Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.12785/amis/070632

Video Forensic Marking Algorithm using Peak Position Modulation

Jihah Nah[1], *Jongweon Kim*[2,*] *and Jaeseok Kim*[1]

[1] Department of Electrical and Electronic Engineering, Yonsei University, Seoul, Korea
[2] Department of Intellectual Property, Sangmyung University, Seoul, Korea

Received: 24 Mar. 2013, Revised: 26 Jul. 2013, Accepted: 28 Jul. 2013
Published online: 1 Nov. 2013

Abstract: Forensic marking, a type of watermarking, is an efficient means of detecting copyright violation and tracing unauthorized access to digital content. In this paper, a new video forensic marking algorithm is proposed which uses peak position modulation (PPM) and can embed sufficient information to trace an illegal distributor. PPM uses the positions of correlation peaks, which are modulated by user information. Furthermore, a forensic mark is embedded into a 2-level 3D discrete wavelet transform (DWT) domain. Experimental results show the robustness against various attack and the capacity improvement of proposed algorithm.

Keywords: Video Forensic Marking, 3D Discrete Wavelet Transform, Peak Position Modulation, Spread Spectrum.

1 Introduction

As the manipulation of digital content becomes easier, copyright infringement becomes a serious problem. It is important to detect copyright violation and control unauthorized access to digital content. Digital watermarking, which embeds copyright information into digital content, is widely used for authentication and copyright protection of digital content [1,2,3,4,5].

Forensic marking is a type of watermarking technology in which a unique mark is embedded in digital materials. If content is distributed illegally, the unauthorized user can be traced using these forensic marks. The advantage of forensic marking is that it can hide content user information, and it can trace a user who has illegally copied and distributed the content.

Many research groups are studying forensic marking of digital images and audio; they have recently extended their research to include digital video content [6,7,8,9]. On the basis of the spread spectrum, an existing video forensic mark is embedded into the motion vector information in the MPEG domain or the discrete cosine transform coefficients. The former exhibits weak re-compression, and the latter yields worse quality at high embedding intensity. Also, the forensic mark capacity is limited because the compressed video stream has little redundancy, and the spread spectrum modulating mark uses each frame for 1 bit. Video forensic marking must ensure sufficient payload and robustness against compression, time editing, and format changes [3,4,5].

In this paper, we propose a video forensic marking algorithm that can embed sufficient information to trace an illegal distributor. The forensic mark is embedded in the 3D discrete wavelet transform (DWT) domain for robustness, and it is generated by peak position modulation (PPM) to ensure adequate information capacity and error tolerance. This guarantees robustness, transparency, and sufficient payload for the video.

2 Proposed Forensic Marking Scheme

Fig. 1 shows a block diagram of the proposed algorithm. The core elements are frame division, 3D DWT, and PPM.

2.1 Forensic Mark Design

The embedding is processed in the 3D DWT domain. The forensic mark is generated using the spread spectrum in order to ensure robustness against signal degradation, compression, and other attacks. The forensic mark is generated using the peak position of the spread spectrum correlation, which is modulated by the user information.

* Corresponding author e-mail: jwkim@smu.ac.kr

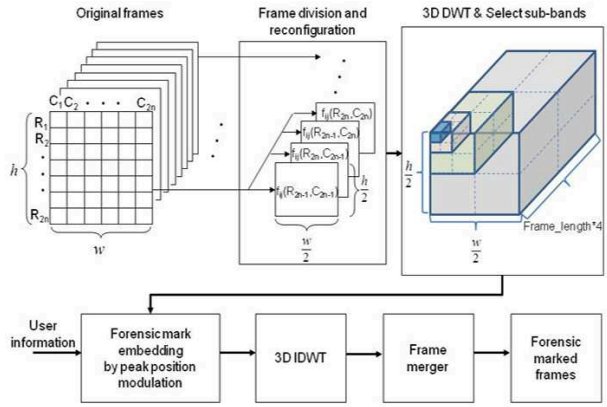

Fig. 1: Overview of the proposed forensic mark scheme

Frame division and alignment

To expand the space available for embedding a forensic mark, we divided each frame into four sub-frames by rearrangement of the odd/even rows and columns.

