

Identifying Manipulations with Dual Video Watermark

S.W. MOON*, K.S. YOON*, H.D. KIM**

*ETRI(Electronics and Telecommunications Research Institute), Republic of Korea

**Department of Computer Science, KAIST(Korea Advanced Institute of Science and Technology), Republic of Korea

moonstarry@etri.re.kr, ksyoon@etri.re.kr, hdkim@mmc.kaist.ac.kr

Abstract— These days, non-professional users can manipulate digital videos easily with a video editing tool. Owing to this situation, many digital videos are distributed illegally in irregular form. In this paper, we propose a dual digital watermarking system which can estimate the distorted region of digital videos to resolve copyright infringement problems. The proposed system can identify geometrical and temporal distortions.

Keywords— Video Watermark, DCT, Forensic, Identifying Manipulation, Spread Spectrum

I. INTRODUCTION

Since video editing software has been widely propagated, non-professional users can change or alter the content of a digital video easily. For this reason, digital video content without authentication should not be used as evidence, as criminals can manipulate the content of a digital video using video editing software. If criminals conduct manipulation such as clipping, cropping, or partial distortion of the video, it may bring about a wrong judgment when the video is used as evidence. In general, fragile or semi-fragile digital video watermarking and forensic video analyses are used to authenticate digital video content and detect manipulations.

Since fragile or semi-fragile watermarks cannot be detected in a manipulated video, fragile or semi-fragile watermarking is used to detect a manipulation in a digital video content [1]-[3]. However, content authentication techniques using fragile or semi-fragile watermarks cannot usually identify either the region or type of manipulation. Furthermore, it is difficult to define whether fragile or semi-fragile watermarks were embedded in the digital video content when the video is manipulated. A forensic video analysis is a scientific technique used to analyze a video used in legal matters. In general, a forensic video analysis is a specialized tool for a specific manipulation. This means that multiple forensic video analyses should be used to identify unknown manipulations in digital video contents. In this paper, we propose a manipulation identification system using dual video watermarks, which can identify geometric and temporal manipulations in digital video contents.

To identify multiple manipulations, the proposed method uses keys based on a temporal index and geometric index to

generate semi-fragile watermarks. Unlike most content authentication systems, which use fragile or semi-fragile watermarks, this system estimates the type and region of target manipulations, while robust watermarks reveal information of the digital video content [4]-[7]. The main target manipulations of the proposed method are clipping, cropping, and partial distortion. These main target manipulations are generally used in crimes because they can modify the video content as desired. If we embed some information in a robust watermark, the proposed content authentication system can identify geometric distortions such as resizing and frame rate changes without modifying the video content. However, the system proposed in this paper does not identify global geometrical manipulations.

This paper is organized as follows. We explain the proposed watermarking scheme in section II. The processes used to identify manipulations are given in section III, and experimental results are given in section IV. Finally, some concluding remarks regarding the proposed method are given in section V.

II. PROPOSED WATERMARKING SCHEME

A. Dual Watermark Design

Figure 1 shows the embedding processes of the proposed dual watermark. Spread spectrum watermarks are embedded

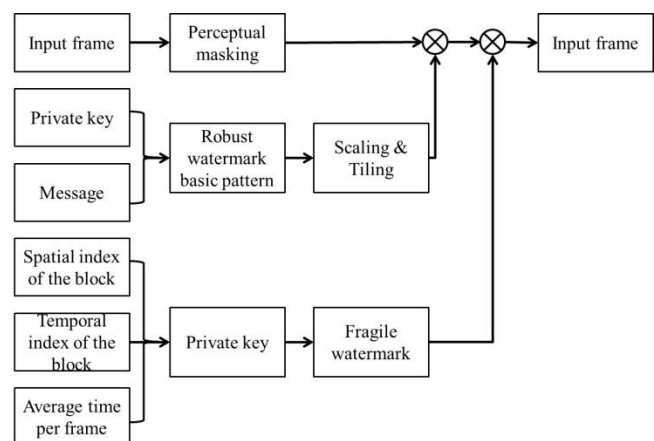


Figure 1. Embedding processes of a dual watermark

into the spatial domain of the video frames as robust watermarks. Robust watermarks are designed using the following steps [6]. First, a $m \times n$ 2-D basic pattern, W_b , is created by a random number generator with private keys. The basic pattern, W_b , is then enlarged k times to enhance the robustness against resizing. Next, constant k is calculated by dividing the height of the frame by l , which is a static integer. Finally, a robust watermark pattern, W_r , is generated by tiling the enlarged basic pattern, W_b . In this paper, we embed 2 bits into the digital video content using reversal patterns created from a robust watermark pattern. Figure 2 shows an example of reversal patterns. In practice, different robust watermark patterns are embedded into frames for each time interval T_r to embed 40 bits of information into the target videos.

