A CHAOTIC AUTHENTICATION TECHNIQUE FOR DIGITAL VIDEO SURVEILLANCE

Siyue Chen and Henry Leung

Department of Electrical and Computer Engineering, University of Calgary
2500 University Drive NW, Calgary, Canada, T2N 1N4

ABSTRACT

Installing video cameras becomes more and more popular in public facilities for the surveillance applications. However, given the ease with which video contents can be manipulated by popular editing softwares, video records do not have any value as legal proofs in applications such as criminal evidence and insurance claims. This paper presents a novel chaotic authentication technique to solve the problem. The temporal information of each frame is modulated into the parameters of a chaotic system. The system output is a random-like signal and used as a watermark embedded into DCT coefficients. The ergodic property of chaotic signals is employed to demodulate the embedded information. It is shown that the proposed scheme not only can survive video compression, but it can also localize both spatial and temporal tampering.

1. INTRODUCTION

Over the last three years, the digital video surveillance (DVS) market has shown an unprecedented boom. A search on the term "digital video surveillance" yielded 1,110,000 results on Google. The reasons for this surge range from the responses to 911 to the rise of IP infrastructure and the availability of high performance, low-cost video processors. However, some practical issues must be taken into account in order to take full advantage of DVS systems. The authenticity of visual data acquired, processed and possibly stored by DVS systems is one of such issues.

The very first research efforts for video authentication is cryptography. The major drawback of such an approach is that it provides a complete verification [1]. In other words, the video data allows no modification after encryption. This constraint might be too restrictive in the DVS application. For example, although compression introduces only imperceptible changes on video data, the digests calculated based on the raw and the compressed data are quite different. Authentication based on watermarking provides an alternative solution. The main advantage of this technique lies on the easy achievements of tamper localization due to the fact that the embedded watermark would be lost or altered as soon as the host data undergoes some modification.

In this paper, we propose a novel semi-fragile watermarking scheme based on chaos theory for video authentication. The information modulated into the watermark is robust to compression, but vulnerable to malicious attacks, thus provides an authenticity check on video data. Furthermore, the watermark is designed to be frame-dependent and time-varying. Therefore, inspection on the difference between the original and the extracted watermark can indicate where and how the tampering is performed. The paper starts with the discussion on the peculiarities of DVS data authentication. We then present our chaotic authentication scheme. Experimental results demonstrate its effectiveness. Finally, the concluding remarks are given.

2. PECULIARITIES OF DVS AUTHENTICATION AND THE PROPOSED SOLUTIONS

Video is distinct in its three dimensional property. Tampering of video thus always falls into the two categories: spatial tampering and temporal tampering. Spatial tampering refers to modification on the image frame, such as content adding and removal. It can be detected by comparing the extracted watermark with the original one. The detection precision depends on where the watermark is embedded, ranging from pixel level to frame level. In our design, we use two 8×8 blocks to embed one watermark. Although the precision of doing so is not as high as that of watermarking in pixels, the computation load is greatly reduced. This is crucial for DVS authentication, since it requires a real-time watermark embedding.

Temporal tampering is manipulation on time axis, such as adding, dropping or reordering the video segments. To make sure the integrity of DVS data as a time sequence, an indissoluble link between the data and the time when it is produced would be created. Though many solutions are possible, the easiest way is to embed frame numbers into the corresponding frames. In our scheme, the frame number is used to generate watermarks by chaotic parameter modulation (CPM) [3]. That is, the information about the frame number is modulated into the bifurcating parameter and the initial condition of a chaotic system. In the retrieval, it is compared with the counted one. If they do not match, tem-

poral tampering can be claimed. It should be noted that the embedded information is able to survive the compression process. Compression thus will not introduce the false alarming of tampering.

