COMBINED FRAGILE WATERMARK AND DIGITAL SIGNATURE FOR H.264/AVC VIDEO AUTHENTICATION

K. Ait Saadi[1]*, A. Bouridane*[2]*, A. Guessoum*[3]

[1]Centre de Développement des Technologies Avancées, Division Architecture des Systèmes, Cité 20 Août 1956, BP 17,
Baba Hassen, 16303, Alger, Algérie.
Phone: +213 21 35 10 18/40, Fax +213 21 35 10 39, Email : ait_saadi@yahoo.com
[2]School of Electronics, Electrical Engineering and Computer Science, Institute of Electronics, Communications and Information Technology, Queen's University Belfast, United Kingdom,
Phone: +44 (0)28 9097 4639 , Fax: +44 (0)28 9097 4879, Email: A.Bouridane@qub.ac.uk
[3]Université Saad Dahlab, Institut d'Electronique, Route de Soumaa BP 270, Blida.

ABSTRACT

Currently the demand of security is getting higher due to easy reproduction of digitally created multimedia data. Digital watermarking aims to embed secret information into content for copyright protection and authentication. This paper proposes a combined digital watermarking and digital signature based on video content to authenticate and to verify the integrity of the compressed H.264/AVC video. Features are extracted as the authentication data from transform domain to generate a unique digital signature using an MD5hash function. The authentication information treated as fragile watermark is embedded in a set of motion vectors belonging to higher motion activities with the best partition mode in a tree-structured motion compensation approach. Watermark detection is blind and does not require an original video. Experimental results show that the proposed watermarking technique can not only achieve the goal of fragility but also verify the authenticity and the integrity of the video.

1. INTRODUCTION

With the recent development of digital technology, various digital contents are widely generated, distributed and stored electronically. However, rapid development of digital contents and infrastructures also raised many problems such as the protection of copyrights or authentication of contents. The importance of copyright protection or authentication of digital contents has increased due to the characteristics of digital contents such that it could be copied easily and the copy is identical to the original [1]. Therefore, technologies for copyright protection or authentication are essential. Digital watermarking and digital signature are two techniques used to address this issue.

Digital watermark techniques embed an invisible signal (for example, company logo or personal symbol) into video so as to attest the owner identification of the media and discourage the unauthorized copying. While watermark techniques emphasize protecting the right of service providers, digital signature focuses on that of the customers. For example, a video purchaser may want to know whether the product he or she bought is from the legal seller and is the authentic one. Digital signature scheme can be used to solve this problem. First the video seller extracts some information dependent on the content of the original video and encrypts it into a small-size file, which is called signature. Then the signature file is sent to the purchaser with the original video [2]. An obvious drawback of these schemes is the extra bandwidth needed for transmission of the signature. Because most digital applications such as Internet multimedia, wireless video, personal video recorders, video-on-demand, videophone and videoconference have a demand for much higher compression to meet bandwidth criteria and best video quality as possible, different video codecs have evolved to meet the current requirements of video application based products. Among various available standards, H.264/AVC Advanced Video Codec is becoming an important alternative providing reduced bandwidth, better image quality in terms of peak-signal-to-noise-ratio (PSNR) and network friendliness [3], but it requires higher computational complexity.

A large number of watermarking schemes have been proposed for copyright protection and authentication for current popular standards such as MPEG-1 and MPEG-2, but only a few for the latest video coding standard H.264/AVC. In addition, as many new features are introduced to H.264, a large number of previous video watermarking algorithms cannot be applied directly, so development of new algorithms is required to address this new standard.

The state-of-the-art watermarking research and technology to authenticate the H.264/AVC video falls into two broad classes: digital watermarking and digital signature.

Digital watermarking directly embeds some information into video. Some of the published H.264/AVC video authentication papers have concentrated on embedding a watermark directly in the compressed domain [4] [5]. In a few others, the embedding process is carried out in the compressed bitstream delivered by the H.264/AVC encoder [6] [7] [8]. Up till now, most of the compressed-domain (during encoding) video authentication systems for H.264/AVC takes into account the temporal dimension of the video and rely on marking the motion vector. In [5], the authors proposed a hard authentication algorithm to authenticate the H.264/AVC video based on the accurate usage of the tree-structured motion compensation, motion estimation and Lagrangian optimisation for mode decision of the H.264/AVC. The algorithm performed well in terms of sensitivity against transcoding and common signal processing but lacked the ability to provide further information necessary to characterise the attack.

