VIDEO WATERMARKING SYNCHRONIZATION BASED ON PROFILE STATISTICS

Shih-Wei Sun and Pao-Chi Chang

Electrical Engineering, National Central University, Chung-Li, Taiwan
e-mail: {swsun, pcchang}@vaplab.ee.ncu.edu.tw

ABSTRACT

We propose a novel temporal synchronization method for video watermarking by matching the profile statistics. The profile statistics, represented by the characteristic parameters such as position mean and variance in x- and y-directions, of a frame in a video sequence can easily be calculated and sent as the side information to the receiver. At the receiving end, temporal attacks such as transposition, dropping, and insertion can be detected by comparing the side information and the characteristic parameters calculated from the received video. The simulation results show that the proposed method can successfully re-synchronize the attacked video back to the original format with the accuracy from 72.41 % to 98.15 % for various video sequences based on single frame matching. After the voting process, the GOP detection accuracy can be improved to the range from 96.30 % to 100 %.

Key words: Video, Watermarking, Synchronization, Profile, Statistics

1. INTRODUCTION

Many video watermarking techniques were proposed in recent years, but most of them are still relatively weak to temporal attacks. Although video watermarking techniques that utilize the video specific characteristics are still under development, the watermark synchronization for temporal distortion is a more challenging problem [1]. However, since no video watermarking standards are determined, it is desirable to develop a re-synchronization technique that can fit most video watermarking systems. Therefore, in this work we propose a temporal re-synchronization method, which is independent of the video watermarking system.

Conventional video watermarking systems use time-invariant key schedules. Lin et al. proposed a temporal redundancy key based video watermarking synchronization system [2] [3] that uses a periodic watermark key to provide higher security. In their work, a pixel value of a certain location is chosen as the feature that is used for synchronization. They declare that if the feature is robust, their key schedule can be detected successfully. However, once the video sequence is under attacks, the feature that they proposed seems weak and may lose synchronization. Therefore, the watermark cannot be extracted successfully.

Feature extraction is a very important topic in face detection. In [4], knowledge-based methods for finding robust feature extraction were surveyed. Yang and Huang used a multi-resolution hierarchy based method for face detection [5]. The multi-resolution hierarchy of images is created by averaging and sub-sampling the image to create the center part of the face. Kotropoulos and Pitas [6] presented a rule-based localization method by calculating the horizontal and vertical projections of image defined as $HI(x)=\sum_{y=1}^{n} I(x,y)$ and $VI(y)=\sum_{x=1}^{m} I(x,y)$ for an $m \times n$ image. The local minimums of the horizontal and vertical profiles are recorded as the feature points in the input image. The 86.5 % successful detection ratio shows that the feature seems robust.

In our proposed method, the profile based feature is enhanced by including the statistics that contain the first and second order moments of the profiles in x- and y-directions, which can be used to represent the coarse scale of an image or a frame. The feature is used to re-synchronize the video back to the original temporal order. Consequently, the watermark can be successfully extracted from the synchronized frames and the keys.

The outline of the paper is as follows. The proposed profile statistics, system architecture, and temporal detection procedure are described in Section 2, 3, 4, respectively. Section 5 shows the experimental results. Finally, the conclusions are presented in Section 6.

2. PROFILE STATISTICS

In digital watermarking, the content of an image or a frame will possibly be attacked and become different from the original ones. So the feature extracted from the profile statistics of a frame should be robust enough under certain kinds of attacks.

0-7803-7882-2/03/$17.00©2003 IEEE

2.1. Profile of an image

Here is an example that shows how to get the profiles from a video frame. Fig. 1 is the first frame of the video source Akiyo. Furthermore, the pixel intensity can be represented by the amplitude axis of the 3-D meshing function, the image can be transformed from a 2-D model to a 3-D model, as shown in Fig. 2. By summing the meshing function along the x- and y-directions, we obtain the profiles of the image, shown as in Fig. 3, where the upper part represents the profile along the x-direction, while the lower part shows the profile in the y-direction.