As addressed in [2], an automated visual inspection (AVI) system is usually used in DVS to analyze video sequences by exploiting the time coherence of natural object motions. This means a certain degree of correlation is assumed to exist between subsequent frames. Uncorrelated frame differences are then filtered out as noise. Therefore, it is important that watermarks embedded in successive frames are highly uncorrelated so that the operation of AVI will not be disturbed by watermarks. When the frame number is used to modulate the watermark, it is difficult for conventional watermarking schemes [4] to generate uncorrelated watermarks since frame numbers of subsequent frames are highly correlated. However, the proposed watermark is the output of a chaotic system, which is sensitive to bifurcating parameters and initial conditions. Even though the same bifurcating parameter is used, only a slight difference in the initial conditions will produce sequences with a low cross-correlation [5].

3. CHAOTIC AUTHENTICATION SCHEME

The process for generating and embedding watermark by the proposed scheme is illustrated in Fig. 1. The frame

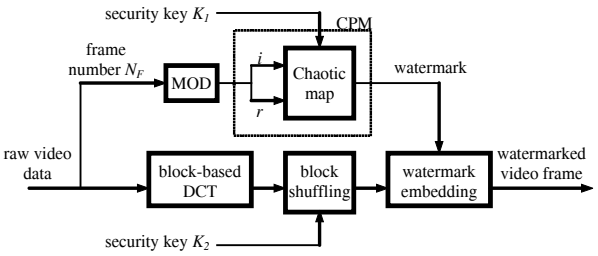


Fig. 1. The chaotic scheme for watermark generation and embedding.

number is first decomposed into $N_F = i \times \Delta + r$ by the operation of "MOD", where N_F is the frame number, and r is the modulus of N_F divided by Δ, ranging from 0 to $\Delta - 1$. Δ is kept secret by the authorized party. It is a parameter to control the robustness performance, as we will show later. Provided i and r known, the frame number can be recovered. Thus, they are used as an representative of the frame number and embedded into the video frame. CPM is then employed to generate watermarks. More precisely, r is modulated into the bifurcating parameter by $\lambda = \mathcal{C}_\lambda(r)$, and i is modulated into the initial condition by $\mathbf{x}_0 = \mathcal{C}_i(i)$. Although there are many choices, one possible way to construct $\mathcal{C}_\lambda$ and $\mathcal{C}_i$ can be

$$\begin{array}{rcl} \mathcal{C}_\lambda(r) & = & \lambda_{min} + \frac{\lambda_{max}-\lambda_{min}}{\Delta-1} r, \\ \mathcal{C}_i(i) & = & (-1)^b \frac{i}{\max(i)}, \end{array} \quad (1)$$

where $\max(i)$ returns the maximum value i can take. b randomly takes one from "1" and "2" based on a certain security key K_1. In (1), we assume that the output of $\mathcal{C}_\lambda$ has to be within the range of $[\lambda_{min}, \lambda_{max}]$, and x_0 falls into $[-1, 1]$. The watermark is then generated by iterating the states of a chaotic system. In this paper, we use the Chebyshev map, which is defined by $x_n = cos(\lambda \arccos(x_{n-1}))$ with $\lambda_{min} = 1.3$, $\lambda_{max} = 2$ and $n = 1, 2, \ldots$.

Embedding of the watermark is performed in the DCT domain for the robustness consideration on compression. The image frame is segmented into 8×8 blocks. The blocks are then formed into block pairs using a pre-determined mapping function $\mathcal{B}$ with another security key K_2. For instance, for a block p, we use $\mathcal{B}$ to choose a counterpart block to form a block pair, i.e., $q = \mathcal{B}(p, K_2)$. It should be noted that the construction of $\mathcal{B}$ has to make sure the selection of block pair is a random process. For each block, the absolute sum of the mid-frequency coefficients is calculated, i.e., $E(p) = \sum_{j \in \text{mid-frequency in } p} |d_j|$, where d_j denotes the DCT coefficients. The difference of them is used to hide watermark x_n. That is

