2018 15th International ISC (Iranian Society of Cryptology) Conference on Information Security and Cryptology (ISCISC)
August 28-29, 2018; Shahid Rajaee Teacher Training University - Tehran, Iran

Tampering Detection and Restoration of Compressed Video

Bardia Azizian
Department of Electrical Engineering, and Electronics Research Institute
Sharif University of Technology
Tehran, Iran
azizian_bardia@ee.sharif.edu

Shahrokh Ghaemmaghami
Department of Electrical Engineering, and Electronics Research Institute
Sharif University of Technology
Tehran, Iran
ghaemmag@sharif.edu

***Abstract*—This paper presents a method to detect tampering of video data and then restore an approximate version of its original contents in compressed H.264/AVC domain using watermarking. In the proposed scheme, a low resolution image from a number of video frames in certain time slots are embedded into the DCT coefficients of the other parts of the video which are adequately far from the reference frames. For detecting temporal/spatial tampering, the index of each frame/macroblock is embedded into itself as an authentication code. If a malicious tampering is detected at the authentication phase, the information used for restoration is extracted to recover the original contents in the altered areas. The extracted images are post-processed to enhance the image quality that could have been affected by channel errors or attacks. Our method provides appropriate transparency and robustness despite large volume of the payload to be handled. The results show that the BER of the watermark signal for recompression attack is about 4.5% on average under QP=24. The main achievement of the proposed watermarking system is the restoration of tampered areas, in addition to high accuracy in detecting malicious tampering in H.264/AVC domain.**

Keywords—digital forensics, video authentication, video self-recovery, H.264/AVC standard, watermarking

I. INTRODUCTION

Vast availability of video recording devices and ease of working with editing tools have made this digital media unreliable or hard to trust as a certain evidence of the recorded data. So, there is a major need for a means to authenticate a digital video and investigate the alterations that might have made to it. Watermarking is one of the most useful approaches for digital forensics. The performance of a watermarking system is evaluated based on three main features: transparency, capacity, and robustness. So far, there are some works that detect tampering and also reconstruct the manipulated areas in raw videos [1-4]. However, such methods cannot resist very low rate video compression, e.g. H.264/AVC, so may not be applicable in most video authentication jobs. This leads to a big challenge and limitation in the area of compressed video authentication that basically come from low redundancy, and consequently restricted capacity of the signal for sufficient payload embedding. The authors of [5] proposed a scheme able to authenticate and recover the lost data of a video using non-negative matrix factorization (NMF) technique, and claim that their method could be applied to any video coding standards. But they have explained no video embedding procedure in compressed domain and no experiments they have reported in the paper for robustness evaluation.

To the best of our knowledge, there is no work for video tampering detection that can also restore the tampered areas in compressed domain, especially for the case of H.264/AVC standard, because of the above-mentioned restrictions. For the same reason, most of the video watermarking methods applied in H.264/AVC video, embed a low payload of data for authentication or copyright protection mostly in intra-coded macroblocks. These watermarking systems need to resist common signal processing attacks e.g. recompression, additive noises, filtering, etc. Embedding in QDCT (quantized DCT) domain could ensure the required robustness. For tampering detection in [6], a content-based code is generated and is embedded into some blocks of an MB (macroblock). The main achievement of the scheme proposed in [6], is a robust method of data embedding in the QDCT coefficients of the MBs. This method is based on the parity of the number of zero elements placed on the secondary diagonal of a 4×4 block. The modulation applied to the coefficients, in case of the incompatibility of the watermark bit with the mentioned feature, changes a nonzero level to zero or vice versa.

One of the most popular embedding algorithms in H.264/AVC domain is inserting the watermark bit into the LNZ (last non-zero) level of QDCT coefficients as used in [7, 8]. Choosing the LNZ QDCT coefficients for data embedding may lead to a lower bit-rate increase, but decreases the robustness as well, because of high rate of changing in the LNZ level index in recompression. The authors of [7], embed k least significant bits of an I4×4 MB's index into the LNZ level of its k different 4×4 SBs (Sub-blocks) by using an LSB-matching like procedure. As expected, their scheme lacks sufficient robustness against common signal processing attacks and the average BER (bit error rate) against recompression attack reported in the paper is about 23.7% under QP=24. In [9], the watermark bit is embedded to a SB by modifying the parity of all non-zero QDCT levels. The extracted bit is determined based on the majority of non-zero level's parity. This method is robust just in case of low capacity watermarking applications.

