALYSIS

In Section 2, we analyzed the influence of temporal forgery on the ENF signal and proposed a forgery detection algorithm. However, we also observe some phenomenon that may weaken the performance of the algorithm, such as the fluctuations of the coefficients mentioned earlier. It is necessary to analyze these phenomena in order to minimize detection failure. In this section, we will analyze some phenomena which have significant impact on the detection results and describe how they can be addressed.

3.1 Influence of ENF Frequency Shifting

Consider the video taken by Canon A620 camera mentioned above as an example. We use the proposed algorithm to detect the forgery, and the Pearson correlation coefficients of the ENF signal are shown in Figure 5.

In order to show the relation between the values of the correlation coefficient and the frequency of the ENF signal, we use the ENF signal without interpolation to calculate its instantaneous frequency.

We use a window with a length of 480 to select data points from the ENF signal which has 54000 data points in total. We calculate the Fourier transform of the 480 data points using a sliding window with a stride of 240, and obtain the approximate instantaneous frequency. The result is shown in Figure 6. In Figure 6, the lighter color represents higher value, and we can see that the instantaneous frequency of the ENF signal waves is around 10 Hz.

Compare Figure 6 with Figure 5, the correlation between coefficients value fluctuation and ENF frequency shifting is obvious. When the frequency of ENF signal is close to 10 Hz, the coefficient value is close to 1, whereas the coefficient value reduces when the frequency deviates from 10 Hz.

In practice, all signals processed in the proposed algorithm are discrete, and the picket fence effect cannot be avoided. In the example above, the center frequency of the filter is 10 Hz. It will extract the ENF signal more accurately when the instantaneous frequency of ENF signal is close to 10 Hz and is less accurate when the frequency shifts. The less accurate data points in the signal will have lower correlation with the data points in adjacent period. On the other hand, the frequency shifting also means the change in instantaneous phase, which will also reduce the value of the correlation coefficient. Because of these reasons, the correlation coefficient fluctuates in value. Fu et al. (2013) proposed a method to obtain accurate spectrum, but the method will also smooth the change caused by forgery and therefore is not suitable for our proposed forgery detection algorithm.

Compare Figure 5 with Figure 3, we can see that the change in value caused by ENF frequency shifting is much smaller than the one caused by forgery. Moreover, this kind of change in value is always continuous in a period of time, while the one caused by forgery is always abrupt. For these reasons, we can use the two step detection mentioned in Section 2.3 to handle this problem and avoid false alarm.

Figure 5. Correlation coefficients between adjacent periods of data in the ENF signal extracted from an original video

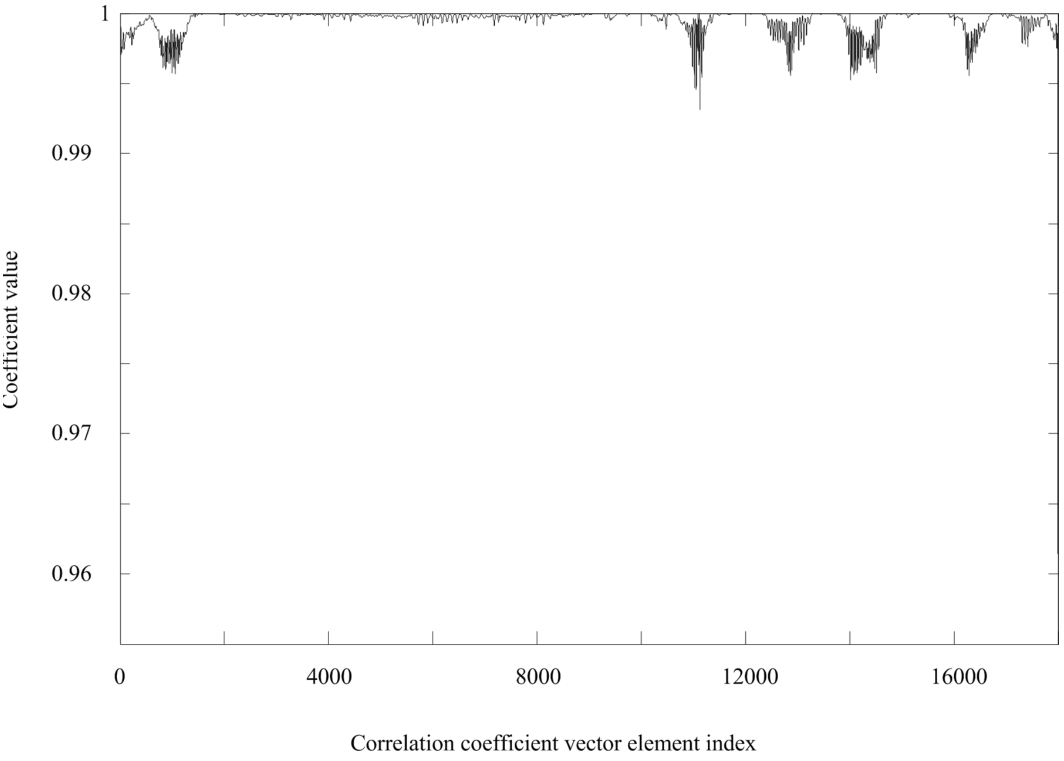
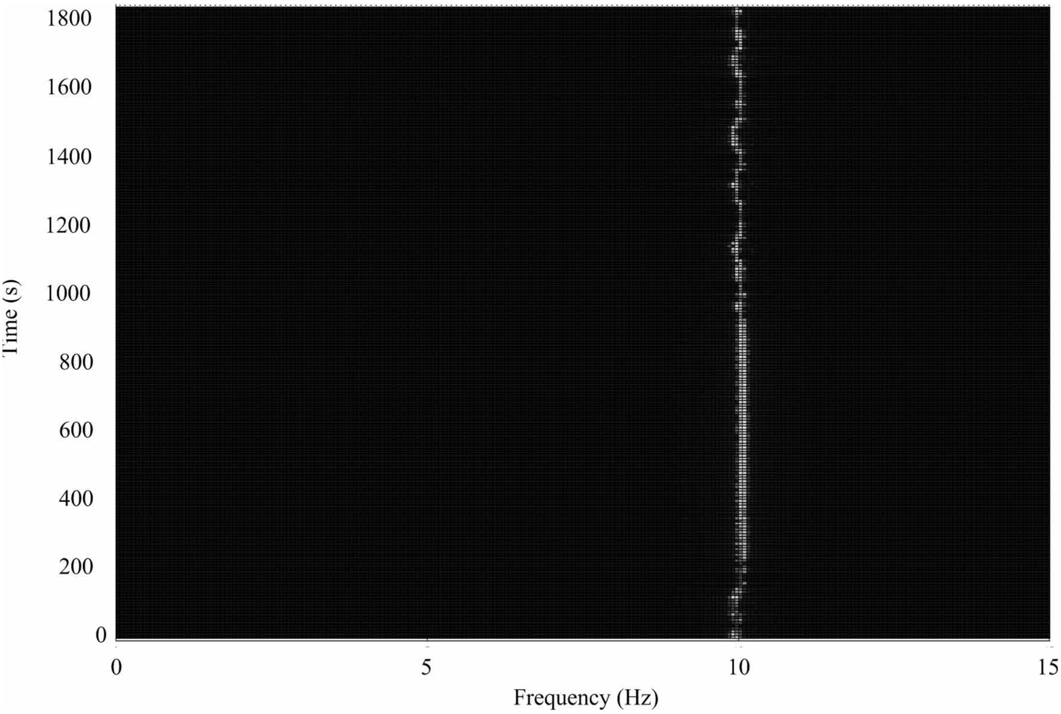


Figure 6. The spectrum of ENF signal extracted from an original video



3.2 Influence of Frame Dropping

In Figure 5 there is an abnormal decrease of coefficient value near the 11000th element, which is different from the change caused by frequency shifting. We examine the data and found that the exact position of this abnormal change occurs at the 11101th element. According to the number of data points in each period, it can be inferred that the abnormal position is between the 33301th and the 33306th data point. We check the average luminance sequence around the suspicious position, and the value of the data points are shown in Figure 7.