$f_{i,j}$ (i: frame number, $j < 4$: sub-frame number)

A new frame set is generated by equation (1).

$$\begin{aligned} S_i &= f_{i,j}(R_{2n-1}, C_{2n-1}), \\ S_{i+1} &= f_{i,j}(R_{2n}, C_{2n-1}), \\ S_{i+2} &= f_{i,j}(R_{2n-1}, C_{2n}), \\ S_{i+3} &= f_{i,j}(R_{2n}, C_{2n}) \end{aligned} \tag{1}$$

R is a row, C is a column, and S is the divided sub-frame. The generated sub-frames are reorganized into a video having four times the length and half the height and width. N original frames are divided into $4N$ sub-frames by rearranging the odd/even rows and columns.

3D discrete wavelet transform

Once the video frame is lengthened by frame division and alignment, the next step is to transform the video sub-frame group by a 3D DWT, which is an iterative procedure of a 1D DWT [10].

$$f(t) = \sum_{j \in Z} C_{0,j} \Phi_{0,t}(t) + \sum_{k \geq 0} \sum_{j \in Z} d_{k,j} \Psi_{k,j}(t) \tag{2}$$

Equation (2) is the DWT, where $\sum_{j \in Z} C_{0,j} \Phi_{0,t}(t)$ is the approximation of $f(t)$, and $\sum_{k \geq 0} \sum_{j \in Z} d_{k,j} \Psi_{k,j}(t)$ are the details of $f(t)$. The 1D DWT is expanded to the 3D DWT by using a product of 1D scaling functions.

$$\Phi(x,y,z) = \Phi(x) \cdot \Phi(y) \cdot \Phi(z) \tag{3}$$

The 3D DWT generates seven details, which we can use to embed the forensic mark. If n unit frames are available for embedding the forensic mark, we can embed user information into the seven sub-bands for each n-frame set by a 3D DWT.

In this paper, a 2-level 3D DWT was used to ensure transparency. Additionally, each frame was divided into four sub-frames to guarantee suitable payloads. We selected three sub-bands, {1,2}{1,1}, {1,2}{1,3}, and {1,2}{1,5}, from the seven details because these sub-bands belonged to the medium-frequency range. We could then embed the forensic mark in these sub-bands.

Peak position modulation

In terms of detecting a forensic mark, video is considered noise. To overcome this problem, the embedded forensic mark must be suitable in either intensity or length. However, the quality of the content can be degraded if the forensic mark is made stronger, and payloads can be decreased if the forensic mark is made longer.

Fig. 2 depicts the relationship between the sequence length and the correlation peak.

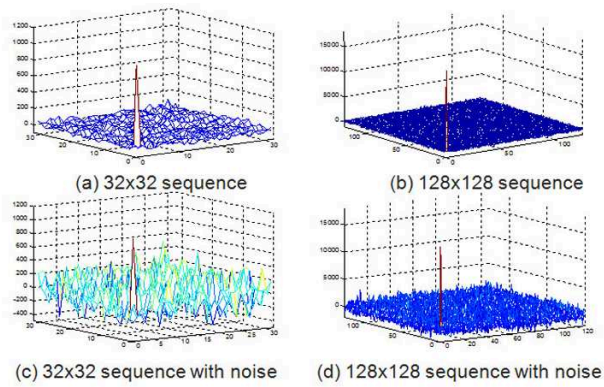

Fig. 2: Correlation peaks by sequence length

Fig. 2(a) shows the correlation peak when the sequence length is 32×32 and (b) when it is 128×128. In both (a) and (b), the correlation peaks are easily distinguished because they are higher than the other values. In both (c) and (d), the noise is three times stronger than in the sequences in (a) and (b). The correlation peak in (c) is hard to discriminate from the other peaks because the noise creates many peaks. However, the correlation peak in (d) stands out from these peaks. Therefore, we can conclude that the quality of digital content and the detection performance of embedded information can be ensured with the use of long sequences as forensic marks.