Fig. 1 Akiyo

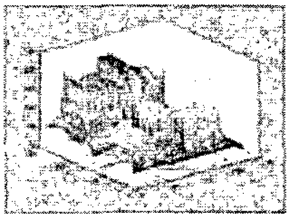


Fig. 2 3-D Meshing Function

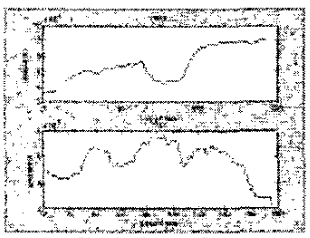


Fig. 3 The Profiles

The profiles of the image intensity $I(x,y)$ can be derived in two directions x and y as in [6]:

$HI(x)=\sum_{y=1}^{n} I(x,y)$, and $VI(y)=\sum_{x=1}^{m} I(x,y)$, respectively.

2.2. Moments of the Profiles

Since the characteristics of a frame will be sent as side information for synchronization, the number of bits representing the characteristics should be kept as few as possible. In the proposed method, the mean and variance that represent the statistics of the first and second order moments are used as follows because they are widely used in evaluating the probability characteristics [7],

$$\mu_x=\sum_{x=1}^{m} x\cdot\frac{HI(x)}{|HI(x)|},\quad \mu_y=\sum_{y=1}^{n} y\cdot\frac{VI(y)}{|VI(y)|} \tag{1}$$

$$\sigma_x^2=\sum_{x=1}^{m} x^2\cdot\frac{HI(x)}{|HI(x)|}-\mu_x^2$$
$$\sigma_y^2=\sum_{y=1}^{n} y^2\cdot\frac{VI(y)}{|VI(y)|}-\mu_y^2 \tag{2}$$

where $HI(x)$ and $VI(y)$ are the profiles, μ_x and μ_y are the means, σ_x^2 and σ_y^2 are variances along the x- and y-directions, respectively. Furthermore, the characteristic vector of the original image C and the attacked one C^a can be presented as

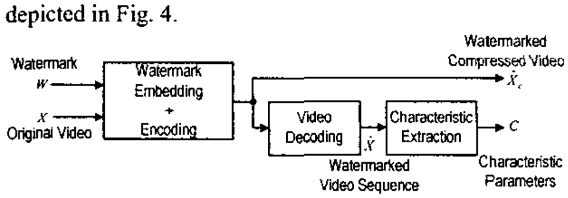
$$C=[\mu_x\ \ \mu_y\ \ \sigma_x^2\ \ \sigma_y^2]$$
$$C^a=[\hat{\mu}_x\ \ \hat{\mu}_y\ \ \hat{\sigma}_x^2\ \ \hat{\sigma}_y^2]. \tag{3}$$

In addition, the similarity, evaluated by the Normalized Correlation (N.C.) [8], can be calculated as follows.

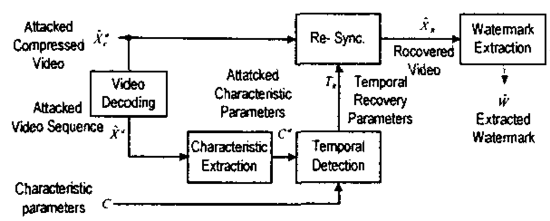
$$Z(C,C^a)=\frac{C\cdot C^a}{|C|\cdot|C^a|} \tag{4}$$

3. SYSTEM ARCHITECTURE

In this paper, we propose a temporal synchronization scheme for video watermarking system. The proposed method can be applied to most of the existing video watermarking scheme to enhance the robustness especially against the temporal attacks. The block diagram is depicted in Fig. 4.



(a) Embedding End



(b) Extracting End

Fig.4 The Proposed System Architecture

Fig. 4 (a) shows the embedding end of the proposed watermarking system. The original video sequence X is first embedded by the watermark W and compressed to produce a watermarked compressed video $\hat{X}_c$. After the video decoding process, the video can be recovered back to the watermarked sequence $\hat{X}$ in the pixel domain, which contains the spatial characteristics. Based on the analysis of profiles in the previous section, the characteristics parameters C can be extracted from each frame, and recorded as the prior knowledge [9].