Digital signature is a conventional scheme used in [9] to authenticate the H.264/AVC. The digital signature is embedded as Supplemental Enhancement Information (SEI) in the H.264/AVC bitstream. The drawback of their scheme is the increase in the bits transmitted by the encoder, so the extra bandwidth needed for transmission of video.

To address the problem of the extra bandwidth needed for transmitting the signature, a combined digital watermark and digital signature for compressed H.264/AVC video authentication and content integrity verification is proposed in this paper. The main idea is motivated by the results obtained in previous works [5] and [9]. The digital signature treated as a fragile watermark is generated from the video contents and then inserted into H.264/AVC stream during encoding process. The watermark is embedded by selecting suitable motion vectors (MVs) which are associated with higher motion activities within P-frames by forcing their Least Significant Bits (LSB) to match the corresponding watermark bits. To authenticate and verify the received compressed video, the receiver performs the same operations as applied on the embedding side in a reversed order to extract the embedded watermark and compares it with the signature generated in the decoder in the same manner as that employed by the embedder. If the signature and the extracted watermark match, the received video is considered to be authentic. Otherwise, the embedded watermark will degrade the original video, which makes the signature extracted from the watermarked video different from the original one. Therefore, a robust feature extraction to generate the signature is of great importance.

The rest of the paper is organized as follows. Section 2 describes the proposed approach. Simulation and results are presented in section 3. Some considerations for further improvements are given in section 4. The conclusion is drawn in section 5.

2. THE PROPOSED VIDEO AUTHENTICATION SYSTEM

To overcome the problem associated with the extra bandwidth needed for transmitting the signature, the proposed approach uses a combined digital watermarking based on embedding fragile watermark in motion vectors and a digital signature generation described in [5] and [9] respectively. Taking the advantages of both approaches and avoiding their drawbacks. The embedding process of the proposed method is shown in Figure. 1. The major difference between the approach proposed in [9] and the proposed method lies in digital signature embedding position. In [9] the digital signature is embedded as Supplemental Enhancement Information (SEI) in the H.264/AVC bitstream while in the proposed method, the digital signature is used as a fragile watermark and is inserted into motion vectors during the coding process. This is done to address the increase number of in bits transmitted by the decoder in [9].

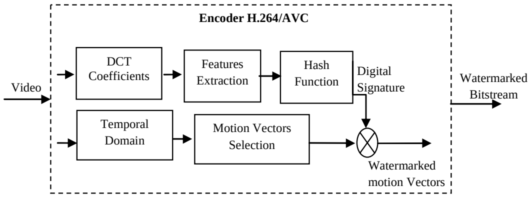


Figure 1 – Watermark generation and embedding process.

2.1 Fragile watermark generation

The core requirement of a fragile watermark is to detect any malicious change. This is easily achieved by using a hashed digest of the original signal to decide the authenticity of the content. Digital signature generation is based on the method proposed in [9] where content-dependent robust bits are extracted from macroblocks and used to authenticate video compressed by H.264/AVC. The H.264/AVC standard supports codes sequences containing I and P slices. I slices contain intra coded macroblocks in which each 16x16 (INTRA 16x16) or 4x4 (INTRA 4x4) luma regions is predicted from previously coded samples in the same slice. The INTER 4x4 is predicted from previously coded samples in the predicted macroblocks from different slices. The INTRA 4x4 mode is based on predicting each luma block separately and is well suited for coding parts of a picture with significant details. The INTRA 16x16 mode, on the other hand, performs prediction and residual coding on the entire 16x16 luma block and is more suited for coding very smooth areas of a picture [3]. To make fragile watermark robust, features of the content that the human eye is sensitive should be used. The features used for digital signature generation are the set of coefficients extracted from INTRA and INTER prediction macroblocks including INTRA 16x16, INTRA 4x4 and INTER 4x4 prediction macroblocks (Figure 2). For INTRA 4x4 and INTER 4x4 macroblocks, the quantized DC coefficient and the first two quantized AC coefficients belonging to low frequency coefficients in zig-zag scan order and surrounding the DC value of every 4x4 block are taken as the feature data for the macroblock. For INTRA 16x16, all the non-zero quantized Hadamard transform coefficients and the first two quantized AC coefficients in zig-zag scan order surrounding the DC value form the feature data for this type of macroblock.