We introduce PPM to ensure appropriate payloads and robustness. We can generally identify the existence of a spread spectrum sequence using correlation. PPM uses the position of a correlation peak as embedded information. The basic equation is given by equation (4).

$$R_{sw}(u,v) = \int\int I_w(x,y)w(x+u,y+v)dxdy \qquad (4)$$

$I_w(x,y)$ is a marked image, and w is a forensic mark generated by a pseudo random number (PRN) sequence. The variables u and v represent the amount of shift in the 2D coordinate system. The peak position denoted by equation (5) represents the embedded information.

$$\begin{aligned} p &= u+v\times width \\ &= B_n 2^n + B_{n-1}2^{n-1} + \cdots + B_1 2^1 + B_0 \end{aligned} \qquad (5)$$

B_n represents the n^{th} bit.

The message is modulated as the peak position using equation (4) and (5). For example, if the message ”A” is embedded, ”A” is changed into the ASCII code value, 41H (=65D) and then the PRN sequence is shifted by the value 65. This method guarantees sufficient payload, transparency, and robustness, all benefits of a long sequence.

Fig. 3(a) shows the correlation peak of the reference sequence and (b) shows the correlation peak of the reference sequence and the PRN sequence shifted by the ASCII value of ”A”. The correlation peak is shifted 65 from origin. 128×128 sequences can represent peak position values from 0 to 16383 $(2^{14}-1)$. In other words, if we use PPM, we can ensure a 14-bit payload for 128×128 sequences.

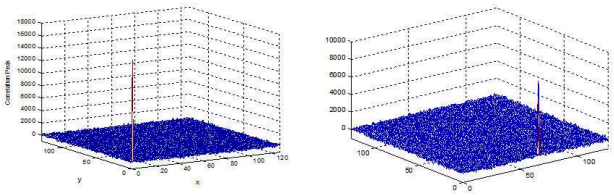

Fig. 3: Peak position modulation

2.2 Forensic Mark Embedding

Forensic mark information can usually include a user ID, Social Security number, time information, and other data. If a user distributes the content illegally, it is possible to detect the user’s information in the content and determine his or her identity.

The message extracted from the user information modulates the PN sequence arrangement. After this occurs, the forensic mark is embedded in the wavelet-transformed frames. The procedure is as follows:

–a) N original frames are divided into 4N sub-frames through a rearrangement of odd/even rows and columns.

–b) If n unit frames are available for embedding the forensic mark, we can embed user information into the seven sub-bands for each n-frame set by a 2-level 3D DWT. We can embed $q\times n$ bits of information per unit frame number per sub-band according to the expression of q bits of information per sub-frame.

–c) The forensic mark has a value of [-1, 1] and a uniform distribution, and the initial arrangement is a reference forensic mark.

–d) The PPM version of the reference forensic mark is generated by dividing the information into q bits.

Using PPM with 1×1 pixels as a unit sample space is not recommended for obtaining a robust forensic mark. It is apt to cause errors due to geometrical modification or asynchronous problems due to changes in frame size or compression format. Therefore, sufficient space in the peak position should be taken to prevent these errors. For this reason, the embedded information is reduced by as much as $(l+m)$ when the unit sample space is increased to $2^l\times2^m$. The error range of the rotation can also be reduced, as described in equation (6).

$$\theta = cos^{-1}\left(\frac{(n\times n)/(2^l\times2^m)}{\sqrt{((n\times n)/(2^l\times2^m))^2+1)}}\right) \qquad (6)$$

That is, even though we can embed 10 bits of information in a 44×36 sub-frame after a 2-level 3D DWT for a 352×288 frame, we can embed only 6 bits of information when we use the 4×4 unit sample space. The total information per sub-frame is 7 bits when the +/-sign bit is added. If we use the 4×4 unit sample space for a 44×36 sub-frame, a rotation range as great as one 4×4 unit sample space can represent a tolerable error range. If we enlarge the unit sample space to greater than 4×4, the rotation error tolerance is improved, but the capacity for embedding information decreases.

3 Performance Evaluation

The proposed algorithm was evaluated using six videos and 16 digits displaying the user ID. The original video was 352×288 in size with 96 frames. Seven detailed sub-bands were obtained by the 2-level 3D DWT. We used three sub-bands for greater robustness against noise.