Fig. 4 (b) shows the extracting end of the proposed watermarking system. The input is the attacked compressed video $\hat{X}_c^a$. The possible attacks include compression, filtering, noising attacks, especially the temporal attacks, such as frame dropping, frame insertion, transposition, etc. Because the temporal attacks may affect the video significantly, the video may not be recovered because of out-of-synchronization. Therefore, we propose a method that offers a re-synchronization technique to recover the attacked image back to the proper position in temporal axis. The attacked video sequence $\hat{X}^a$ in the pixel domain must be obtained first by video decoding. The characteristic extraction process is similar to that in the embedding end, and then characteristic parameters set C^a is extracted. By comparing the two sets of the characteristic parameters, C received from the side information and C^a computed from the received video, the temporal recovery parameters T_R can be obtained and used for re-synchronization. Finally, the re-synchronized video $\hat{X}_R$ is ready for extracting the watermark $\hat{W}$.

4. TEMPORAL DETECTION

The main purpose of the proposed method is to re-synchronize the temporal attacked video back to the proper time relation for accurate watermark extraction. The procedure using a simple set of characteristic parameters for temporal re-synchronization is detailed in this section.

The parameter C_i represents the characteristics of frame i. The original video characteristics CV and the attacked video characteristics CV^a can be formed by the set of C_i and C_j^a, respectively, as

$$\begin{aligned} CV &= \{C_i; i = 0, 1, 2, \cdots, N-1\} \\ CV^a &= \{C_j^a; j = 0, 1, 2, \cdots, M-1\} \end{aligned} \tag{5}$$

where N and M represent the number of frames of the original and attacked video sequences.

By comparing the characteristics contained in CV and CV^a, the received frames can be re-synchronized. The frame index re-synchronization is determined by the nearest neighborhood rule in which the attacked frame should be fit to the most similar frame in the original video

$$K^a = \left\{ K_j; K_j = \max_{i \in \{0, 1, \cdots, N-1\}}{}^{-1} Z(C_i, C_j^a); \text{for all j} \right\} \tag{6}$$

where K_j is the re-synchronized index for the j-th frame and K^a is the set of the index sequence of K_j.

Moreover, video compression and attacks are often processed based on the group of picture (GOP). In the case that the watermark embedding procedure uses the same key for all frames in a group of picture (GOP) and the temporal distortion is also based on GOP, the re-synchronization accuracy can be further improved by a voting process. The re-synchronization accuracy can be relaxed from a frame to a GOP. The GOP index for frame i is determined by the majority rule voting results of all frames in the same GOP. The resulting index sequence V^a represents the re-synchronized GOP indices for all frames j, $j = 0, 1, 2, \cdots, M-1$.

Finally, the temporal recovery parameter T_R

$$T_R = \begin{cases} K^a & \text{, without voting process} \\ V^a & \text{, with voting process} \end{cases} \tag{7}$$

is sent to the re-sync. block to produce the recovered video, and then the watermark can be extracted successfully.

5. EXPERIMENTAL RESULTS

In the experiments, the video watermarking system presented in [10] is used as an experimental system in the embedding and extracting ends. Our proposed re-synchronization method is used with the watermarking system improving the robustness against the temporal distortion, especially on transposition, dropping, and insertion.

In the simulation, the video sources Akiyo, Foreman, and Stefan of class A, B, and C are tested. The frame detection results without voting are shown in Table 1. The accuracy of frame detection is from 72.41 % to 98.15 %. After the voting process, the GOP detection accuracy, shown as in Table 2, is improved to the range from 96.30 % to 100 %. The overall performance in Fig. 5 shows that, after resynchronization following temporal attacks, the N.C. values of watermark detection are improved from 0 to 0.4 or even more. In other words, the un-extractable watermark in the original watermarking system becomes recognizable with the proposed method incorporated.

6 CONCLUSIONS

In our proposed method, we have developed a novel video watermarking synchronization technique based on profile statistics. The experimental results show that our technique can efficiently improve the watermark extraction results because the characteristics of profile statistics are very suitable for re-synchronization. The technique can also improve the robustness for video watermarking against temporal attacks, such as transposition, dropping, and insertion.

7. ACKNOLOWLEDGEMENT

Special thanks to C.S. Lu, J.R. Chen, H.Y. Liao, and K.C. Fan, at Institute of Information Science, Sinica, and National Central University in Taiwan to provide the watermarking programs in [10].