Semi-fragile watermarks are embedded into the DCT domain of the video frames to identify any manipulations. Semi-fragile watermarks are designed using the following steps [8]. First, the target video frame is divided N_w times horizontally and N_h times vertically to generate block B_i . A semi-fragile watermark, W_f , which follows a Gaussian distribution with zero mean and unit variance, is embedded into the DCT domain of block B_i . To generate a semi-fragile watermark, W_f , a private key is made from three factors based on the information of target block B_i . The private key K for block B_i is computed as follows:

$$K = K_1 + 10^3 \times K_2 + K_3 \quad (1)$$

where K_1 is a spatial index of block B_i , K_2 is a temporal index of block B_i , and K_3 is the frame rate of the target video. Three factors are shown in Fig. 3. We use scaling factors to avoid the duplication of private keys.

B. Watermark Embedding

We embed a robust watermark and semi-fragile watermark into different domains to minimize interference between two watermarks. To improve the imperceptibility of robust

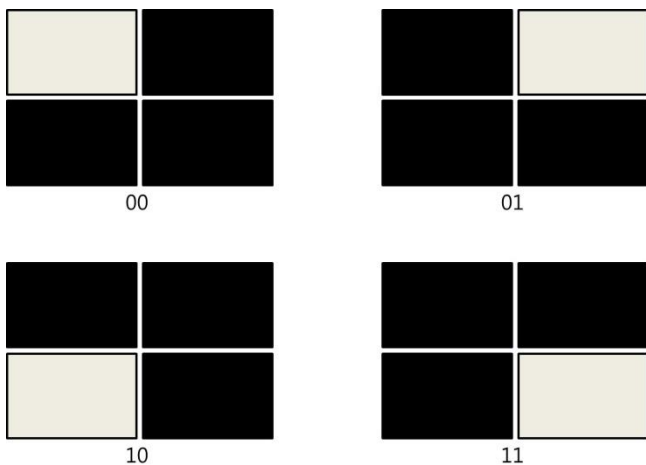


Figure 2. 2-bit expression with reversal watermark patterns

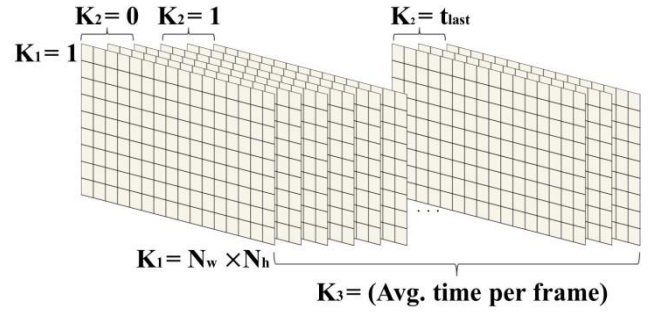


Figure 3. Three factors to generate private key

watermarks, we apply perceptual masking after embedding robust watermarks into the video frames. The perceptual masking consists of noise visibility function (NVF) masking, motion masking, and luminance masking [1], [6]. An NVF masking filter, M_n , is computed as follows:

$$M_n(i, j) = S_0 + (S_1 - S_0) \cdot NVF(i, j), \quad (2)$$

where S_0 is the maximum value of the masking filter, S_1 is the minimum value of the masking filter, and NVF is a measurement of the noise in the frame. A motion masking filter, M_m , and a luminance masking filter, M_l , uses the characteristic of the HVS, which is insensitive to noise and brightness changes when a scene has more motion or brighter or darker regions. Table 1 shows the value of masking filters M_m and M_l where $P(i, j)$ indicates the value of each pixel, and $D(i, j)$ is the difference between the past and present values of $P(i, j)$. After perceptual masking, a robust watermarked frame, X_r , can be expressed as

$$X_r = X + (W_r \cdot M_n \cdot M_m \cdot M_l), \quad (3)$$

where X is the original target the video frame. Each dot product indicates a pixel-wise product.