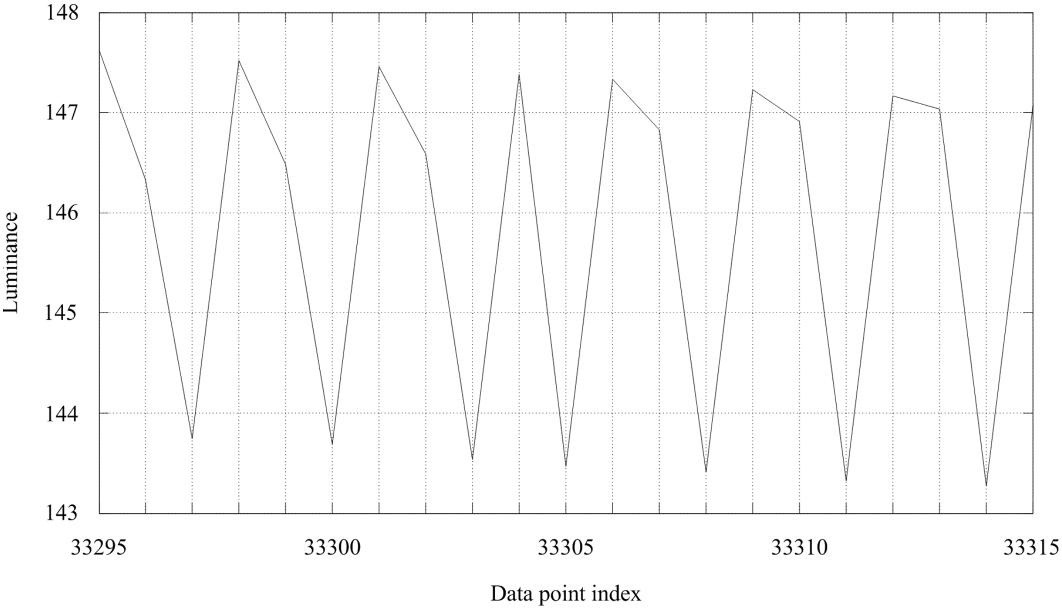
In Figure 7 it can be observed that every three data points compose one period approximately, and it is also consistent with the frame rate of the video and the frequency of the ENF signal. However, the 33305th data point does not satisfy this pattern. By analyzing the trend of the data points, we can infer that the original 33305th frame has been dropped. The frame drop undermines the periodicity of the average luminance sequence and changes the phase of the ENF signal extracted from the sequence at the corresponding position, which causes the abnormal decrease of the coefficient value in Figure 5.

Because of the limitation of the environment and hardware, the frame dropping problem cannot be completely avoided. The actual effect of frame dropping is similar to the frame deleting forgery, and both of them will interrupt the continuity of the video. Fortunately, frame dropping only involves a few frames, while the frame deleting forgery will delete a considerable number of frames. For this reason, this abnormal value change can be handled by setting suitable thresholds in the algorithm.

3.3 Influence of Filter Bandwidth

When using the band pass filter to extract the ENF signal, the bandwidth of the filter has a significant influence on the result. Using the video taken by Canon A620 camera mentioned above, changing the bandwidth of the band pass filter from 0.6 Hz to 1.2 Hz, and keeping the other parameters constant, we obtain the correlation coefficient values as shown in Figure 8.

Figure 7. The average luminance sequence around the suspicious position



In Figure 8, we can observe that the reduction of coefficient value caused by frame dropping mentioned in Section 3.2 becomes quite significant and is close to the situation caused by inter-frame forgery shown in Figure 3. We can also find that the change in value caused by frequency shifting is bigger than the one mentioned in Section 3.1. Both of the phenomena will bring deterioration to the performance of the proposed detection algorithm and may give rise to false alarm.

As analyzed above, the direct cause of reduction in correlation coefficient value is the phase difference between the adjacent periods in the ENF signal. When the difference increases, the reduction in coefficient value will become more significant. Moreover, the difference of phase always means the difference of instantaneous frequency. Assume that the ENF signal is $S(t) = A\cos(2\pi f_0 t+\varphi)$, and to account for the fact that the phase of the signal will change slowly with time, i.e. the phase of the signal is a function of time, the ENF signal can be rewritten as follows:

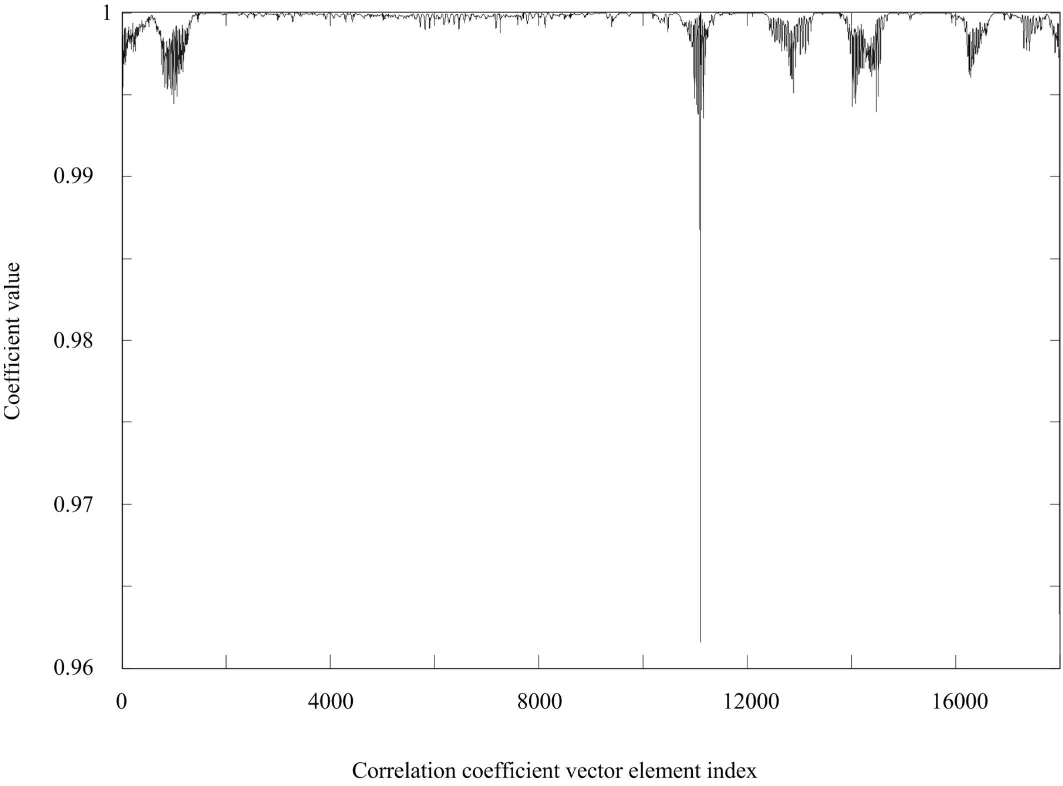
$$S\left(t\right) = A\cos\left[2\pi f_0 t + \varphi\left(t\right)\right] \tag{25}$$

Let $\Phi(t)=2\pi f_0 t+\varphi(t)$, the instantaneous frequency of the ENF signal can be calculated as follows:

$$f\left(t\right) = \frac{\Phi'\left(t\right)}{2\pi} = f_0 + \frac{\varphi'\left(t\right)}{2\pi} \tag{26}$$

In normal situation, the phase changes with time very slowly and the derivative $\varphi'(t)$ is close to 0, and the instantaneous frequency of the signal is close to the centre frequency of the filter. In this situation, the correlation coefficient value will not change significantly when the filter bandwidth is increased. When frequency shifting happens, the phase change will become faster with time and the derivative is greater than 0, which causes the instantaneous frequency to deviate from the centre frequency. When the filter bandwidth becomes wider, more energy in the signal with phase change

Figure 8. Correlation coefficients between adjacent periods of data in the ENF signal extracted from an original video by a band pass filter with 1.2 Hz pass band



will be kept, and the correlation coefficient will decrease correspondingly. Moreover, at the position of frame dropping, the decrease in value will become much more serious. The frame dropping will bring sudden change to the phase of the signal and makes the derivative *φ'*(*t*) large. When the bandwidth of the filter becomes wider, more energy from the signal with phase change will be kept, and the energy will be much stronger than the one caused by frequency shifting, which leads to a sudden reduction in peak similar to the situation of forgery. In summary, in order to avoid the false alarm and improve the performance of the proposed detection algorithm, a pass band with narrow bandwidth should be selected when designing the filter for ENF signal extraction.