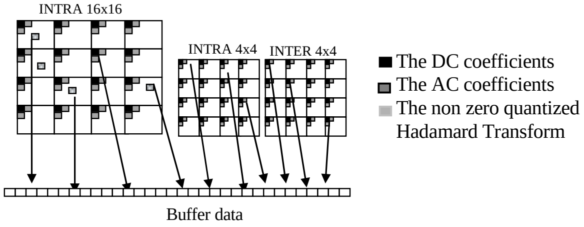


Figure 2 – Features extraction

The DC coefficient represents the mean of every block and contains most of the energy.

If the DC coefficients themselves are changed, the perceptual quality of each frame could be preserved but the digital signature is changed. This would make authentication impossible.

These feature data are collected in a buffer for every coded macroblock within every frame until the end of sequence is reached and then are hashed using Hash function. In this work we used the MD5 (Message-Digest algorithm 5) which produce 128 bits message digest [10], but the method is also valid by using other secure Hash algorithms such as SHA-160 or 256. Finally, this digest acts as a fragile watermark and is used to be and embedded in motion vectors of H.264/AVC.

2.2 Watermark embedding

Because a motion vector can changed easily even for trivial attacks, it is appropriate for fragile watermarking. Therefore, the usage of motion vector information in fragile watermarking makes it possible to judge whether the content is authenticated and its integrity is verified. In this proposed watermarking system, the digital signature used as fragile watermark bits are inserted into the motion vectors MVs of H.264/AVC by changing one Last LSB bit of the two components MV(x, y) noted also MVx and MVy of a set of selected MVs. Each component is an integer number specifying either the horizontal or vertical dimension of the MV. To limit the watermark distortions, two restrictions are performed to select carefully the motion vectors to be embedded:

- Skipped macroblocks and their neighboring blocks are discarded because the motion vector of skipped macroblocks is derived only from motion prediction MVP and there is no MVD (MV) data in the bitstream. The insertion in the neighboring blocks causes motion vector errors in the skipped macroblock thus they could not be compensated.
- The second applied restriction comes from the motion estimation and mode decision delivered by H.264/AVC standard [5]. The embedding is performed on the higher motion activity frames by selecting higher motion activity macroblocks with the best mode 8x8 including four sub modes chosen from 4x4, 8x4, 4x8 and 8x8. By selecting motion vectors of these small partition modes, which represent areas with a lot of detailed motion, it would be difficult for Human Visual System (HVS) to detect distortions introduced by the watermark embedding process.

The remaining motion vectors after the application of the above restrictions and belonging to *k* higher motion activities P-frames are embedded if each component MV(x,y) of the MVs verify the following condition:

$$-T1 \leq MV(x,y) \quad or \quad MV(x,y) \geq T2 \tag{1}$$

where T1 and T2 are two positive thresholds that determine the selected motion vector to be embedded.

The embedding process is performed by changing the last LSB bit of the two components MVx and MVy of the selected embedded MVs. Before it is watermarked, the motion vector component MVx and MVy are quantized as follows:

$$Q(MVx) = \begin{cases} (MVx)\&(0xFFFE) & MVx \geq 0 \\ -(-MVx+1)\&(0xFFFE) & MVx < 0 \end{cases} \tag{2}$$

where the logical AND operation (&) is used to clear the last LSB bit of MVx

The watermark bits w_i are embedded by replacing the last original LSB bit of MVx as follows:

$$\overline{MVx} = \begin{cases} Q(MVx) \mid (0x0001) & if\ w_i = 1 \\ Q(MVx) & otherwise \end{cases} \tag{3}$$

where w_i is the watermark and the vertical bar ($\mid$) represents logical OR operation.