The major watermarking challenge addressed in our work is robustness, which is influenced by high payload capacity embedding. Modifying the video contents due to the high capacity of data embedding changes the prediction modes in recompression, alters the new DCT coefficients totally, and increases the BER. Accordingly, we have optimized our scheme based on its robustness against recompression.

978-1-5386-7582-3/18/$31.00 ©2018 IEEE

The rest of the paper is organized as follows: In section II, our proposed scheme is presented in detail. Experimental results and discussions are given in section III and concluding remarks can be found in section IV.

II. Proposed Scheme

In the scheme proposed here, the video frames are segmented into equal consecutive groups, called R-GOP (restoration group of pictures). From the first frame of each R-GOP a low resolution image is extracted that represents the corresponding R-GOP. Each image extracted from an R-GOP is embedded into the frames (excluding the first frame) of another R-GOP, with a reasonable distance from the current R-GOP, to increase the possibility of extracting an intact image to be employed to restore the current R-GOP in case of attack. For spatial tampering detection of an R-GOP, a specific ID is embedded into the MBs of its first frame. We set the number of R-GOP's frames equal to the integer multiple of the intra-period. By this constraint the first frame of each R-GOP becomes an I-frame.

The temporal length of the R-GOP affects the robustness, capacity and transparency of the watermarking system. A shorter R-GOP yields a higher temporal resolution for tampering examination, while the payload for data embedding per frame is increased and, as a consequence, transparency and robustness are decreased. Due to our experiments, the proposed scheme works well with an R-GOP of length 1-second or longer for a video with 30fps frame rate, which is a reasonable time interval for detecting a visible tampering changes like object addition or removal. Also, to detect the temporal tampering (e.g. adding or dropping frames) the number of each frame in the video is embedded into some MBs of the current frame. Fig. 1 shows the block diagram of the whole scheme when an attack happens. Details of the proposed method are given below.

A. MB's ID Embedding

The first frame of each R-GOP is used for embedding the IDs of MBs. The ID of each MB is composed of the k least significant bits of its index. These k bits are embedded into k 4×4 SBs of the related MB of both I_4×4 and I_16×16.

B. Frame Number Embedding

The n least significant bits of the number of each frame are inserted into some of its specific MBs, d times. The greater d, yields higher robustness for frame number extraction and can distinguish the malicious temporal tampering from the errors that occur randomly by performing a typical signal processing modification or recompression. So, in each frame (both I-frame and P-frame) $n.d$ bits are embedded into $n.d$ different MBs (at most 1 bit in each MB).

C. Image Generation

The image's luminance component of the first frame of each R-GOP is derived, and then this image is decimated in both horizontal and vertical directions by factor 2. So the resolution of the resulting image is a quarter of the original one. Next, 8 bits of each pixel are re-quantized to q bits, so we get a grayscale image with q-bit pixels (Fig. 4 (d)). Thus, the generated image contains q/4 bits information per each pixel of the original frame.