4. EXPERIMENTAL RESULTS AND ANALYSIS

To evaluate the performance of the proposed algorithm, we use fixed digital cameras to shoot videos indoor under fluorescent light and obtain the videos which can simulate the surveillance videos. Surveillance video forgery detection is one of the most important applications of our proposed algorithm. The original videos are altered with frame deletion, duplication and insertion to generate the forged videos. Then the proposed algorithm is used to detect the positions of forgery. There are no other ENF based inter-frame forgery detection methods using ENF signal independently, so we compare our method with two state-of-the-art inter-frame forgery detection methods based on velocity field estimation and variation of prediction footprint (Huang et al., 2017) and multi-level subtraction (Sitara et al., 2017). We abbreviate the two methods to VFE-VPF and MLS for short. The setting of the experiments and the analysis of the results will be elaborated on in the following sub-sections.

4.1 Settings of the Experiments

The cameras used in the experiments include Canon A620, Canon A710 and Canon G12. The power grid frequency of the shooting location is 50 Hz. The frame rate of the videos is 30fps, and the resolution is 320×240. All the original videos are cut to 30 minutes, contains 54000 frames in total. During the detection of static area in the video, the block size is set to 4×4. The small block size will allow us to extract static area more accurately. According to the general condition of the videos, the thresholds in (13) are set to $Th_H = 235$, $Th_L = 20$ and $Th_D = 40$.

According to the power grid frequency, the frame rate of the video, and the analysis in Section 3.3, we design a band pass filter with centre frequency of 10 Hz and bandwidth of 0.6 Hz to extract the ENF signal from the videos. This filter can prevent false alarm caused by frequency shifting and frame dropping during detection effectively.

Cubic spline interpolation is applied to interpolate 4 data points between each two adjacent original data points. In order to avoid the boundary error shown in Figure 4, we control the forgery position and ensure the forgery position will not be located at the first or the last second of the video. In the experiments, 1 second contains 10 periods of ENF signal, and the range of i in $C_t(i)$ is [0,17997]. We set the discarding thresholds to $Th_{\text{front}} = 9$ and $Th_{\text{back}} = 17988$. Based on the investigation of numerous forgery videos, the N_{check} used for further detection is set to 10.

When calculating the thresholds $Th_1$and Th_2, the selection of a_1 and a_2 is very important and will affect the performance of the algorithm directly. In order to evaluate the overall performance of the algorithm, we fine tune a_1 and a_2 to get different false positive rates and true positive rates, and plot the ROC curve of the algorithm.

To the two algorithms for comparison, all the parameters except thresholds are set following the references. The thresholds in the two algorithms will be changed in steps in order to get different false positive rates and true positive rates and plot the ROC curve.

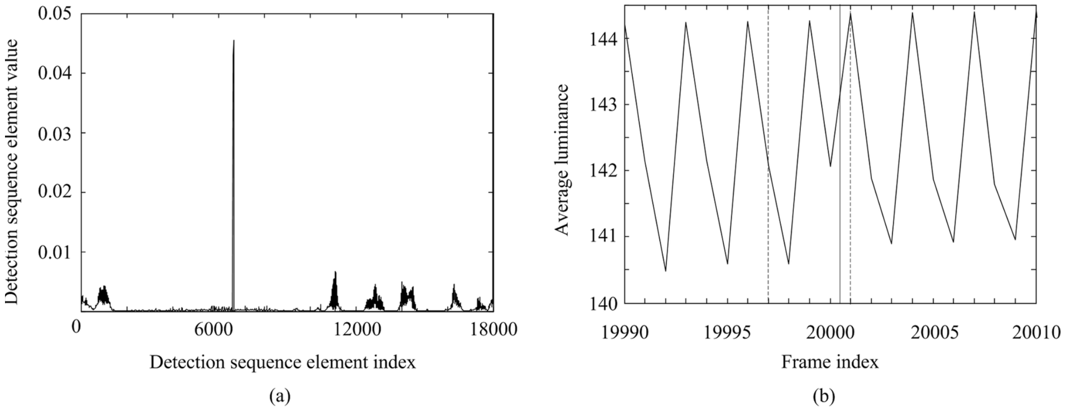
4.2 Detection of Frame Deletion Forgery

For video with large part of static scene, the frame deletion forgery can hardly be noticed by the naked eye. However, the detection algorithm based on ENF signal can solve this problem. In this experiment, the 30-minute videos will be deleted 10, 100 or 500 frames at random position between 01:00 and 29:00 of the video to generate forgery videos. Then the proposed algorithm is used to detect the forgery location in the videos.

First, we will examine a specific example. We select a forgery video taken by Canon A620 and delete the 20001st to 20010th frames in the video. The forgery position in this video is between the 20000th and the 20001st frame. Using the proposed detection algorithm, we have the detection sequence $C_t(i)$ shown in Figure 9(a). By checking the position of the peak in Figure 9(a), we obtain the detection result shown in Figure 9(b). In Figure 9(b), the solid line represents the actual position of forgery and the dotted lines mark the range of the detection result. In this example, the detection range is (19997, 20001), which includes the actual position of forgery. It indicates that the proposed algorithm is effective in this example.

Then we change the parameters a_1 and a_2 to calculate the thresholds Th_1 and Th_2 and use different thresholds to detect the videos in batch. In the experiment, the range including the actual forgery position will be considered as a positive sample, and the other part of the video will be considered as a negative sample. When the algorithm detects a positive sample successfully, the number of true positive samples increases by 1. When the algorithm detects a negative sample as forgery position wrongly, the number of false positive samples increases by 1. The number of true positive samples divided by the number of positive samples is the true positive rate, and the number of false positive samples divided by the number of negative samples is the false positive rate. For different a_1 and a_2 we will have different true positive rates and false positive rates, and the ROC curve can be obtained. The ROC curves for different length of deleted frame are shown in Figure 10 (a).

Figure 9. (a) The detection sequence used in the frame deletion forgery example; (b) the forgery position detection result of this example.