$$E(p) - E(q) = \alpha x_n, \quad (2)$$

where α is a scalar to control the watermark strength. However, the original $E(p) - E(q)$ might not satisfy (2). Therefore, d_j in one of the blocks has to be modified. The decision which one is to be modified depends on the comparison of the variances in coefficients. A block with a larger variance is selected since it usually can tolerate more modifications without affecting its visual quality. Without loss of generality, the block p is assumed to have a larger variance. d_j in the block p is thus changed to

$$d'_j = d_j + \frac{E(q) + \alpha x_n - E(p)}{N_{\text{mid}}}, \quad (3)$$

so that (2) is satisfied. N_{mid} is the number of the mid-frequency coefficients. Using the mid-frequency coefficients to embed watermarks is because it can achieve a good trade-off between robustness and imperceptibility. As known, although watermarking in low-frequency coefficients is more robust, it is more likely to produce visible noises. Watermarking in high frequencies has the inverse results.

The watermark detection is sort of an inverse of the embedding process. The image frame is reconstructed with the possibly compressed data sequences. Provided K_2 is known, the block pair (p, q) can be still found by applying $\mathcal{B}$. The watermark extraction can be expressed by

$$\hat{x}_n = \frac{\hat{E}(p) - \hat{E}(q)}{\alpha}. \quad (4)$$

To demodulate the embedded information, the ensemble average of the extracted watermark is calculated, i.e., $\hat{\mu} = \frac{\hat{x}_1+\hat{x}_2+\cdots+\hat{x}_L}{L}$, where L is the length of the watermark. For the Chebyshev map, it is known that the ergodic measure $\mu = \lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} x_n$ is a monotonic increasing function of the bifurcating parameter λ, and independent of initial conditions [6]. If we can interpret this relationship by the so called mean value function $\mathcal{M}$, i.e., $\mu = \mathcal{M}(\lambda)$, the bifurcating parameter can be decoded by inverting $\mathcal{M}$, i.e., $\hat{\lambda} = \mathcal{M}^{-1}(\hat{\mu})$. And the modulus of the frame number, r, can be demodulated by

$$\hat{r} = \lfloor \mathcal{C}_\lambda^{-1}(\hat{\lambda}) \rceil = \lfloor \mathcal{C}_\lambda^{-1}(\mathcal{M}^{-1}(\hat{\mu})) \rceil, \tag{5}$$

where $\lfloor\cdot\rceil$ denotes a round operator, since $\hat{r}$ can only take Δ discrete values according to the generation of watermarks.

Observing (5), we find that if $\mathcal{C}_\lambda^{-1}(\hat{\lambda})$ falls into the region of $[r-0.5, r+0.5)$, $\hat{r} = r$. The accurate estimation actually relies on wether

$$\lambda - \frac{\lambda_{max}-\lambda_{min}}{2(\Delta-1)} \le \hat{\lambda} < \lambda + \frac{\lambda_{max}-\lambda_{min}}{2(\Delta-1)} \tag{6}$$

can be satisfied or not. Recalling $\hat{\lambda} = \mathcal{M}^{-1}(\hat{\mu})$, and $\mathcal{M}$ is a monotonic increasing function for the chebyshev map, it can be further deduced from (6) that $\hat{\mu}$ must also be within a certain region centering around μ, i.e., approximately $\mu - \delta \le \hat{\mu} < \mu + \delta$, where

$$\delta = \min \left(\begin{array}{c} \mu - \mathcal{M}\left(\lambda - \frac{\lambda_{max}-\lambda_{min}}{2(\Delta-1)}\right), \\ \mathcal{M}\left(\lambda + \frac{\lambda_{max}-\lambda_{min}}{2(\Delta-1)}\right) - \mu \end{array} \right). \tag{7}$$

It is seen that a larger value of δ indicates a more robust performance of the scheme, since the scheme can tolerate more distortions in detection. We further find from (7) that δ is determined by Δ. For the robustness consideration, a small value of Δ would be desirable. However, it reduces the effctiveness of tampering detection. The smaller the value of Δ is, the more chance a multiple of Δ frames can be dropped or added without being detected by comparing $\hat{r}$ with $\hat{r}'$, since the period of r from 0 to $\Delta-1$ is still maintained in the check. Although we can also depend on the analysis of i to detect such tampering, it would be better to reduce such chances as much as possible.