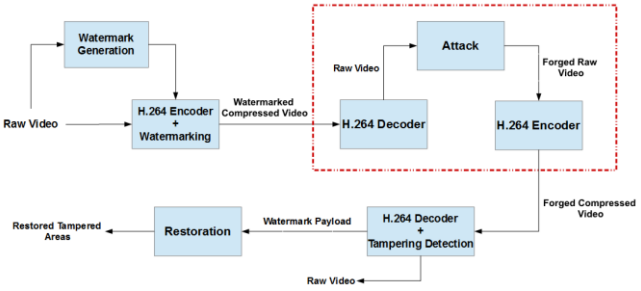

Fig. 1. Block diagram of proposed method when an attack happens

D. Image Embedding

For deriving a WS (watermark signal) based on the images generated in the previous stage, some preprocesses are performed on them. First, the q bits of each pixel are encoded to its gray-code to prevent significant alterations to extracted image when one bit is corrupted. Then, all of the encoded bits of the image are placed in a one dimensional array and permuted pseudo-randomly by a key shared between sender and receiver. Now, the WS of an R-GOP is derived and should be embedded into its destination. The destination R-GOPs are selected based on a condition that temporal distance between source and destination R-GOPs turns into its maximum. If total number of the R-GOPs in a video is r, the WS of R-GOP number x is embedded into the R-GOP number $Dest(x)$ according to (1).

$$Dest(x) = x + \left\lfloor \frac{r}{2} \right\rfloor \ (mod\ r)\ ;\ (0 \le x \le r-1) \qquad (1)$$

The procedure of the embedding in destination R-GOP is MB-based. To reduce sensitivity of the system to the drifts in the extraction phase, we choose the number of the embedded bits in an MB independent of changes that may occur by an attack. So, in each MB of a specific frame of the destination R-GOP, a constant number of bits are embedded based on the frame type. Therefore the payload capacity is inserted uniformly among the MBs, where the MBs with the specific index along an R-GOP (excluding the ones belonging to the first frame) should carry $q/4 \times 16 \times 16 = 64q\ bits$ data, totally. Because of higher capacity of I_MB (Intra-coded MB) than P_MB (Inter-coded MB), I_MB assures higher robustness. If the capacity of a SB is at most one bit, we embed the maximum possible number of bits in MBs of I-frames which is 15 (one extra bit for frame number embedding). MBs of P-frames carry lower part of the payload. It should be noted that spreading the mentioned $64q$ bits along with the R-GOP's MBs is based on the type of the frame, not on the type of the MB prediction that could be altered by an attack. In other words, it may happen that an I_MB exists in a P-frame, but it carries the same payload capacity as a P_MB does.

E. Embedding Procedure

As mentioned earlier, our main concern in this work is robustness of the watermarking system designed. Hence, an embedding algorithm is proposed which employs a robust property of a mid-frequency QDCT level of the macroblocks' residuals for both I_MB and P_MB as demonstrated in (2), where C_i is the i^{th} quantized DCT level (after zig-zag scan) in which w (the watermark bit) should be embedded. p is a uniform random variable in [0,1]. Equation (3) shows the extraction procedure of a watermark bit ($\bar{w}$).

$$C_i = \begin{cases} 0 & if \quad w = 0 \\ +1 & if \quad w = 1 \quad and \quad C_i = 0 \quad and \quad p \geq \frac{1}{2} \\ -1 & if \quad w = 1 \quad and \quad C_i = 0 \quad and \quad p < \frac{1}{2} \\ C_i & if \quad w = 1 \quad and \quad C_i \neq 0 \end{cases} \tag{2}$$

$$\bar{w} = \begin{cases} 0 & if \quad C_i = 0 \\ 1 & if \quad C_i \neq 0 \end{cases} \tag{3}$$

Changing a QDCT level of a SB with all-zero elements to a non-zero value causes a big increase in the video bit-rate so, in MBs of P-frames, the SBs with at least one non-zero element have higher priority to carry the watermark bits. Since the MBs of I-frames carry the maximum possible number of bits, all of their SBs are selected for the embedding.

Now, we want to find the optimized *i* based on increasing the robustness against recompression which is the inevitable part of each attack as illustrated in the red block of Fig. 1. Since the low frequency coefficients have a higher effect on the contents of video, modifying these coefficients for data embedding leads to more prediction modes to be changed and consequently higher BER in recompression. Besides, high-frequency coefficients are more vulnerable to the common signal processing attacks. To find the proper DCT level, a random watermark signal is embedded into 4 test sequences, i.e. foreman, mobile, hall and mother-daughter, according to the above-mentioned embedding algorithm. The number of bits embedded into the MBs of I-frames and P-frames is 15 and 3, respectively. The average BER of the extracted watermark for MBs of I-frames and P-frames, after recompression, versus *i* (index of DCT level) for different QPs is shown in Fig. 2 and Fig. 3, respectively. Apparently, in these figures, the BER has a significant drop in *i=11* in both I-frames and P-frames generally. The reason is that the 11th DCT level is placed on the diagonal position of the SB after zig-zag scan and diagonal elements, in the mid-frequency range, are more robust against recompression, since they are symmetric in horizontal and vertical directions. Therefore, C_{11} is chosen to carry the payload in our scheme.