(a)

(b)

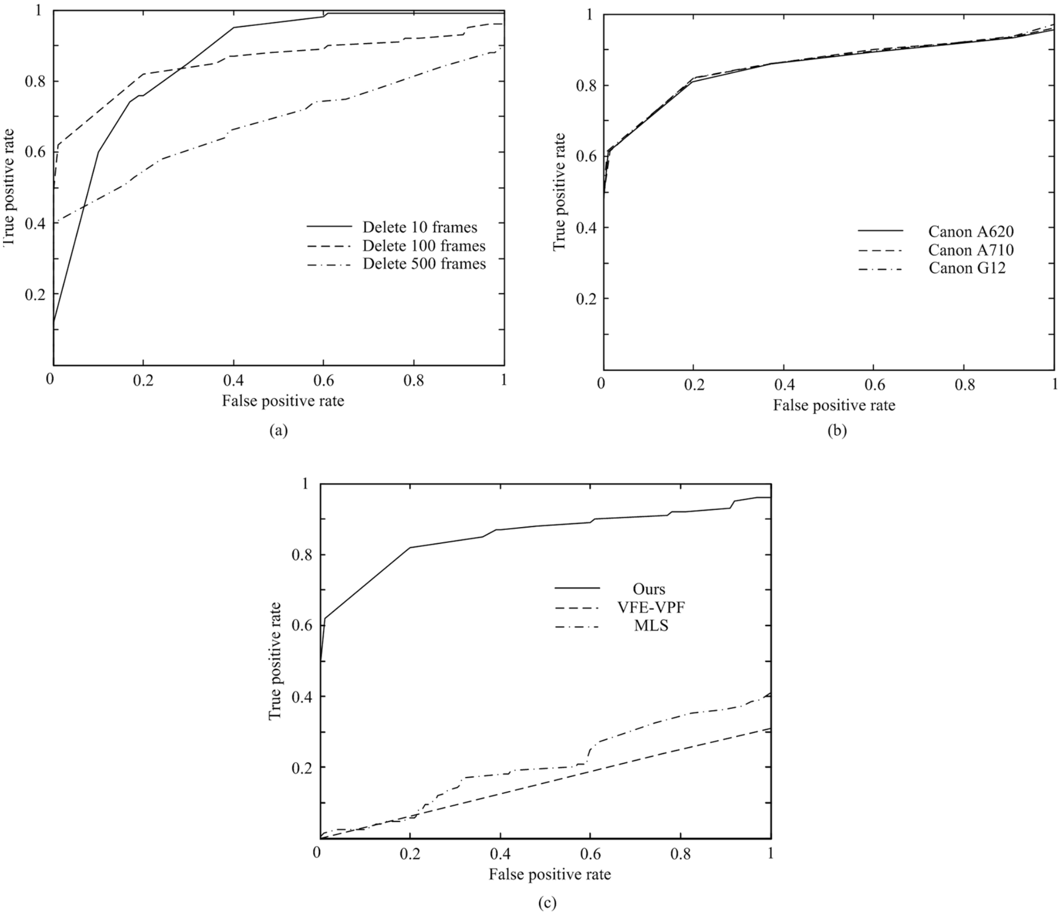
From Figure 10 (a) it can be observed that when the false positive rate is low, the true positive rate of 10 frames deletion forgery videos is the lowest among the three situations. The reason is that when the number of deleted frames is small, the difference between the frames at the forgery position is also small, and the sudden change of phase in the signal is not as significant as when the number of deleted frames is large. As a result, the peak in $C_t(i)$ is lower when the number of frame deletion is small. Low false positive rate means having a high detection threshold, and the peak in $C_t(i)$ of the 10 frames deletion forgery videos may not be detected by using these thresholds, so the true positive rate of 10 frames deletion forgery videos is the lowest in this situation.

Compare 100 frames deletion forgery videos and 500 frames deletion forgery videos, Figure 10 (a) shows that the performance of the proposed detection algorithm to 100 frames deletion forgery is better than the performance to the 500 frames deletion forgery. The reason for this phenomenon is the accumulation of phase change. As the analysis in Section 3.1 showed, the frequency shift will cause the phase to change faster than normal. Although the proposed algorithm can avoid the false alarm caused by frequency shift, the abnormal change of phase cannot be eliminated and will accumulate along the time. The longer the deleted sequence is, the higher the chance that the difference of phase at the forgery position is close to 2π. When the difference of the phase is close to 2π, the correlation coefficient value between the two periods will be close to 1, and the peak in $C_t(i)$ would disappear, which makes it impossible to detect the forgery. For this reason, the proposed algorithm has better performance for the 100 frames deletion forgery videos than the 500 frames.

With the increase of the false positive rate, the true positive rate of the 10 frames deletion forgery videos grows much faster than the other two forgery videos. This is also caused by the accumulation of phase change. When the length of the deleted sequence is short, the difference of the phase at the forgery position will be far from 2π. The peaks in $C_t(i)$ may not be very high due to the short deletion length, but they will not decrease to 0 either. When the false positive rate increases, the detection threshold decreases, and more peaks will be detected by our algorithm. On the other hand, when the peaks in $C_t(i)$ disappear because of the phase change accumulation, the forgery will not be detected no matter what the thresholds are set. For this reason, the performance of the 10 frames deletion forgery videos becomes the best with the increase of the false positive rate.

In summary, when the length of the deleted sequence is short, the phase difference of the periods at the forgery position is not very big, and the corresponding peak in $C_t(i)$ is not very high, which will reduce the true positive rate when the false positive rate is low. On the other hand, the short length of the deletion sequence will not produce a phase difference of 2π. In contrast, a long-deleted sequence will give a high peak in $C_t(i)$, but will also increase the possibility of peak disappearance

Figure 10. (a) The ROC curves of forgery videos deleted different number of frames; (b) the ROC curves of 100 frames deletion forgery videos taken by different camera models; (c) the ROC curves of 100 frames deletion forgery videos detected by different methods.



(a)

(b)

(c)

caused by phase change accumulation. Overall, the proposed algorithm is effective to detect frame deletion forgery.

We also compare the performances of the proposed algorithm on the three camera models. To the 100 frames deletion forgery videos, the ROC curves for different camera models are shown in Figure 10 (b). From Figure 10 (b) it can be observed that the proposed method has very close detection ability to the forgery videos taken by the three different camera models. That is to say, the proposed method has robustness to different camera models.

To compare our algorithm with VFE-VPF and MSL, we use the three methods to detect the same videos with 100 frames deletion forgery and get the ROC curves. The results are shown in Figure 10 (c). The figure indicates that the proposed algorithm has much better performance than the other two algorithms when detecting the forgery in surveillance videos with a large part of static scene. The VFE-VPF algorithm tries to detect the abnormal motions in the video, so its performance decreases seriously when most of the frames in video are static. The MSL algorithm uses the luminance difference between frames to detect the forgery, so it can catch the fluctuation of ENF signal to some degree and has better performance than VFE-VPF. However, the static content covers the weak ENF signal most of the time, and MLS cannot work well in this situation. On the other hand, the proposed algorithm extracts the ENF signal and eliminates the interference of content, so its performance exceeds the other two's significantly.

4.3 Detection of Frame Duplication Forgery

The frame duplication forgery only involves one video, which is similar to the frame deletion forgery. When the video has a large part of static scene, the frame duplication forgery is also hard to be noticed by the naked eye. In this situation, the proposed algorithm based on ENF signal will be effective to detect the forgery. In this experiment, a sequence of 100 frames, 500 frames or 1000 frames will be selected at a random position between 01:00 and 29:00 of a 30 minutes video. The selected sequence will be copied and used to replace another sequence at another random position to generate the forgery video. The two sequences do not have any overlap. We have several original videos, and each of them can be used to generate several forgery videos. The proposed detection algorithm is used to detect the forgery locations of the forgery videos.