Each was 44×36 in size and had eight sub-frames. The unit sample spaces were configured as 4×4 arrays. We can express 198 peak positions ($11 \times 9 \times 2$ [+/- sign bit]), and we embedded 56 bits of information extracted from 16 digits using PPM. If we use seven detailed sub-bands, 392 bits of information can be embedded in every eight sub-frames. Although it is difficult to compare the proposed algorithm directly with that of Kim et al. [11], our algorithm, which embedded 392 bits in a 352×288 video during a 0.54-s interval, performed 64 times better than when 192 bits were embedded in a 1920×1080 HD video in a 17-s interval.

Fig. 4 ∽ Fig. 8 show the experimental results. Fig. 4 shows the average peak signal-to-noise ratio (PSNR) and Fig. 5 shows the average bit error rate (BER) of forensic marked videos with embedding intensities of 5, 7, 9, and 11; one video was uncompressed, and various codecs were applied to the others (MPEG-4 = 669 kbps, MSVC = 26.9 Mbps, and CIVD = 7.2 Mbps).

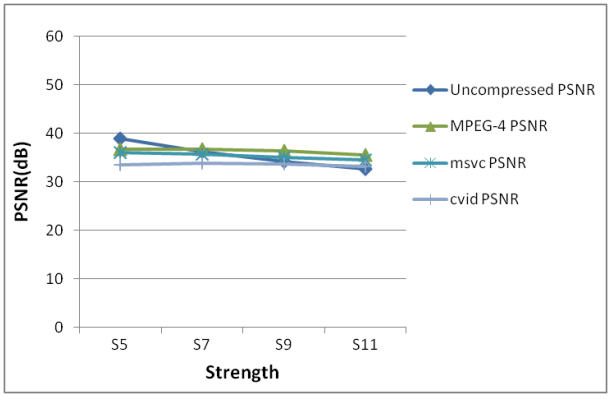

Fig. 4: Average PSNR under compression attacks

The quality of the marked video decreased with increasing embedding strength. MPEG-4, which has a high compression rate, required a higher embedding strength of S11; its average PSNR was 35.6 dB, and the average BER was 0. In contrast, the BERs for the other codecs were 0% when the embedding strength was greater than or equal to S7, because their compression ratios are lower than that of MPEG-4. The experimental result for the MPEG-4 codec was compared with that of Chouse and Tokur [12]. When the PSNR was about 36 dB, the proposed algorithm's BER was 0%, but for Chouse and Tokur's algorithm, the BER was 18.8% when the PSNR was 36.9 dB. In some cases, the PSNR of compressed video was better than that of uncompressed video. This is because of a synergistic effect on the PSNR upon removal of the forensic mark. The reason is that high-frequency components of the forensic mark are

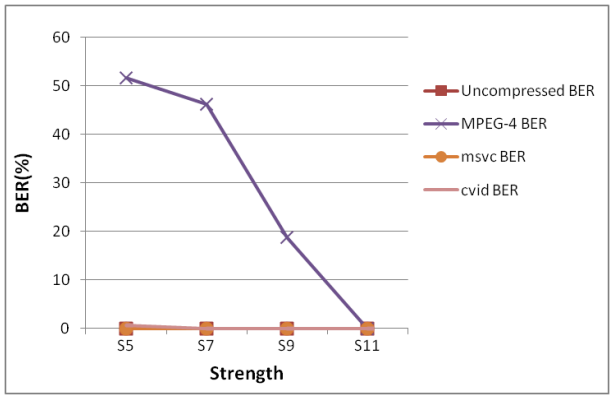

Fig. 5: Average BER under compression attacks

removed, and low- and medium-frequency components affect the video quality.

Fig. 6 shows the PSNR and BER when an additive white Gaussian noise (AWGN) was applied. The BER is 0% for PSNRs of 15 ∽ 40 dB, which indicates that the algorithm is very robust to noise attack. Actually, at a PSNR of 15 dB, the quality of the video is too poor to allow content identification. However, even at a PSNR of 10 dB, the BER is just 4.17%.