F. Tampering Detection

In the decoder, first, the *n.d* bits related to the number of frames are extracted and the bits of the frame numbers are obtained by means of the majority rule applied to its *d* repeats. If a temporal tampering is detected, it should be compensated to avoid the drift in I-frames and P-frames. For example, if some extra frames are inserted into the video, they will be deleted, or if some frames are removed, some extra frames will be replaced to return the I-frames and P-frames into their main position.

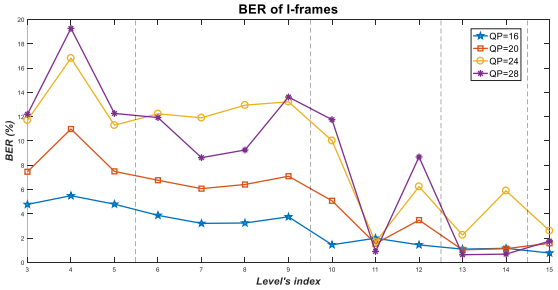

Fig. 2. BER of I-frames – level's frequency, for four different QPs

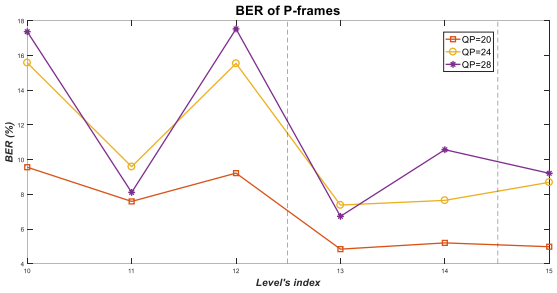

Fig. 3. BER of P-frames – level's frequency, for three different QPs

After this step, the MBs' ID of the R-GOPs' first frame are extracted and the BER of each MB is calculated using (4). Now we can generate a tampering pattern image in which the intensity of each MB is directly affected by the related BER (Fig. 4 (e) and (f)). According to this pattern, we can detect the suspicious tampered areas.

$$BER_i = \frac{number\ of\ errors\ in\ i^{th}\ MB}{k} \tag{4}$$

G. Image Extraction

The procedure of image extraction is exactly reverse of its embedding method. After extracting the WS, permuting it by the related security key, and converting the q-bit gray codes to their main values, a noisy image is obtained. To improve the image quality a post-processing is done on these images by the DWM (directional weighted median) filter [10]. This filter detects the spots that could be an impulse noise with random value based on a threshold, set before. After detecting these spots, they are substituted with the weighted median of their neighbors lied on one of four main directions which is closer to the shape of that region. The filtering process is performed iteratively by decreasing threshold to reach to an appropriate quality.

III. Experimental Results

Our proposed method is implemented on x.264 which is an H.264/AVC encoder. For extracting the watermark data, we used libav codec as the decoder of H.264/AVC video stream. To evaluate the performance of our watermarking system, four standard video sequences (hall, foreman, mother-daughter, mobile), are employed. Table I, shows the configuration and initialization parameters of the x.264 and our method. By choosing *q=3* the payload capacity per macroblock gets to 192 bits. Embedding 15 and 3 bits in MBs of I-frames and P-frames, respectively, provides the required capacity in the destination R-GOPs for *Intra-period=3* and *|R-GOP|=30*.