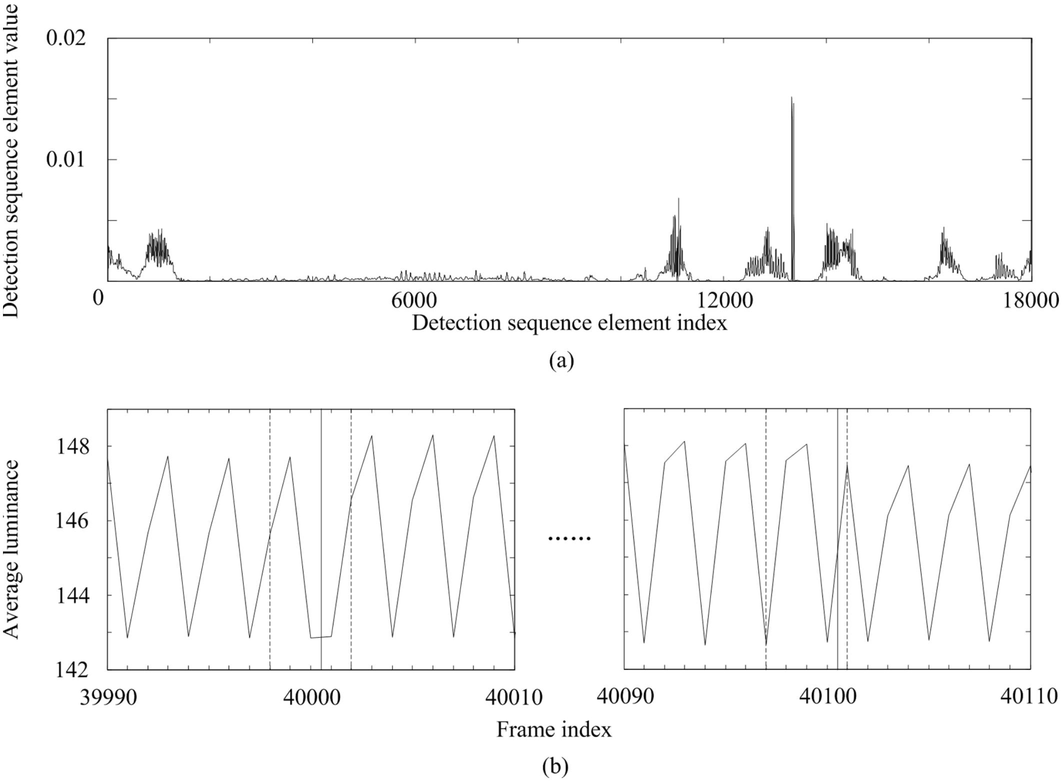
We select a video taken by Canon A620 and generate a forgery video by copying its 10001th to 10100th frames to replace its 40001th to 40100th frames. In this forgery video with 100 frames duplication there are two forgery positions, one is between the 40000th and the 40001st frame, i.e. at the start of the duplication, and the other is between the 40100th and the 40101th frame, at the end of the duplication. Using the proposed detection algorithm, we have the detection sequence $C_t(i)$ shown in Figure 11(a). By checking the positions of the peak in Figure 11(a), we have the detection result shown in Figure 11(b). The lines in Figure 11(b) have the same meaning as the ones in Figure 9(b). In this example, the detection ranges are (39998, 40002) and (40097, 40101), and both of them include the actual position of forgery. It shows the effectiveness of the proposed algorithm in detecting frame duplication forgery.

Then the thresholds are changed to calculate different true positive rates and false positive rates to generate the ROC curve. It should be noted that in a frame duplication forgery video there are two forgery positions, so the number of positive samples is twice the number of negative samples, and it must be considered during the calculation of true positive rates and false positive rates. The ROC curves for different length of duplicated sequence are shown in Figure 12 (a).

Compare Figure 12 (a) with Figure 10 (a), it can be observed that the performance of the proposed detection algorithm to frame duplication forgery videos is similar to the performance for the frame deletion forgery videos with long deleted sequence, and their ROC curves have similar shape. The reason for this is related to the process of frame duplication forgery. An example of frame duplication forgery is shown in Figure 13. The source sequence contains K frames and begins from the I_Cth frame in the video, and the replaced sequence begins from the I_Pth frame. In this situation, the two forgery positions in the video are between the I_P-1th and the I_Pth frame, the I_P+K-1th and the I_P+Kth frame respectively. We assume $I_C > I_P+K-1$ in this example. For the first forgery position, the frame behind it is the I_Cth frame from the same video, so the situation of this forgery position is the same as deleting the sequence from the I_Pth frame to the I_C-1th frame. The situation of the second forgery position is similar, and the frames beside it are the I_P+Kth frame and the I_C+K-1th frame in the video. Because the frame duplication forgery only involves one video, its forgery positions always have the same situation as the forgery position in frame deletion forgery, so the performance of the proposed algorithm for frame duplication forgery can be analyzed in the same way as in Section 4.2.

Most of the time the number of frames between the I_Pth and the I_Cth frame is much larger than K, so the performance for frame duplication forgery is similar to the performance for frame deletion forgery with long deleted sequence. The true positive rate reaches a certain level when the false positive rate is low and increases slowly with the rise of the false positive rate. In a few cases the positions of the source and replaced sequence are close, and the difference between I_P and I_C is close to K. In this situation the number of duplicated frames will begin to affect the detection performance. When K is large, the detection result has higher probability of been affected by the accumulation of phase change. So it can be seen in Figure 12 (a) that the proposed algorithm has better performance for the 100 frames duplication forgery than the other two situations. Overall, the proposed algorithm is effective to detect frame duplication forgery.

Figure 11. (a) The detection sequence used in the frame duplication forgery example; (b) the forgery position detection result of this example



Similar to Section 4.2, we compare the performances of the proposed algorithm to the 100 frames duplication forgery videos taken by the three camera models and give the ROC curves for different camera models in Figure 12 (b). The figure also shows the robustness to different camera models when using the proposed algorithm to detect frame duplication forgery. The 100 frames duplication forgery videos are also used for comparison of the three algorithms, and the ROC curves are shown in Figure 12 (c). Compare Figure 12 (c) with Figure 10 (c), the proposed algorithm still has a great advantage in detection performance, and the MLS algorithm has a little better performance than the VFE-VPF algorithm. The reason to this phenomenon is discussed in Section 4.2. Moreover, as we analyzed in this section earlier, both the VFE-VPF method and MLS method have similar performance when detecting 100 frames duplication forgery and 100 frames deletion forgery.

4.4 Detection of Frame Insertion Forgery

The frame insertion forgery involves at least two videos, so it is generally easier to be detected than the frame deletion and duplication forgery. However, when the videos are shot at the same place and have a large part of static scene, the forgery is also hard to be noticed by the naked eye, and the proposed algorithm based on ENF signal will be useful in this situation. In this experiment, several videos are selected as the fundamental videos. For each fundamental video, another video with similar content is selected as a material video. A random sequence is copied from the material video and inserted into the fundamental video at a random position between 01:00 and 29:00 of the 30 minutes video to generate the forgery video. The length of the inserted sequence can be 100 frames, 500 frames

Figure 12. (a) The ROC curves of forgery videos with different number of duplicated frames; (b) the ROC curves of 100 frames duplication forgery videos taken by different camera models; (c) the ROC curves of 100 frames duplication forgery videos detected by different methods