8. REFERENCES

[1] I.J. Cox and Matt L. Miller, "Electronic watermarking: the first 50 years," Proc. IEEE Fourth Workshop on Multimedia Signal Processing, pp. 225 -230, 2001

[2] E. Lin, C. Podilchuk, T. Talker, and E. Delp, "Streaming video and rate scalable compression: What are the challenges for watermarking? ", Proc. SPIE Security and Watermarking of Multimedia Contents III, Vol. 4314, pp. 116-127, San Jose, January, 2001

[3] E. Lin and E. Delp, "Temporal synchronization in video watermarking, " Proc. SPIE Security and Watermarking of Multimedia Contents IV, Vol. 4675, pp. 478-490, San Jose, January, 2002

[4] M.H. Yang, D.J. Krieman, and N. Ahuja, "Detection faces in images: a survey, " IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 24, No. 1, pp. 34 -58, Jan. 2002

[5] G. Yang and T.S. Huang, "Human face detection in complex background," Pattern Recognition, vol. 27, no. 1, pp. 53-63, 1994

[6] C. Kotropoulos and I. Pitas, "Rule-based face detection in frontal views, " Proc. Int'l Conf. Acoustics, Speech and Signal Processing, Vol. 4, pp. 2537-2540, 1997

[7] Henry Stark, John W. Woods, "Probability and random processes with applications to signal processing", 3rd ed., Prentice Hall, New Jersey, 2002

[8] I.J. Cox, M.L. Miller, and J.A. Bloom, "Digital watermarking", Morgan Kaufmann Publishers, 1/e, USA, 2002

[9] C.S Lu, H.Y. Liao, and M. Kutter, "Denoising and copy attacks resilient watermarking by exploiting knowledge at detector", IEEE Trans. on Image Processing, Vol. 11, No. 3, pp. 280-292, 2002.

[10] C.S. Lu, J.R. Chen, H.Y. Liao, and K.C. Fan, "Real-time MPEG2 video watermarking in the VLC domain", Proc. 16th IAPR Int. Conf. on Pattern Recognition, Quebec, Canada, Vol. II, August 11-15, 2002.

Class	Accuracy	Transposition	Dropping	Insertion
A	Akiyo	75.56	72.59	72.41
B	Foreman	98.15	93.70	95.17
C	Stefan	84.44	81.85	83.45

Table 1. Frame Synchronization without Voting

Class	Accuracy	Transposition	Dropping	Insertion
A	Akiyo	100	96.30	96.55
B	Foreman	100	96.30	96.55
C	Stefan	100	96.30	96.55

Table 2. GOP Synchronization with Voting

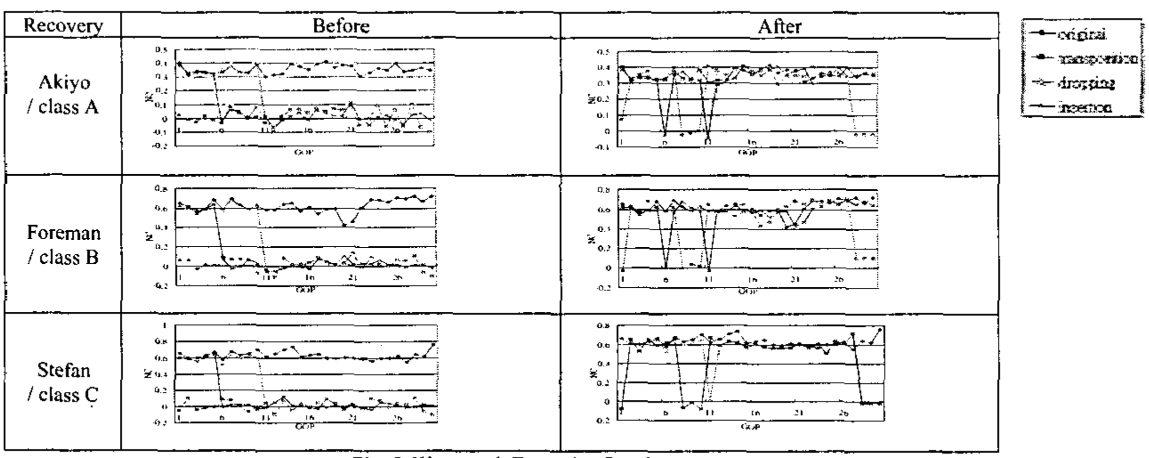


Fig. 5. Watermark Extraction Results