The same principle is applied to embed the vertical component MVy where the two components of motion vector are modified as follows:

$$\overline{MV} = \begin{cases} \{\overline{MVx}, \overline{MVy}\} & if\ w_i = 1 \\ \{Q(MVx), Q(MVy)\} & otherwise \end{cases} \tag{4}$$

Note that last original LSB bits of MVx and MVy are not simply replaced by the watermark bits in this scheme. This is to ensure that the distortion of watermarked motion vectors is minimized. This embedding process ensures the synchronization condition:

$$Q(\overline{MVx}) = Q(MVx) \quad and \quad Q(\overline{MVy}) = Q(MVy) \tag{5}$$

2.3 Watermark extraction

The proposed algorithm extracts the watermark in a blind manner, i.e. the original video sequence is not required for watermark extraction. This process is very similar to the embedding process including the following steps: decoding motion vectors, applying the same restriction performed during the encoding process to select the set of embedding MVs and calculating the average motion vector in each P-frame within the sequence. By using the key *K*, the thresholds T1 and T2 the embedded P-frames are determined and the watermark bits inserted are extracted from the LSB bit of both motion vector components.

2.4 Digital signature verification

In the verification process, the same features are extracted from the H.264/AVC decoder and stored in a buffer. The H/264/AVC decoder stores in a buffer the DC coefficient and the two first AC coefficients in zig-zag scan order before inverse quantization and inverse transform of every 4x4 block of a macroblock for INTRA 4x4 and INTER modes. For INTRA 16x16 modes, all the non-zero Hadamard coefficients. When the end of the sequence is detected, the data is then hashed by the MD5 to produce a digital signature which is compared to the extracted watermark. If the signature and the watermark match, the received video is considered to be authentic and has not been tampered. Otherwise, the video is not authenticated. However, in the case of authentication failure it is not possible to identify frame tampering.

3. EXPERIMENTS AND RESULTS

The proposed algorithm has been integrated into the H.264 JM-10.1 reference software [11]. The most important configuration parameters of the reference software are given in table I. The rest of the parameters have retained their default values. All tests are performed on two groups of well known representative video sequences in QCIF format (Y UV 4:2:0) such as Akio, Miss America and Claire (group A), Foreman, flowers, Carophone and Table (group B). Both groups have the following characteristics:

- Group A: Low spatial detail and low amount of motion;
- Group B: Medium spatial detail and low amount of motion, also high spatial detail and medium amount of motion or vice versa.

Figure 3 shows the payload *Nw* of every P-frame of the sequences belonging to the two groups such as Claire (group A) and Table (group B). The insertion is performed into the two components of the motion vector.

Table II shows the simulation results with the maximum payload for all test sequences belonging to the two groups. In the first group, the motion is low and homogeneous (the sequences are mostly static with limited motion in some frames), so there are less MVs belonging to the best mode 8x8 (with sub-modes chosen from 4x4, 4x8, 8x4 and 8x8) to embed all bits of watermark (128 bits delivered by the MD5).

Table I. Configuration parameters of the encoder.

Profile	Baseline
Number of frames	150 for all test sequences except for Table includes 99 frames
Frame rate	30 fps
Motion estimation	Full Search
Search range	16
Quantification parameter	28
Intra period	O: only the first frame is intra

Table II Simulation Results for sequences belonging to the two groups.

	Sequences	Nw (bits)	PSNR (dB)		Bit Rate (kb/s)	
			original	marked	original	marked
Group A	Miss America	88	40.04	40.04	30.71	30.71
	Claire	135	39.67	39.67	30.51	30.51
	Bridge-close	264	34.85	34.85	87.52	87.53
	Akio	133	38.17	38.17	26.47	26.47
Group B	Carphone	1495	37.29	37.29	86.96	87.02
	Coastguard	2806	34.02	34.02	236.77	236.85
	Flower	4304	34.31	34.31	689.79	689.89
	Foreman	2569	35.75	35.75	113.70	116.51
	Suzie	1349	37.12	37.12	78.91	78.95
	Table	5182	34.91	34.91	563.83	600.44

For example for the sequence Miss America, only 88 bits of the watermark can be embedded in the video containing 150 frames. For Akio, the watermark generated by the MD5 is inserted but if the hash function is changed by other hash functions such as SHA-256 which delivers 256 bits, the insertion of the watermark is truncated and the authentication fails. In the second group, such as Table sequence, the forward and the background are in moving, the video starts with movement of a table tennis player's hand followed by a scene change, there is considerable motion. So there are more best mode 8x8 and more large values of MVs in a frame to watermark (unless high distortions are tolerable). Figure 4 shows one frame with higher motion activities and the partition mode selection map of Akio and Table sequences.