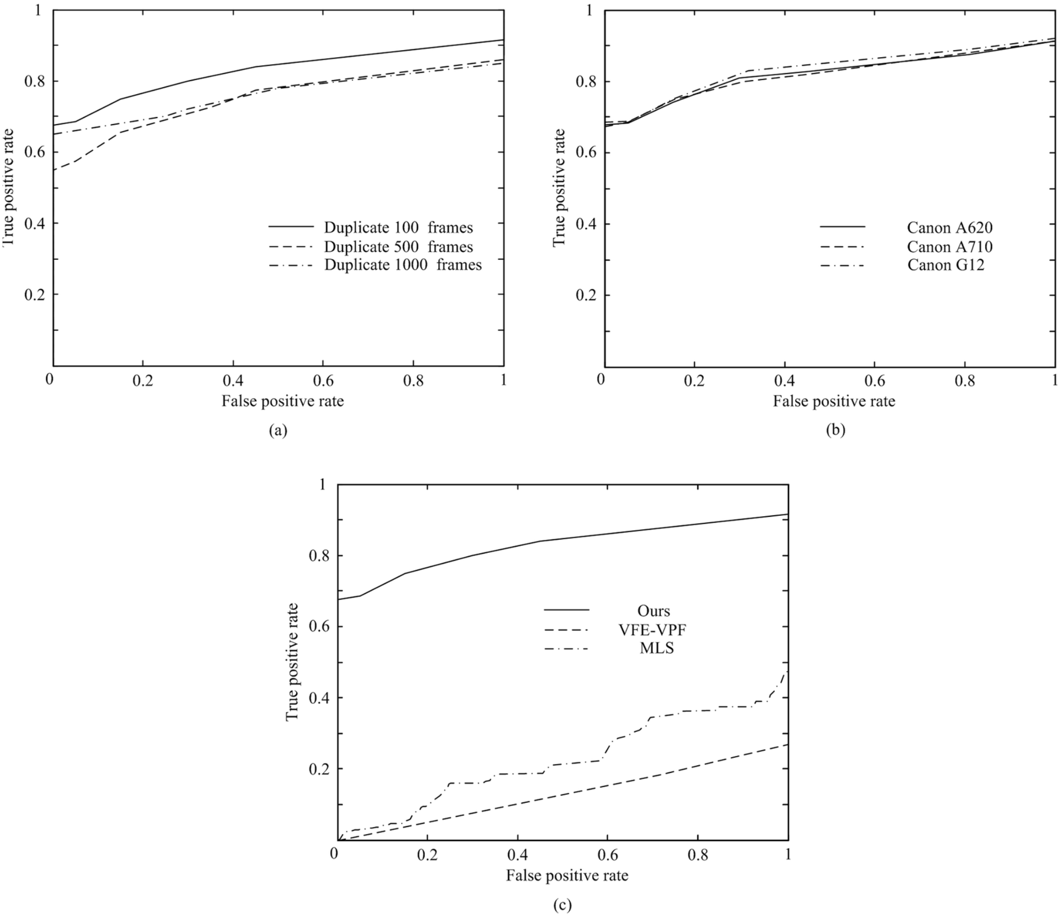
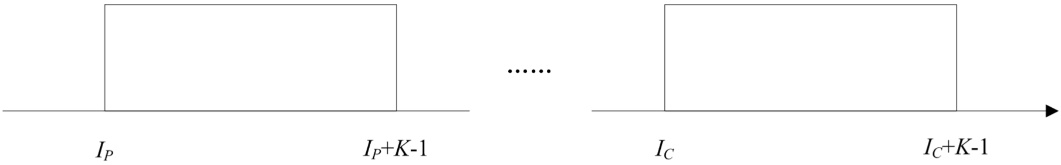


Figure 13. A frame duplication forgery example



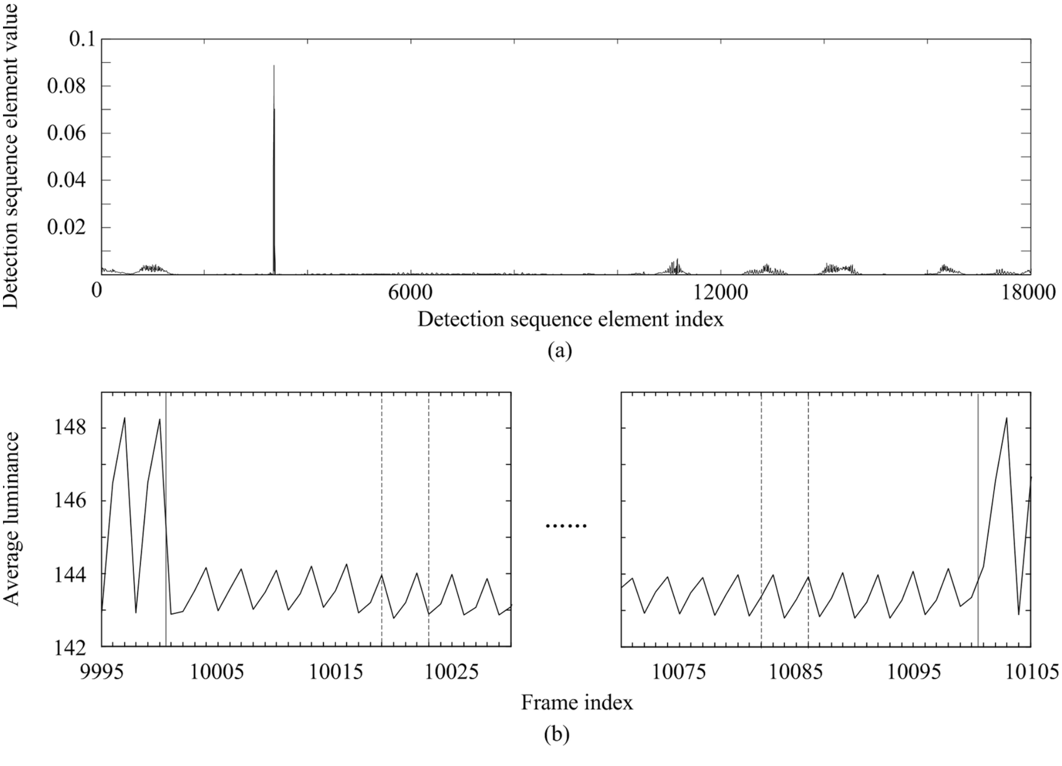
or 1000 frames, one fundamental video can generate several forgery videos using different material videos. Then the proposed detection algorithm is used to detect the forgery locations.

Once again, we will examine a specific example. We select a video taken by Canon A620 as a fundamental video, and select a video taken by the same camera at the same place but at a different date as a material video. A sequence from the 10001st to the 10100th frame in the material video is copied and inserted into the fundamental video after the 10000th frame to generate a forgery video. In this forgery video there are two forgery positions, one is between the 10000th and the 10001st frame, the other is between the 10100th and the 10101th frame. Using the proposed detection algorithm, we have the detection sequence $C_t(i)$ as shown in Figure 14(a). By checking the positions of the peak in Figure 14(a), we have the detection result as shown in Figure 14(b). The lines in Figure 14(b) also have the same meaning as the ones in Figure 9(b). In this example, the detection ranges are (10019,10023) and (10082,10086), which deviate by nearly 20 frames from the actual position of forgery. In summary, the proposed algorithm can detect the existence of frame insertion forgery but gives some deviation in the detection of forgery positions.

The reason for the deviation is that the average luminance of the fundamental video and the material video are significantly different, which can be observed in Figure 14(b). The sudden change of the average luminance can be considered as a step signal superposed on the original signal. Because of the significant difference, the accuracy of ENF signal extracted by the band pass filter will be reduced at the boundary of the step signal, and the additional data points generated by cubic spline interpolation will also be less accurate. As a result, the correlation coefficients value will also be incorrect. The boundary of the step signal is always at the forgery position, so the incorrectly extracted ENF signal will cause the deviation when detecting the forgery position. Although the forgery positions detected by the proposed algorithm are nearly 20 frames away from the actual positions, the algorithm can still detect the existence of the forgery. In other words, the proposed algorithm can be used to confirm the authenticity of the videos for frame insertion forgery.

Similar to the two experiments mentioned above, the thresholds are also changed to calculate different true positive rates and false positive rates and to generate the ROC curve. As in frame duplication forgery, in frame insertion forgery there are also two forgery positions in the forgery video, so the number of positive samples is also twice the number of negative samples. On the other hand, considering the deviation caused by the step signal effect, the detection range should be extended.