After perceptual masking, we embed semi-fragile watermarks into the DCT domain of the video frames. First, semi-fragile watermarks, W_f , are divided into sub-watermarks, W_s . We then apply a $b \times b$ DCT to each block B_i and reorder the DCT coefficients using a zigzag scan to embed sub-watermarks W_s into DCT coefficients. We embed semi-fragile watermarks into mid-range coefficients of the DCT block to minimize the inference from noise and to maximize the fidelity of the watermarked frames. Watermarked coefficient C_i^w can be expressed as follows:

$$C_i^w = C_i + \alpha x_s |C_i| \quad (4)$$

where C_i is the target DCT coefficient, α is the scaling factor of a semi-fragile watermark, and x_s is the value of sub-watermark, W_s .

TABLE 1. Value of the motion masking filter $M_m(\text{top})$ and luminance masking filter $M_l(\text{bottom})$

Value of $M_m(i, j)$	Value of the difference $D(i, j)$
1.0	$D(i, j) < 20$
1.1	$20 \leq D(i, j) < 50$
1.2	$50 \leq D(i, j) < 150$
1.5	$150 \leq D(i, j)$
Value of $M_l(i, j)$	Value of the pixel $P(i, j)$
3.0	$P(i, j) < 25$
1.5	$25 \leq P(i, j) < 50$
1.0	$50 \leq P(i, j) < 200$
1.3	$200 \leq P(i, j) < 225$
1.5	$225 \leq P(i, j)$

III. MANIPULATION IDENTIFICATION

A. Watermark Detection

We estimate robust watermarks embedded into target video as noise using an adaptive Wiener filter as a noise reduction filter. A denoised frame, X_d , with the adaptive Wiener filter can be expressed as follows:

$$X_d(i, j) = \mu(i, j) + \frac{\sigma^2(i, j) - s^2}{\sigma^2(i, j)} \cdot (X(i, j) - \mu(i, j)) \quad (5)$$

where $\mu(i, j)$ is the local mean of X , $\sigma(i, j)$ is the local variance of X , and s^2 is the approximated mean of the local variance. After denoising, an estimated watermark, W_e , can be extracted by

$$W_e = X(i, j) - X_d(i, j) \quad (6)$$

An estimated robust watermark pattern, W_e , is accumulated for t_r seconds and normalized. We fold the normalized estimated watermark pattern W_e to strengthen the tendency of the robust watermark. To extract a message from a robust watermark, we use a normalized cross correlation (NCC). The normalized cross correlation, Z , of the estimated robust watermarks is computed as follows:

$$Z = \frac{IFFT(FFT(W_e) \cdot FFT(W_r)^*)}{|W_e| \cdot |W_r|} \quad (7)$$

where W_r is the reference watermark pattern. The threshold T_r , which is compared with Z is defined as follows [1]:

$$T_r = \mu_z + \alpha_z \sigma_z \quad (8)$$

where μ_z is the mean of Z , σ_z is the standard deviation of Z , and α_z is the controlling factor of a false positive error.

To detect semi-fragile watermarks, we generate a reference semi-fragile watermark, W_f , from the target frame using the same watermark embedding process. We then divide the target frame into block B_i and extract mid-range $b \times b$ DCT coefficients from B_i . We construct a vector $C =$

$\{C_1^w, C_2^w, C_3^w, \dots, C_N^w\}$, where C_i^w is the mid-range $b \times b$ DCT coefficients from B_i . The correlation between the reference semi-fragile watermark, W_f , and vector C can be computed as follows [8]:

$$\text{Corr}(W_f, C) = \frac{W_f \cdot C}{N} \quad (9)$$

where N is the number of blocks, B_i , in a target frame. We use threshold T_f to decide whether the semi-fragile watermark is detected. A generally used threshold, T_f , is given by [8]:

$$T_f = \frac{\alpha_f}{3N} \sum_{i=1}^N |C_i^w| \quad (10)$$

where α_f is a controlling factor of false positive errors. If more than P_f percentages of blocks in a target frame have been manipulated, we can determine that the target frame has been manipulated.

B. Content Authentication

The proposed method authenticates digital video content using the following steps. First, we bind adjacent frames into a frame group. We then detect semi-fragile watermarks from each frame group, and if more than P_{global} percent of frames in the same frame group are manipulated, we can determine that the frame group has been manipulated. If more than $P_{partial}$ percentages of blocks have the same spatial index in the same frame group, we can determine that the region has been manipulated. If every frame group and region is determined as having not been manipulated, we can determine that the target video is pure.

The proposed method identifies manipulations through the following processes. If every frame group is manipulated, we can determine that the target video has been geometrically manipulated or the frame rate of the target video has been changed. If the target video is clipped, semi-fragile watermarks are not detected from the frame group that is clipped or from the following frame groups. We use the same key for the end of a video to identify whether the first frame group of the video has been clipped. If a criminal distorts a certain region of a target video, semi-fragile watermarks cannot be detected from the manipulated blocks.