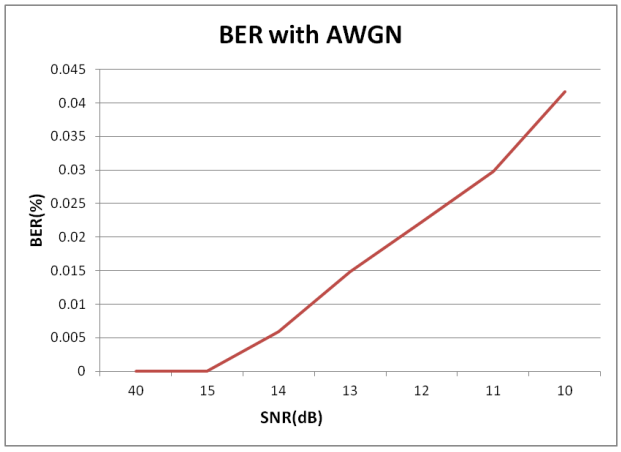

Fig. 6: BER under Gaussian noise attack

Frame dropping is often common when the video size is reduced for hand-held devices. Fig. 7 shows the average BER when video frames are dropped. The frame length of the sample videos is 96, so we examined 1, 2, 4, 8, 16, 32, and 48 dropped frames. The BER is 5.8% and 17% when 16 and 48 frames are dropped, respectively. The results

demonstrate that the algorithm is also robust against frame dropping.

Fig. 8 shows the BER for various frames other than 352×288. When the frame size is expanded, no error occurs. When the frame size is reduced to 320×240, the BER is also 0%, but the average BER is about 45.7% when the size is downscaled to 176×144. When the scale is less than 1/2 of the original frame size, the BER is about 22.3%.

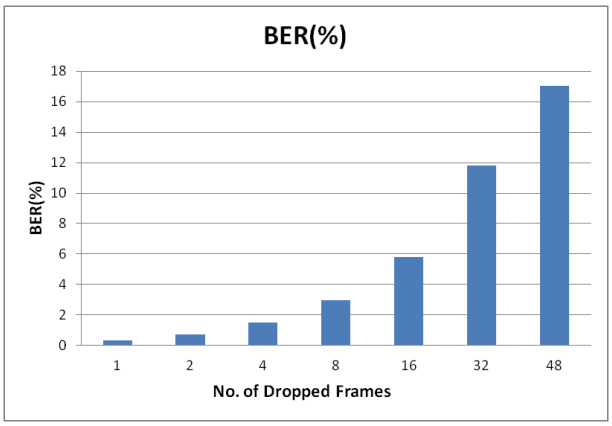

Fig. 7: BER versus number of dropped frames

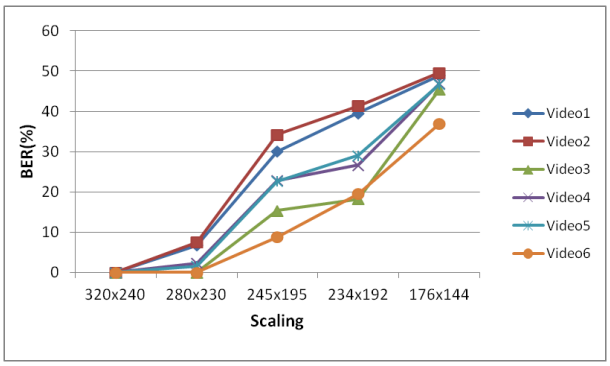

Fig. 8: BER versus frame size

4 Conclusions

In this paper, we proposed a new video forensic marking algorithm to hide specific information related to content users. To evaluate the algorithm's performance, we tested six videos 352×288 in size with embedded information. We detected 56 bits of information without error in the uncompressed mode. We also conducted four experiments focused on compression, additive Gaussian noise, frame dropping, and scaling. The results demonstrated that the proposed algorithm outperforms the algorithms suggested by Kim et al [11]. and Chouse and Tokur [12]. In particular, this algorithm shows robustness to noise attack, frame dropping and scaling.

An advantage of the proposed algorithm is that there are only $1 \backsim 2$ bits errors because PPM causes decimal errors in the range of $1 \backsim 2$. Therefore, error correction is very simple when this algorithm is used. This algorithm enables the embedding of Social Security numbers, since we can embed 56 bits information into eight unit sub-frames. Therefore, we can embed user information and then trace illegal distributors. In this paper, we embedded only 168 bits of information into eight unit sub-frames in three selected sub-bands; however, we can extend the maximum payload to $(\lfloor 7 \times 8 \rfloor \times 7$ sub-bands$) \times 12 = 4,704$ bits for $(352 \times 288) \times 96$ frames by using all seven embedding spaces with their corresponding eight sub-frames in the 3D DWT domain.