In table II, the bit rate and the PSNR are used as comparative metrics. These two metrics were read from the log.date file created by the encoder. From the results, it can be seen that the PSNR remained unaffected and the bit rate varies slightly. A video authentication signature has to be robust to the transcoding. In some editing process, the compressed videos are transformed to the uncompressed bitstreams which are then re-encoded with one of different compression standards such as MPEG1, MPEG2, H.261 and H.263. In this case, the GOP structure of frames and the motion vectors may change. This is demonstrated by an experiment, which we simply recompress the video with an H.264 encoder, which has the same encoding parameters as the original encoder used to watermark the video. The watermark embedded in MVs is fragile and is removed. Figure 5 and Figure 6 show the visual quality result of the insertion performed on the motion vectors MVs belonging to the 16x16 partition mode and the insertion process described above, respectively. Figure 7 illustrates the visual quality degradation caused by the embedding process without taking into account the conditions of insertion.

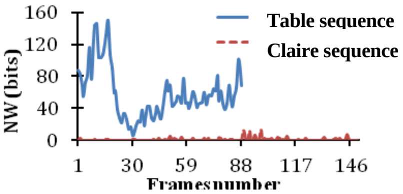
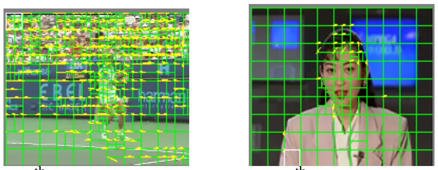
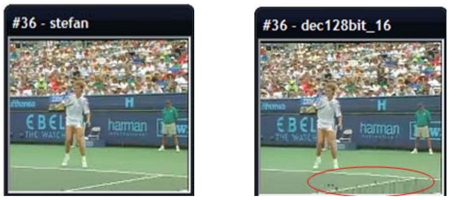


Figure 3 – The payload *Nw* of every P-frames of Claire and Table sequences.



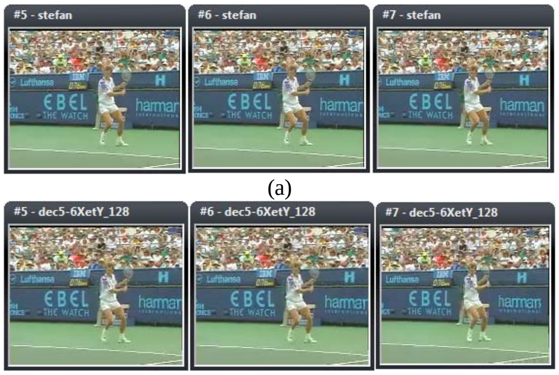
5th frame of Table 18th frame of Akio

Figure 4 - One frame with higher motion activities and the partition mode selection map.



Original Frame watermarked frame

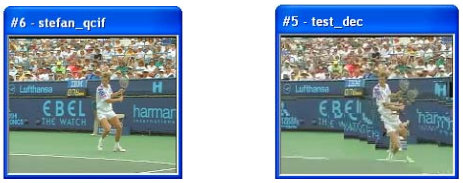
Figure 5 - Insertion in MVs belonging to 16x16 partition mode.



(a)

(b)

Figure 6 – Frames 5, 6 and 7 from Table sequence (a) original frames (b) watermarked frames.



Original Frame watermarked frame

Figure 7 - Insertion without observing the embedding condition.