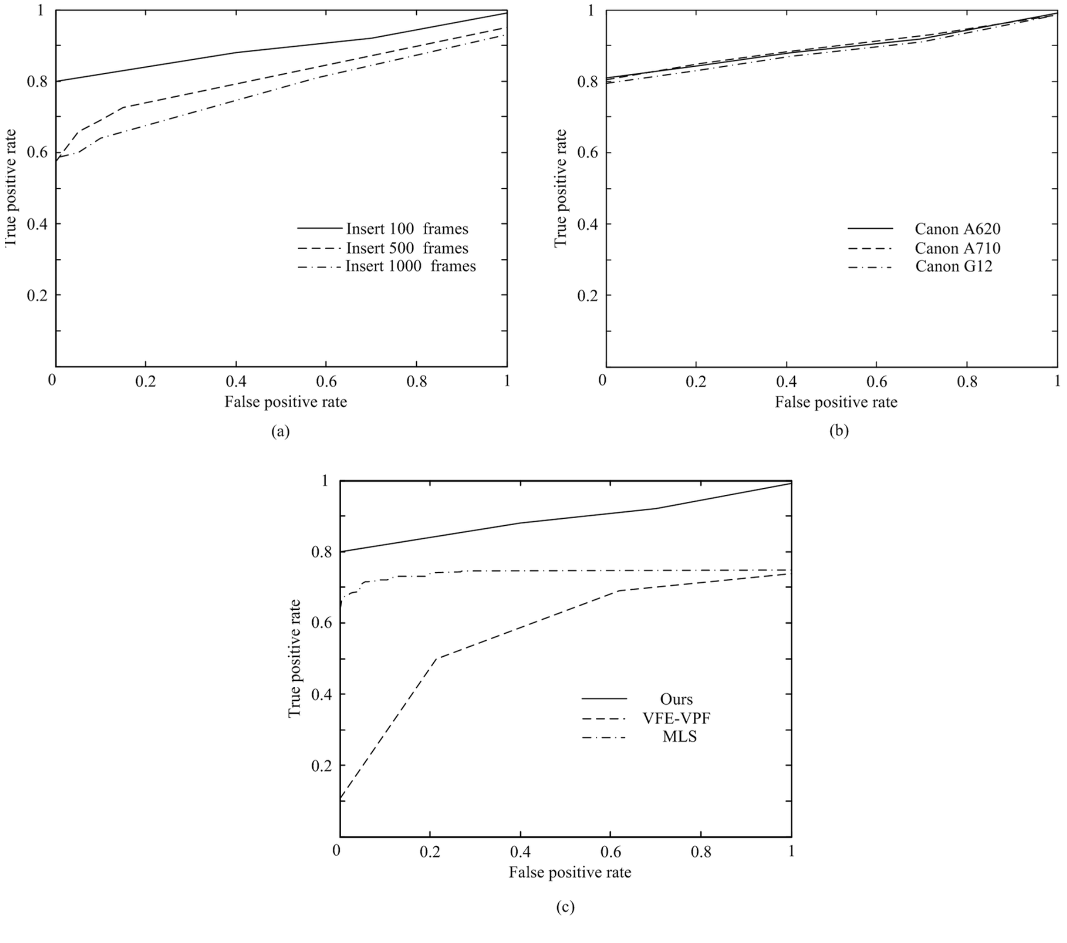
Figure 14. (a) The detection sequence used in the frame insertion forgery example; (b) the forgery position detection result of this example



Based on the experiment results, we find that the deviation will not exceed 30 frames, which is 1 second in time, so we extend 30 frames before and after the detected range respectively. The ROC curves for different length of inserted sequence are shown in Figure 15 (a).

Compare Figure 15 (a) with Figure 12 (a), it can be observed that the shape of the ROC curves of the proposed detection algorithm for frame insertion forgery videos is similar to the frame duplication forgery videos. Moreover, because the inserted video sequence is from another video, the difference between the frames at both sides of the forgery position is more significant than that in the frame duplication forgery videos, so the true positive rate is higher for the frame insertion forgery videos. The ENF signal is always associated with time, so even though the fundamental video and the material video are taken by the same camera and at the same place, the different shooting time will produce different ENF signals. When the fundamental video and the material video are taken by the same camera at the same position, such as in surveillance videos, the frames beside the forgery position have similar characteristics as the very long sequence deletion forgery, so their properties can be analyzed in the same way. Similar to the situation in frame duplication forgery videos, the true positive rate is high when the false positive rate is low, but increases slowly with the rise of the false positive rate. The length of the inserted sequence also has the same influence as analyzed earlier. The more frames

Figure 15. (a) The ROC curves of forgery videos with different number of inserted frames; (b) the ROC curves of 100 frames insertion forgery videos taken by different camera models; (c) the ROC curves of 100 frames insertion forgery videos detected by different methods



(a)

(b)

(c)

are inserted, the higher the probability the performance will decline because of the accumulation of phase change. The peak in the detection sequence $C_t(i)$ will decrease because of the accumulation, which has negative impact on the true positive rate. For this reason, it can be observed from Figure 15 (a) that the performance of the proposed algorithm to the 100 frames insertion forgery videos is better than the performance of the 500 frames and 1000 frames insertion. The performance of the proposed algorithm is also affected by the fluctuation of ENF signal in the fundamental and material videos, and the performance has some randomness when the forgery videos are generated randomly. However, Figure 15 shows that the overall performance of the proposed algorithm is still good and is effective for frame insertion forgery detection.

Figure 15 (b) shows the ROC curves of the proposed algorithm detecting the 100 frames insertion forgery videos taken by the three different camera models, and it indicates the robustness of the proposed algorithm to different camera models again. We also use the methods for comparison to detect the 100 frames insertion forgery videos, and the ROC curves are shown in Figure 15 (c). Because of the more significant difference between the frames at both sides of the forgery position in the frame insertion forgery videos mentioned earlier in the section, the VFE-VPF and MLS also have much better performance when detection frame insertion forgery. However, as we analyzed in Section 4.2 and 4.3, the proposed algorithm still has the best detection performance to frame insertion forgery.

5. CONCLUSION

An inter-frame video forgery detection algorithm is proposed in this paper. The algorithm extracts the ENF signal from the suspected video by band pass filtering and uses cubic spline interpolation to handle the problem of lack of data. In the proposed algorithm, the correlation coefficient between each pair of adjacent periods in the interpolated ENF signal is first obtained, and the sudden decrease in the correlation coefficient value is used to detect the existence and the exact position of the forgery. Different from other ENF signal based detection algorithms, the proposed algorithm uses only the extracted ENF signal for forgery detection and does not need a reference ENF signal from the power grid. Nowadays, there are few ENF signal databases available, and this situation makes the reference needed forgery detection algorithms unable to be used most of the time. Moreover, building an ENF database will cost a lot of resources. The proposed reference free algorithm can overcome these shortcomings and is more convenient to be used in practice. The results of the experiments show that the proposed algorithm has good performance for inter-frame video forgery detection, such as frame deletion, frame duplication, and frame insertion. The experiments also indicate that the proposed algorithm has superior performance compared with some state-of-the-art inter-frame forgery detection algorithms when detecting the surveillance videos with static scenes. In future research, we will extend the application scope to other type of video forgery.

ACKNOWLEDGMENT

This work was partially supported by Guangzhou City Science and Technology Foundation (Project No. 201510010275), Science and Technology Foundation of Guangdong Province (Project No. 2017A050501002), Sino-Singapore International Joint Research Institute (Project No. 206-A017023 and 206-A018001), and EU Horizon 2020 project under Grant 690907.

REFERENCES

Brixen, E. B. (2008). ENF; Quantification of the magnetic field. In *Audio Engineering Society Conference: 33rd International Conference: Audio Forensics-Theory and Practice.* Audio Engineering Society.

Cooper, A. J. (2008). The electric network frequency (ENF) as an aid to authenticating forensic digital audio recordings–an automated approach. In *Audio Engineering Society Conference: 33rd International Conference: Audio Forensics-Theory and Practice.* Audio Engineering Society.