TABLE I. Configuration Parameters

Parameter	Configuration	Parameter	Configuration
Profile	Baseline	Frame number	300
Frame rate	30 fps	Video resolution	CIF (352×288)
Intra-period	3	QP	16, 24
Entropy coding	CAVLC	*|RGOP|*	30
k	6	*n*	8
q	3	*d*	5

TABLE II. QUALITY AND BIT-RATE ANALYSIS

QP	Video	Foreman		Hall		Mobile		Mother-daughter		Average	
	Method	Proposed	LNZ based	Proposed	LNZ based	Proposed	LNZ based	Proposed	LNZ based	Proposed	LNZ based
QP = 16	PSNR drop (dB)	*0.69*	*0.26*	*0.52*	*0.22*	*0.50*	*0.21*	*0.47*	*0.35*	***0.55***	***0.26***
	SSIM drop	*0.0013*	*0.0008*	*0.0011*	*0.0009*	*0.0005*	*0.0003*	*0.0012*	*0.0011*	***0.0010***	***0.0008***
	BIR (%)	*3.0*	*2.9*	*1.7*	*2.9*	*3.0*	*1.0*	*7.6*	*8.3*	***3.8***	***3.8***
QP = 24	PSNR drop (dB)	*0.77*	*0.52*	*0.62*	*0.34*	*1.22*	*0.31*	*1.07*	*0.88*	***0.92***	***0.51***
	SSIM drop	*0.0070*	*0.0049*	*0.0065*	*0.0037*	*0.0043*	*0.0019*	*0.0094*	*0.0067*	***0.0068***	***0.0043***
	BIR (%)	*11.9*	*5.7*	*14.9*	*13.0*	*4.7*	*2.5*	*39.8*	*28.9*	***17.8***	***12.5***

To make a fair comparison, we have implemented two embedding algorithms with the same capacity and the same procedure of watermarking, described in section II. One of these algorithms is the LSB-matching like procedure modifying the LNZ level of QDCT as used in [7, 8], and the other one is the method proposed here.

A. Quality, Bit-rate, and Robustness

Table II, shows the results of PSNR (peak signal to noise ratio) drop, SSIM (structural similarity) index drop, and BIR (bit-rate increase ratio) of the watermarked compressed videos, as compared to the original compressed ones. As expected, the LNZ based method performs better than our proposed algorithm in terms of transparency, because LNZ based method mostly changes the higher frequency components and the modifications it makes to DCT coefficients is one QP step at most. In general, the proposed scheme has an appropriate transparency and it could be seen under a subjective comparison between the watermarked and original video in Fig. 4 (a) and (b). The large payload of the WS, should be embedded into the video, enforces to employ some SBs with all-zero elements for data embedding in both methods. This causes a high increase in the bit-rate, especially for higher QPs, for which there are more such SBs.

The robustness of the watermarking scheme against common signal processing attacks is shown in Table III for QP=24. In this table, the average BCR of the watermark signal and the mean SSIM between the extracted and embedded restoration images after DWM filtering for the test sequences is presented. The SSIM greater than 0.6 assures a perceptible image. As shown in Table III, the proposed scheme can resist the most common signal processing attacks, especially recompression, where the extracted hidden images could have an acceptable quality. Another important point in this table is the advantage of our scheme, as compared to LNZ embedding method. Also, the results show that robustness of the proposed method excels the most of the similar tampering detection schemes in H.264/AVC domain, despite the high payload capacity offered by our method.

B. Spatial Tampering Detection

The main achievement of our proposed method is that it not only detects the spatial tampering, but also can restore the tampered areas. As an example, we show a forgery attempt on the hall video sequence, in which the man appearing in the video at the right side of frame number 72 has been removed (see Fig. 4 (c)).