To illustrate the process of tampering detection, let's assume the frame in investigation has the frame number of $\hat{N}_F$. Provided Δ is known to the authorized party, $\hat{N}_F$ can also be decomposed into $\hat{N}_F = \hat{i}' \times \Delta + \hat{r}'$. Comparing $\hat{r}$ with $\hat{r}'$, temporal tampering can be claimed if they are not matching. Otherwise, $\hat{r}'$ and $\hat{i}'$ are used to generate a sequence of chaotic signal $\hat{x}'_1, \hat{x}'_2, \ldots, \hat{x}'_L$. Comparing $\hat{x}'_1, \hat{x}'_2, \ldots, \hat{x}'_L$ with the extracted one as in (4), i.e., $\hat{x}_1$, $\hat{x}_2$, …, $\hat{x}_L$, they would be identical if no tampering is applied. However, if the difference exists between them, we can set a threshold T_x to determine whether a detection error occurs. That is, if $|\hat{x}_n - \hat{x}'_n| > T_x$, it is decided that a detection error occurs and $e_n = 1$, otherwise $e_n = 0$. If temporal tampering by chance results in $\hat{r}' = \hat{r}'$, such as dropping or adding a multiple of Δ frames, there will be a high probability of $e_n = 1$ due to $\hat{i}' \neq \hat{i}$ and the sensitivity of a chaotic systems to initial conditions. Meanwhile, the detection error is spread all over the frame due to the random selection of block pairs. When spatial tampering is applied, such as replacing, adding or removing objects, we will have specially dense detection errors found in a certain regions. The effect of compression distinguished from of tampering lies in the fact that the detection has $\hat{r}' = \hat{r}$ and the probability of $e_n = 1$ is very low.

4. EXPERIMENTAL RESULTS

Experiments were conducted to evaluate the performance of the proposed scheme. A representative color DVS sequence composed of 265 frames of size 240×320 has been used in all our experiments. Experimental results suggest that the mid-frequency would better be set as the positions from 9 to 28 in the zig-zag order, because watermarking in these coefficients can achieve a good tradeoff between robustness and imperceptibility.

It should be noted that since VS data is usually acquired by inexpensive, low quality devices, and is used for identify behaviors or classify events, the requirement of imperceptibility is not that crucial compared to that for arts and media. Unless the distortion introduced by watermarking does not disturb the performance of AVI, the value of α can be set relatively larger in DVS than in other classical authentication applications. Furthermore, the selection of α is not necessarily adaptive based on human visual system. In this study, we just use a global α which can satisfy PSTN $\ge$ 40dB for each frame. The proposed scheme is also shown fast enough to achieve 28 frame per second in watermark embedding.

4.1. Video compression

The MPEG-2 coding algorithm, which is one of the most popular video compression technique, is employed in this experiment. As addressed in Section 3, the robustness of the proposed scheme relies on the selection of Δ. A larger Δ is better for the security consideration, however usually provides less robustness. The maximum value of Δ for the embedded r to survive compression with a certain quality factor is thus the interest of investigation. Fig. 2 reports the results. It is seen that when the quality factor is larger than 50%, the maximum value of Δ remains stable at 119 or even larger. Since the frame rate for DVS is usually 30 frames per second, 119 indicates the tampering has to add or drop a multiple of 4-second video sequence to remain the

complete period of $\hat{r}$. However, this is very difficult to implement without introducing abrupt visual changes. Therefore, the proposed scheme can survive the high quality compression with a proper selected Δ, and maintain its function on authenticity check.