4. FURTHER IMPROVEMENTS

In order to determine the location of the tampered frames in case of frame tampering, the video is segmented in Group Of Pictures (GOP), and everyone is authenticated independently. The same robust features are extracted from every GOP and then hashed using Secure Hash Algorithm such as SHA-256 to generate the digital signature. The above result is encrypted by public-key cryptosystem. Finally, the encrypted signature acts as a fragile watermark and is inserted into motion vectors within every GOP. The same embedding process is performed on every GOP without using the key *K*. To achieve this, the video sequence must be decomposed into GOP according to the temporal activities and the number of blocks 4x4 in the best mode 8x8 and its sub mode 8x8, 8x4, 4x8 and 4x4. The number of GOP is given by:

$$N_{GOP} = \sum_{f=1}^{n} N_w(f) \Big/ 2(Hash \quad digest) \qquad (6)$$

where N_w *is* the payload of every P-frame of sequence and is equal to twice the number of blocks 4x4 belonging to the best mode 8x8 and its four sub modes. Hash digest is delivered by hash functions.

To detect the location of the tampered frames, the decoder computes the hash of features extracted of every received GOP and matches it with the corresponding decrypted watermark extracted from the embedding MVs. If the signature verification fails, it can be ascertained that the corresponding frames within GOP have been tampered. Furthermore, the proposed scheme not only can verify the authenticity and the integrity of the video, but also can detect the accurate location of the tampered GOPs.

5. CONCLUSION

Conventional digital signature schemes usually encode the signature in a file separated from the original video or embedded it in the compressed bitstream. This increase the bits transmitted by the encoder and require extra bandwidth to transmit it. The proposed approach is based on a combined use of both the fragile watermarking and the digital signature approaches, taking the advantages of both them while avoiding their drawbacks. The proposed scheme extracts features from transform domain of H.264/AVC encoder to generate a unique digital signature and embeds them back into a set of motion vectors belonging to higher motion activities. Furthermore, the scheme not only can verify the authenticity and the integrity of the video, but also avoid an increase in the bit transmitted by the encoder. Experimental results show that the proposed scheme achieves the goal of fragile watermarking with a slight variation in the bit rate while keeping good perceptual quality. In the future, we will focus on authenticating every GOP separately, on the research of covering the tampered frames within a GOP and enhancing the robustness of the watermarking.

REFERENCES

[1] . M. Kim, I. H. Hwan Cho, A. Young Cho and D. S. Jeong, "Semi Fragile watermarking Algorithm for detection and localization of tamper using hybrid watermarking method in MPEG-2 video," *Computational Intelligence and Security*, International Conference, pp. 623–628, 2005.

[2] T. Chen, J. Wang and Y. Zhou, "Combined digital signature and digital watermark scheme for image authentication," *International Conferences on, Info-tech and Info-net*, ICII 2001 – Beijing, Vol. 5, pp. 78-82, 2001.

[3] T. Wiegand, G.J. Sullivan, G. Bjøntegaard and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol*., Vol. 13, No. 7, pp. 560–576, 2003.

[4] G. Qiu, P. Marziliano, A. T. S. Ho, D. J. He and Q. B. Sun, "A hybrid watermarking scheme for H.264/AVC video," in Proc. *17th Int. Conf. Pattern Recogn*., U.K., 2004.

[5] J. Zhang and A. T.S. Ho, "Efficient video authentication for H.264," *IEEE Proc. of the first International Conference on Innovative Computing, Information and Control* (ICICIC'06), 2006.

[6] C. V. Nguyen, D.B.H. Tay, and G. Deng, "A Fast Watermarking System for H.264/AVC Video," *IEEE Asia Pacific Conference on Circuits and Syst.*, APCCAS, pp. 81–84, December 2006.

[7] D. Pröfrock, H.Richter, M. Schlauweg, and E. Muller, "H. 264/AVC video authentication using skipped macroblocks for an erasable watermark," Proc. *Of the VCIP*, Beijing, China, 2005.

[8] S. Ueda, H. Shigeno and K. I. Okada, "NAL Level Stream Authentication for H.264/AVC," *IPSJ Transactions on Database,* Vol. 48, No. 2, pp. 635–643, 2007.

[9] N. Ramaswamy and K. R. Rao, "Video Authentication for H.264/AVC using Digital Signature Standard and Secure Hash Algorithm," *NOSSDAV'06*, Newport, Rhode Island, USA, May 2006.

[10] Rivest R.: The MD5 Message Digest Algorithm. RFC1321 (1992). ftp://ftp.rfc-editor.org/in-notes/rfc1321.txt.

[11] JVT reference Software homepage http://iphome.hhi.de/suehring/ttml.