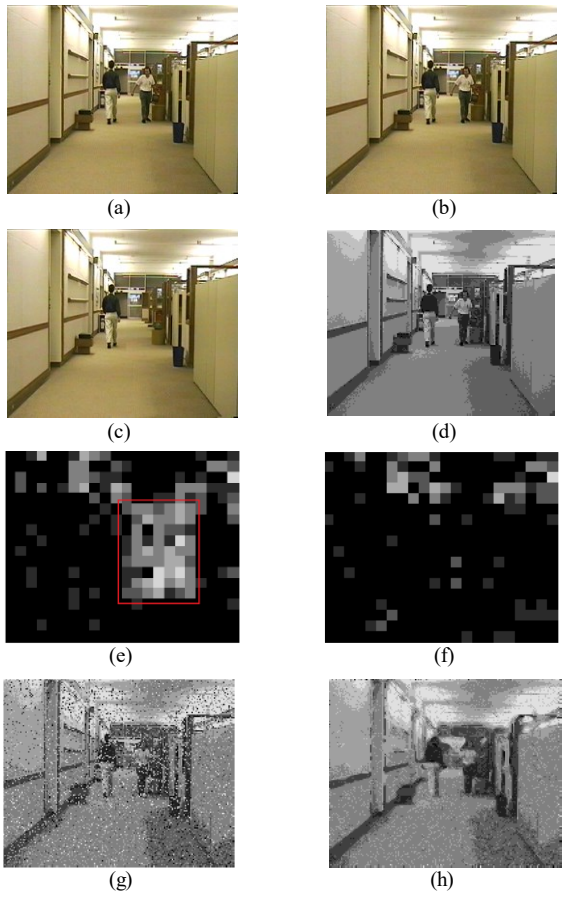

Fig. 4. Spatial tampering, (a) the original frame 181 (b) the watermarked frame 181 (c) the forged frame 181 (d) the embedded image in R-GOP 2 (e) the tampering pattern of R-GOP 7 (f) the tampering pattern of R-GOP 2 (g) the extracted image from R-GOP 2 (h) the DWM filtered image

Since the number of frames in the R-GOPs is 30, the first two R- GOPs remain intact. In Fig. 4, the procedure of restoring the 181st frame is given. As shown, the tampering pattern of the 7th R-GOP (frame 181) is suspicious. The pattern of the other R-GOPs except R-GOP 1 and 2 is the same as that shown in the Fig. 4 (e), so we figure out that the possibility of adding or removing an object in red box is high. By observing the pattern of 1st and 2nd R-GOP, and the difference between them and the other R-GOPs, we guess that the corruption of the extracted data from these two R-GOPs would be less. The restored frame of R-GOP 7, is obtained by extracting its related data from R-GOP 2 as shown in Fig. 4 (g) and (h). It should be noted that the image of the R-GOP 7 is embedded into the R-GOP 2 (see (1)).

TABLE III. ROBUSTNESS AGAINST COMMON SIGNAL PROCESSING ATTACKS

QP = 24	Attack	Recompression			AWGN	Salt & Pepper		Gaussian Filtering		Brightness Increase	
	Method	$QP_2 = 22$	$QP_2 = 24$	$QP_2 = 26$	$\sigma = 0.01$	$d = 0.001$	$d = 0.01$	$\sigma = 0.3$	$\sigma = 0.4$	+10	+20
BCR (%)	Proposed	*96.32*	*95.42*	*84.93*	*79.65*	*90.72*	*74.32*	*92.77*	*85.40*	*92.46*	*90.68*
	LNZ based	*58.33*	*79.37*	*75.9*	*63.77*	*72.12*	*63.41*	*72.83*	*58.21*	*72.79*	*71.13*
SSIM	Proposed	*0.7966*	*0.7673*	*0.4794*	*0.4731*	*0.6883*	*0.4331*	*0.7051*	*0.5236*	*0.7110*	*0.6676*
	LNZ based	*0.5445*	*0.4685*	*0.0893*	*0.2783*	*0.4019*	*0.2992*	*0.4042*	*0.2000*	*0.4056*	*0.3721*

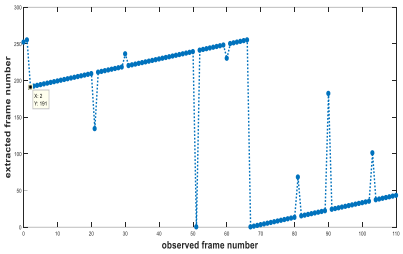

Fig. 5. Temporal tampering detection plot

(a) (b) (c)