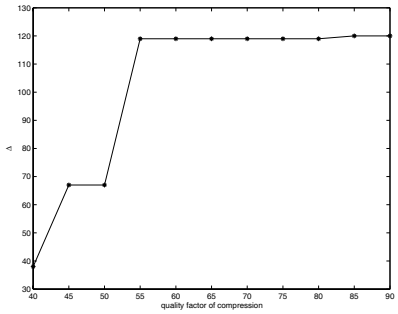


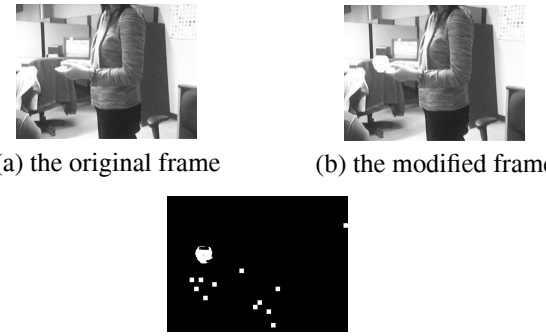
Fig. 2. The maximal value of Δ that can take for the proposed scheme to survive the compression with a certain quality factor.

4.2. Spatial tampering

In this experiment, we modify the original frame by adding a lighting bulb on the girl's hand, as shown in Fig. 3(a) and (b). The detection result is illustrated in a tampering detection image, which is formed by setting all pixel value of the blocks having $e_n = 1$ to 1 (white), and those having $e_n = 0$ to 0 (dark). The detection result by the proposed scheme is shown in Fig. 3(c). It can be seen that the tampering is successfully localized. It should be pointed out that the proposed scheme has the drawback that it does not know which block in the pair should take the major responsibility to the detection error. Therefore, besides the location where the tampering is really applied, some extra blocks are also marked with white. However, because of the block shuffling, the wrong marked block should be uniformly distributed all over the frame. In this case, more attention should be paid for the area where the marked block is dense.

4.3. Temporal tampering

In this experiment, Δ is set as 119, and we randomly drop the frames from 160 to 170. It is found that after one period from 0 to 118, the extracted value of r is missing from 41 to 51. We also duplicate the frames from 1 to 119, and insert them between the original 119 and 120 frame. Although $\hat{r}'$ always matches with $\hat{r}$, the large number of detection errors all over the frame, which results in an almost white tampering detection image, reveals such modifications.



(a) the original frame (b) the modified frame

(c) the tampering detection image

Fig. 3. Illustration of a spatial tampering example and the corresponding tampering detection image by the use of the proposed scheme

5. CONCLUSIONS

This paper presents a novel chaotic authentication scheme for digital video surveillance (DVS). The chaotic watermark is frame-dependent, time-varying, thus can be used to detect spatial and temporal tampering. In addition, the proposed scheme can survive the video compression. Because of its simplicity to implement, it can be successfully applied on raw video data within cameras. Experimental results illustrate its effectiveness.

6. REFERENCES

[1] G. Doërr and J-L. Dugelay, "A guide tour of video watermarking", *Signal Processing: Image Communication*, vol. 18, no. 4, pp. 263–282, 2003.

[2] F. Bartolini, A. Tefas, M. Barni and I. Pitas, "Image authentication technique for surveillance application," *Proceedings of the IEEE*, vol. 89, no. 10, pp. 1403–1418, 2001.

[3] S. Chen and H. Leung, "Chaotic parameter modulation with application to digital watermarking," *Proceedings of ICASSP*, Orlando, FL, 2002.

[4] J. R. Hernandez, F. Perez-Gonzales, J. M. Rodriguez and G. Nieto, "Performance analysis of a 2-D-multipulse amplitude modulation scheme for data hiding and watermarking of still images," *IEEE Journal on Selected Areas in Communications*, vol. 4, no. 5, pp. 510–524, 1998.

[5] G. Heidari-Bateni and C.C. McGillem, "A chaotic direct-sequence spread spectrum communication system," *IEEE Trans. Commun.*, vol. 42, pp. 1524–1527, 1994.

[6] H. Leung, H. Yu and K. Murali, "Ergodic chaos-based communication schemes," *Physical Review E*, vol. 66, no. 036203, pp. 1–8, Sept 2002.