Cooper, A. J. (2011). Further considerations for the analysis of ENF data for forensic audio and video applications. *International Journal of Speech Language and the Law*, *18*(1), 99–120. doi:10.1558/ijsll.v18i1.99

Elmesalawy, M. M., & Eissa, M. M. (2014). New forensic ENF reference database for media recording authentication based on harmony search technique using GIS and wide area frequency measurements. *IEEE Transactions on Information Forensics and Security*, *9*(4), 633–644. doi:10.1109/TIFS.2014.2304838

Fu, L., Markham, P. N., Conners, R. W., & Liu, Y. (2013). An improved discrete fourier transform-based algorithm for electric network frequency extraction. *IEEE Transactions on Information Forensics and Security*, *8*(7), 1173–1181. doi:10.1109/TIFS.2013.2265088

Garg, R., Varna, A. L., Hajj-Ahmad, A., & Wu, M. (2013). “Seeing” ENF: Power-signature-based timestamp for digital multimedia via optical sensing and signal processing. *IEEE Transactions on Information Forensics and Security*, *8*(9), 1417–1432. doi:10.1109/TIFS.2013.2272217

Garg, R., Varna, A. L., & Wu, M. (2011). Seeing ENF: natural time stamp for digital video via optical sensing and signal processing. In *Proceedings of the 19th ACM international conference on Multimedia* (pp. 23-32). ACM. doi:10.1145/2072298.2072303

Grigoras, C. (2005). Digital audio recording analysis–the electric network frequency criterion. *International Journal of Speech Language and the Law*, *12*(1), 63–76. doi:10.1558/sll.2005.12.1.63

Grigoras, C. (2007). Applications of ENF criterion in forensic audio, video, computer and telecommunication analysis. *Forensic Science International*, *167*(2), 136–145. doi:10.1016/j.forsciint.2006.06.033 PMID:16884872

Grigoras, C. (2009). Applications of ENF analysis in forensic authentication of digital audio and video recordings. *Journal of the Audio Engineering Society*, *57*(9), 643–661.

Hajj-Ahmad, A., Berkovich, A., & Wu, M. (2016). Exploiting power signatures for camera forensics. *IEEE Signal Processing Letters*, *23*(5), 713–717. doi:10.1109/LSP.2016.2537201

Hajj-Ahmad, A., Garg, R., & Wu, M. (2013). ENF based location classification of sensor recordings. In *Proceedings of the 2013 IEEE International Workshop on Information Forensics and Security (WIFS)* (pp. 138-143). IEEE. doi:10.1109/WIFS.2013.6707808

Hajj-Ahmad, A., Garg, R., & Wu, M. (2015). ENF-based region-of-recording identification for media signals. *IEEE Transactions on Information Forensics and Security*, *10*(6), 1125–1136. doi:10.1109/TIFS.2015.2398367

Huang, C. C., Zhang, Y., & Thing, V. L. L. (2017). Inter-frame video forgery detection based on multi-level subtraction approach for realistic video forensic applications. In *Proceedings of the 2017 IEEE International Conference on Signal and Image Processing (ICSIP)* (pp. 20-24). IEEE. doi:10.1109/SIPROCESS.2017.8124498

Huijbregtse, M., & Geradts, Z. (2009, August). Using the ENF criterion for determining the time of recording of short digital audio recordings. In *Proceedings of the International Workshop on Computational Forensics* (pp. 116-124). Springer. doi:10.1007/978-3-642-03521-0_11

Kajstura, M., Trawinska, A., & Hebenstreit, J. (2005). Application of the electrical network frequency (ENF) criterion: A case of a digital recording. *Forensic Science International*, *155*(2), 165–171. doi:10.1016/j.forsciint.2004.11.015 PMID:16226153

Liu, Y., Yuan, Z., Markham, P. N., Conners, R. W., & Liu, Y. (2012). Application of power system frequency for digital audio authentication. *IEEE Transactions on Power Delivery*, *27*(4), 1820–1828. doi:10.1109/TPWRD.2012.2198892

Nicolalde, D. P., & Apolinario, J. A. (2009). Evaluating digital audio authenticity with spectral distances and ENF phase change. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP 2009* (pp. 1417-1420). IEEE. doi:10.1109/ICASSP.2009.4959859

Rodríguez, D. P. N., Apolinário, J. A., & Biscainho, L. W. P. (2010). Audio authenticity: Detecting ENF discontinuity with high precision phase analysis. *IEEE Transactions on Information Forensics and Security*, *5*(3), 534–543. doi:10.1109/TIFS.2010.2051270

Sanders, R. W. (2008). Digital audio authenticity using the electric network frequency. In *Audio Engineering Society Conference: 33rd International Conference: Audio Forensics-Theory and Practice.* Audio Engineering Society.

Sitara, K., & Mehtre, B. M. (2017). A comprehensive approach for exposing inter-frame video forgeries. In *Proceedings of the 2017 IEEE Colloquium on Signal Processing & its Applications (CSPA)* (pp. 73-78). IEEE. doi:10.1109/CSPA.2017.8064927

Su, H., Hajj-Ahmad, A., Garg, R., & Wu, M. (2014a). Exploiting rolling shutter for ENF signal extraction from video. In *Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP)* (pp. 5367-5371). IEEE. doi:10.1109/ICIP.2014.7026086

Su, H., Hajj-Ahmad, A., Wu, M., & Oard, D. W. (2014b). Exploring the use of ENF for multimedia synchronization. In *Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* (pp. 4613-4617). IEEE. doi:10.1109/ICASSP.2014.6854476

Yufei Wang received a B.E. degree in information engineering from South China University of Technology in 2010. Currently, he is a Ph.D. candidate in School of Electronic and Information Engineering at South China University of Technology. His research interests include information forensics, steganography and steganalysis, computer vision, and machine learning.

Yongjian Hu received a Ph.D. degree in communication and information systems from South China University of Technology in 2002. He is a professor with the School of Electronic and Information Engineering, South China University of Technology. Dr. Hu is also a senior member of Chinese Institute of Electronics (CIE) and a senior member of China Computer Federation (CCF). He has published more than 60 peer reviewed papers since 2000. His research interests include information hiding, multimedia security and machine learning.

Alan Wee-Chung Liew is currently an Associate Professor with the School of Information and Communication Technology, Griffith University, Australia. Previously, he has been an Assistant Professor in the Department of Computer Science and Engineering at The Chinese University of Hong Kong, and a senior research fellow in the Department of Electronic Engineering at the City University of Hong Kong. His research interests include pattern recognition and machine learning, medical imaging, computer vision, and bioinformatics.

Chang-Tsun Li received the BEng degree in electrical engineering from National Defence University (NDU), Taiwan, in 1987, the MSc degree in computer science from U.S. Naval Postgraduate School, USA, in 1992, and the PhD degree in computer science from the University of Warwick, UK, in 1998. He was an associate professor of the Department of Electrical Engineering at NDU during 1998-2002 and a visiting professor of the Department of Computer Science at U.S. Naval Postgraduate School in the second half of 2001. He was a professor of the Department of Computer Science at the University of Warwick, UK, until Dec 2016. He is currently Professor of Cyber Security at Deakin University, Australia. His research interests include multimedia forensics and security, biometrics, data mining, machine learning, data analytics, computer vision, image processing, pattern recognition, bioinformatics, and content-based image retrieval. He is currently Associate Editor of the EURASIP Journal of Image and Video Processing (JIVP) and Associate of Editor of IET Biometrics. He has been involved in the organisation of a number of international conferences and workshops and also served as member of the international program committees for several international conferences. He was the Lead and PI of the international joint project entitled Digital Image and Video Forensics (acronym: DIVEFOR) funded through the Marie Curie Action under the EU's Seventh Framework Programme (FP7) from June 2010 to May 2014. He is currently the Lead and PI of the EU Horizon 2020 project, entitled Computer Vision Enabled Multimedia Forensics and People Identification (acronym: IDENTITY).