Fig. 6. Extracted images from the tampered video after filtering

C. *Temporal Tampering Detection*

To examine accuracy of the proposed scheme in temporal tampering detection, a forgery attempt has been run on the foreman video sequence. The first 190 frames of this video in which a foreman is looking to the camera are dropped. Since the intra-period is 3 and 190 is not an integer multiple of 3, there should be a shift in the order of I-frames and P-frames. This drift converts all of the I-frames to P-frames and disrupts the extraction process. After compensation of this drift by duplicating the first frame and recompressing the video, the hidden frame numbers are extracted. Fig. 5 shows the extracted frame number versus the observed frame number. By neglecting the errors happened due to the recompression, this plot is linear and its y-intercept is 189. So, we conclude that the first 190 frames of this video were removed. The first frame of R-GOPs number 3, 4, and 5 of the original video could be restored from the last 90 frames of the 110 available frames in the forged video, as shown in Fig. 6.

IV. CONCLUSION

We have presented a method that detects tampering in a video compressed by H.264/AVC standard and restores the tampered areas as well. In the proposed scheme, we are able to recover some manipulated frames which are the first frame of an R-GOP using the data embedded into the other frames of the video. The most important challenge in this work has been to create a high capacity for data embedding into the highly compressed H264.AVC video. So, we have chosen to embed the data in both intra-coded and inter-coded macroblocks. To achieve higher robustness, we have selected a mid-frequency diagonal component of the QDCT coefficients of macroblocks for hiding the data bits. The watermark bits are embedded in the selected QDCT level. The results show that the proposed scheme is more robust to common attacks, as compared to other methods for detecting tampering of video in H.264/AVC format.

REFERENCES

[1] M. U. Celik, G. Sharma, A. M. Tekalp, and E. S. Saber, "Video authentication with self-recovery," pp. 531-541, 2002.

[2] B. G. Mobasseri and A. T. Evans, "Content-dependent video authentication by self-watermarking in color space," pp. 35-44, 2001.

[3] A. M. Hassan, A. Al-Hamadi, Y. M. Hasan, M. A. Wahab, and B. Michaelis, "Secure block-based video authentication with localization and self-recovery," *image,* vol. 21, p. 23, 2009.

[4] Y. Shi, M. Qi, Y. Yi, M. Zhang, and J. Kong, "Object based dual watermarking for video authentication," *Optik - International Journal for Light and Electron Optics,* vol. 124, pp. 3827-3834, 2013.

[5] M. Tong, J. Guo, S. Tao, and Y. Wu, "Independent detection and self-recovery video authentication mechanism using extended NMF with different sparseness constraints," *Multimedia Tools Appl.,* vol. 75, pp. 8045-8069, 2016.

[6] D. Xu, R. Wang, and J. Wang, "A novel watermarking scheme for H.264/AVC video authentication," *Signal Processing: Image Communication,* vol. 26, pp. 267-279, 2011.

[7] M. Fallahpour, S. Shirmohammadi, M. Semsarzadeh, and J. Zhao, "Tampering Detection in Compressed Digital Video Using Watermarking," *IEEE Transactions on Instrumentation and Measurement,* vol. 63, pp. 1057-1072, 2014.

[8] L. Tian, N. Zheng, J. Xue, and C. Li, "Authentication and copyright protection watermarking scheme for H.264 based on visual saliency and secret sharing," *Multimedia Tools and Applications,* vol. 74, pp. 2991-3011, 2015.

[9] Z. Ma, J. Huang, M. Jiang, and X. Niu. A Video Watermarking DRM Method Based on H.264 Compressed Domain with Low Bit-Rate Increasement. *Electronics 25(4),* 641-647, 2016. Available: at: http://digital-library.theiet.org/content/journals/10.1049/cje.2016.07.010

[10] Y. Dong and S. Xu, "A New Directional Weighted Median Filter for Removal of Random-Valued Impulse Noise," *IEEE Signal Processing Letters,* vol. 14, pp. 193-196, 2007.