Multimed Tools Appl (2014) 72:2469–2495
DOI 10.1007/s11042-013-1561-2

A low cost fragile watermarking scheme in H.264/AVC compressed domain

Shi-Jinn Horng · Mahmoud E. Farfoura · Pingzhi Fan · Xian Wang · Tianrui Li · Jing-Ming Guo

Published online: 27 June 2013

Abstract H.264/AVC-based products have grown tremendously in social networks; issues of content-based authentication become increasingly important. This paper presents a blind fragile watermarking scheme for content-based H.264/AVC authentication, which enjoys high sensitivity to typical video attacks. A spatiotemporal analysis is exploited to guarantee a minimum impact on perceptual quality and bit-rate increment. The watermark features are extracted from intra/inter prediction modes of intra/inter macroblocks, constituting the content-based Message Authentication Code (MAC) which is embedded/extracted in a Group-of-Pictures GOP-based fashion utilizing the syntactic elements of the Network Application Layer (NAL) units from the compressed bitstream. It's unnecessary to fully decode a compressed bitstream before the embedding or detection processes. A content-based key is generated to control fragile watermark generation, embedding, extraction, and verification algorithms. Additionally, fragility is ensured by selecting the last nonzero

S.-J. Horng · P. Fan · X. Wang · T. Li
School of Information Science and Technology, Southwest
Jiaotong University, 610031 Chengdu, People's Republic of China

P. Fan
e-mail: pingzhifan@gmail.com

X. Wang
e-mail: drwangxian@gmail.com

T. Li
e-mail: trli@home.swjtu.edu.cn

S.-J. Horng (✉) · M. E. Farfoura
Department of Computer Science and Information Engineering,
National Taiwan University of Science and Technology, Taipei 106, Taiwan
e-mail: horngsj@yahoo.com.tw

M. E. Farfoura
e-mail: mfarfora@yahoo.com

J.-M. Guo
Department of Electrical Engineering, National Taiwan
University of Science and Technology, Taipei 106, Taiwan
e-mail: jmguo@seed.net.tw

quantized ac residuals for watermark embedding. The embedded watermark can be detected and verified by means of partially decoding intra/inter prediction modes from syntactic elements of the bitstream without the prior knowledge of the original video or complete decoding. Experiment results demonstrate that the performance of the proposed scheme is excellent in terms of bit-rate and perceptual quality. Furthermore, various types of content-preserving and/or content-changing attacks can be detected efficiently.

Keywords H.264/AVC · Video watermarking · Intra/inter prediction mode · Discrete cosine transform · Content-authentication

1 Introduction

The ever evolving social networks, blogs, and video sharing websites make it rather easy to access, edit and redistribute digital multimedia contents such as images, video and audio, multimedia content-authentication has become an ongoing and constant requirement for protecting the media. In multimedia context, video authentication aims to establish its authenticity in time, sequence and content. A video authentication scheme ensures the integrity of digital video, and verifies that the video presented into use has not been tampered with. Digital watermarking provides a promising method of protecting digital data from illicit copying and manipulation by embedding a secret code directly into the data [19–21, 33].

Recently, video transcoding [2] which is a core technology for providing universal multimedia access by the Internet users with different access links and devices has become a disturbing issue for some video stream owners or producers. Consequently, there is a crucial need to protect the video streams from being forcibly transcoded and illegally re-distributed. One class of authentication watermarks is hard authentication [38] which rejects any intentional or unintentional modifications to the video bitstream and can be considered as a form of lossless authentication. The inserted watermark is so weak that any manipulation to the video content disturbs its integrity. Digital signatures are one way of achieving hard (lossless) authentication.

The well-established H.264/AVC video coding standard [1] was jointly developed by the ITU-T VCEG and the ISO/IEC MPEG standards committees. It achieves clearly higher compression efficiency, often quoted as, up to a factor of two over the MPEG-2 video standard [11]. A number of error-resilience tools to tolerate errors in H.264/AVC have been developed [13]. However, no specific tool and/or mechanism considers data integrity and authenticity of the transferred bitstreams. Therefore, data integrity and authenticity is still an open issue in H.264/AVC.

The early video watermarking approaches originated from still image watermarking techniques [5], which were extended to video by hiding the watermark in every frame separately, are infeasible to be applied directly to the H.264 standard, since it needs a full decoding and re-encoding for embedding or watermark detection. In this paper, the watermark embedding is applied in the compressed domain [18], in which the original video is provided to the embedder as a stream of bits. The embedder partially decodes the stream and parses the syntactical elements of the compressed video, such as transform coefficients, motion vectors, and intra/inter prediction modes. The elements of the partially decoded video are modified to insert the watermark, and then reassembled to form the compressed watermarked video stream. In doing so, the embedder has sufficient information, such as prediction, motion, and quantization parameters, which allows informed decisions to improve fragility, imperceptibility, and bit-rate control.