Acknowledgement

This research project was supported by the Ministry of Culture, Sports and Tourism (MCST) and the Korea Copyright Commission in 2011.

References

[1] Husrev T. Sencar, M. Ramkumar and Ali N. Akansu. Data Hiding Fundamentals and Applications: Content Security in Digital Multimedia. Elsevier, (2009).

[2] K. J. Ray Liu, Wade Trappe, Z. Jane Wang, Min Wu, and Hong Zhao. Multimedia Fingerprinting Forensics for Traitor Tracing. EURASIP Book Series on Signal Processing and Communications, **4**, (2005).

[3] I. J. Cox, J. Kilian, T. Leighton and T. Shamoon. Secure Spread Spectrum Watermarking for Multimedia. IEEE Transactions on Image Processing, **6**, 1673-1687 (1997).

[4] C. Busch, W. Funk, and S. Wolthusen. Digital watermarking: From concepts to real-time video applications. IEEE Transactions on Computer Graphics and Applications, **19**, 25-35 (1999).

[5] C. I. Podilchuk and E. J. Delp. Digital watermarking: Algorithms and applications. IEEE Signal Process. Magazine, **18**, 33-46 (2001).

[6] Natarajan Meghanathan and Lopamudra Nayak. Steganalysis Algorithms for detecting the hidden information in image, audio and video cover media. Int. Journal of Network Security & Its Application(IJNSA), **2**, (2010).

[7] H. Y. Huang, C. H. Fan, and W. H. Hsu. An effective watermark embedding algorithm for high JPEG compression. Proceeding of 8th IEEE Int. Conf. Computer and Information Technology, 256-259 (2008).

[8] S. T. Chen, H. N. Huang, C. J. Chen and G. D. Wu. Energy-proportion based scheme for audio watermarking. IET Signal Processing, **4**, 576-587 (2010).

[9] C. H. Yang, H. Y. Huang, and W. H. Hsu. An adaptive video watermarking technique based on DCT domain. Proceeding of 8th IEEE Int. Conf. Computer and Information Technology, 589-594 (2008).
[10] Kwang-il Kim, Cui Jizhe, Jong-Weon Kim and Jong-Uk Choi. Digital Video Watermarking Using Frame Division and 3D Wavelet Transform. KIISC, **18**, 155-162 (2008).
[11] Kyung-Su Kim, Dong-Hyuck Im, Young-Ho Suh and Heung-Kyu Lee. A Practical Real-Time Video Watermarking Scheme Robust Against Downscaling Attack. LNCS, **5041**, 323-334 (2008).
[12] Chasan Chouse and Yüksel Tokur. A Wavelet-Based Video Watermarking. Eleco, **2**, 180-184 (2005).

Jihah Nah received the MS degree in electronics engineering from University of Seoul and Ph.D. degree in Electrical and Electronic Engineering from Yonsei University, Korea. From 1993 to 2003, she was a member of the research staff of the ATM Switching Lab at ETRI. She was manager of the Embedded Software Team at KIPA from 2003 to 2008. Her research interests are in the areas of copyright protection, image processing and embedded software development.

Jongweon Kim received the Ph.D. degree from University of Seoul, major in signal processing in 1995. He is currently a professor in Department of Intellectual Property at Sangmyung University in Korea. He has considerable practical experience in digital signal processing and copyright protection technology in the institutional, the industrial, and academic environments. His research interests are in the areas of copyright protection technology, digital rights management, digital watermarking, and digital forensic marking.

Jaeseok Kim received a Ph.D. degree in electronic engineering from RPI, NY, USA in 1988. From 1988 to 1993, he was a member of the technical staff at AT&T Bell Labs, USA. He was director of the VLSI Architecture Design Lab of ETRI from 1993 to 1996. He is currently a professor in the electrical and electronic engineering department at Yonsei University, Seoul, Korea. His current research interests include communication IC design, high performance digital signal processor VLSI design, CAD S/W, and implementation of H.264/AVC codec.