IV. EXPERIMENTAL RESULTS

Twenty video clips were used to evaluate the proposed system. A 40-bit message was embedded with a robust watermark. We used ten SD videos, five HD videos, and five Full-HD videos, and each video was divided 16 times vertically and 9 times horizontally to generate the blocks. The parameters used for the test were set to $L = 10$, $N = 1$, $\alpha = 0.2$, $\alpha_f = 0.2$, $P_f = 50$, $P_{global} = 90$, and $P_{partial} = 100$.

We used three types of manipulations to test the proposed method. The type 1 manipulation is rescaling and a frame rate change, the type 2 manipulation is clipping, and the type 3 manipulation is partial distortion. We clipped less than 10

TABLE 2. Comparison between the proposed method and a method by Kim in terms of the manipulation identification rate and PSNR

	Proposed Method	Kim's Method[1]
Type 1	100%(20/20)	100%(20/20)
Type 2	100%(20/20)	100%(20/20)
Type 3	100%(20/20)	75%(15/20)
PSNR(dB)	44.29	43.12

percent of a video for a type 2 manipulation, and distorted less than 5 percent of the video for type 3 manipulations. We compared the proposed dual watermark scheme with Kim's hybrid watermarking scheme [1], which was used to embed two spatial watermarks.

Table 2 shows that proposed method has a higher fidelity than Kim's method since we embedded a semi-fragile watermark into the mid-range of the DCT domain. Kim's method [1] cannot detect type 3 manipulations for 5 five of the target videos. The proposed method, on the other hand, identified every type of manipulation successfully. Table 3 shows that the manipulation identification rate of the proposed method is over 95 percent for a target frame.

V. CONCLUSION

In this paper, we proposed a manipulation identification system with a dual watermark scheme for content authentication. To use digital video content as evidence in a court case, the integrity of the content should be guaranteed. The proposed method can identify the type of manipulation and measure a manipulated region in an altered video. Our experimental results show that proposed method guarantees better fidelity and accuracy than an existing method.

TABLE 3. Manipulation identification rate of the proposed method for a target frame

	Identification rate for a frame
Type 1	95.57%
Type 2	95.15%
Type 3	95.33%

In future research, we will try to identify complex manipulations in a video using a video forensic technique and video forensic mark. To guarantee better fidelity, we will embed watermarks in various domains.

ACKNOWLEDGMENT

This work was supported by MCST(Ministry of Culture, Sports & Tourism) (10041539 High Compression, Low Loss Content Creation/Distribution /Display Technology Development for 8K-Video Service).

REFERENCES

- [1] H.D. Kim, T.W. Oh, J.W. Lee, *et al.* "A Hybrid Watermarking Scheme for CCL-applied Video Contents," in *Proc. EUVIP*, pp. 194-204, 2011.
- [2] J. Fridrich, "Security of Fragile Authentication Watermarks with Localization," in *Proc. SPIE*, vol. 4675, pp. 691-700, 2002.
- [3] C.C. Wang and Y.C. Hsu, "Fragile Watermarking for H.264 Video Stream Authentication," *Intelligent Systems Design and Application*, vol. 1, pp. 77-80, 2008.
- [4] I. J. Cox, J. Kilian, F. T. Leighton, *et al.*, "Secure Spread Spectrum Watermarking for Multimedia," *IEEE Trans. on Image Processing*, vol. 6, no. 12, pp. 1673-1687, 1997.
- [5] M. Barni, F. Bartolini, V. Cappellini, *et al.*, "A DCT-domain System for Robust Image Watermarking," *Signal Processing*, vol. 66, pp. 357-372, 1998.
- [6] K.S. Kim, H.Y. Lee, D.H. Im, *et al.*, "Practical, Real-time and Robust Watermarking on the Spatial Domain for High-definition Video Contents," *IEICE Trans. on Inf. Syst.*, vol. E91-D, pp. 1359-1368, 2008.
- [7] S.H. Joo, Y.H. Suh, J.H. Shin, *et al.*, "A New Robust Watermark Embedding into Wavelet DC Components," *ETRI Journal*, vol. 24, no. 5, pp. 401-404, Oct. 2002.
- [8] K. Al-Dhmani and M. E. Al-Mualla, "A Block-based Barni Watermarking Method for Improved Robustness under Video Compression Attacks," *Proc. GCC*, pp. 1-5, 2006.