In digital video watermarking, due to the unrestrained watermark embedding, two major perceptual artifacts, spatial noise and temporal flicker, may arise. Embedding watermark in smooth areas, renders perceptual distortion in most cases, while inserting different watermarks into video frames independently without taking the temporal dimension into account, usually yields a flicker effect in video which is induced from the differences between the intensities of pixels at the same position in two successive video frames [17].

Many researchers have tackled the copyright protection and content-authentication issues in the well-established H.264/AVC coding standard which adopts many new features [29]. In the open literature, most of the proposed watermarking schemes operate during the encoding phase. In [9], a blind robust DCT-based watermarking method was proposed. This method is robust against compression. In [24], a non-blind robust watermarking method based on Human Visual Model (HVM) was proposed. This method solves the error pooling effect discussed in [33], yet it suffers from two problems: 1) The payload capacity was not convincing, and 2) the original video must be available for watermark extraction. Guo et al. [27] proposed a hybrid scheme including a robust watermark embedded in the DCT domain and a fragile watermark in motion vectors during the H.264 encoding. Their scheme suffers from security problems due to exploiting just the diagonal coefficients for embedding. In [25], the robust algorithm proposed in [24] is extended for embedding a watermark in P-frames by considering the HVM and temporal domain analysis to preserve the visual quality.

In [26], an authentication scheme for H.264 video was proposed, in which the watermark is embedded by reactivating some of the skipped macroblocks (skip-MBs). This scheme has several pitfalls: 1) Skipped MBs are 16×16 macroblocks and embedding in which may induce prominent artifact. 2) Skipped macroblocks are sent to the decoder with no coded coefficients, no header, and no prediction information, and reactivating them will cause additional overhead which thus increases the corresponding bit-rate. 3) Reactivating skipped MBs may render some security breaches which help an attacker to distinguish reactivated skipped MBs with simply one nonzero ac residual.

Kapotas et al. [14] proposed a fragile method utilizing the intra IPCM-block type for watermark embedding. The embedding is conducting over the Least Significant Bits (LSB) of the luma and chroma components in the spatial domain. This method cannot detect malicious content modification outside the IPCM-blocks, and it also increases the bit-rate. Kim [15] devised an entropy coding based data hiding method to embed a watermark bit in the sign bit of the trailing ones in Context-Adaptive Variable Length Coding (CAVLC) of the H.264 bitstream. This method suffers from flickering artifacts in the temporal direction, since the errors incurred by the bit-modification are accumulated throughout the intra frame in the raster scan order by the intra predictions. Liu et al. [22] embedded watermark bits by changing the best block type determined by the Rate-Distortion Optimization (*RDO*). By constraining the prediction modes on different block sizes, the bits of “0” and “1” can be hidden, which inevitably will have a negative effect the final PSNR and bit-rate.

In [12, 37], authors exploited the intra prediction modes of qualified intra 4×4 luminance blocks to hide information data based on mapping rules and matrix coding. The methods of Hu [12] and Yang [37] hide the secret data by modifying the best intra prediction mode for some intra 4×4 luminance blocks, along with the mapping rules. However, the rules were derived from the statistical analysis of certain testing sequences which are not always optimal for any video sequence. In addition, the mapping rules should be sent to the decoder to perform the watermark extraction. Wang et al. [32] proposed a fragile watermarking scheme, in which the watermark embedding is performed into the last nonzero quantized coefficient of each DCT block during the encoding process. Unfortunately, the results show a high distortion induced by the watermark insertion due to the unacquainted watermark embedding.

In [30], watermark features are extracted as the authentication data from DCT domain to generate a unique digital signature using an MD5 hash 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. This method suffers from high visible distortion perceived in their subjective evaluation results. Kim et al. [16] proposed a fragile watermarking method scheme that inserts a watermark bit on the motion vectors' LSB for inter-coded MBs or on the mode number for intra-coded MBs. For skip-MB type MBs, a watermark bit is inserted at the first nonskip-MB type MB with the same coordinate in the following frames. This method suffers from security problem in which an attacker can easily localize the watermark embedding positions, and thus guessing the embedded watermarks. Xu et al. [36] proposed a semi-fragile authentication technique that extracts a self-authentication code and then embeds into the diagonal DCT coefficients in the encoding phase of H.264/AVC. This method suffers from a security problem due to exploiting just the diagonal coefficients for watermark embedding.

In the compressed domain, Xu and Wang [35] proposed a fast fragile watermarking algorithm for the H.264/AVC using Exponential-Golomb (Exp-Golomb) code words mapping. Watermark embedding is performed by modulating the Exp-Golomb coded reference frame index in the bitstream. The algorithm is claimed to be fast and preserves the video coding efficiency and high payload capacities. However, since the optimal reference frames are modified due to watermark embedding, maintaining perceptual transparency is not achievable. This can be drawn clearly from their reported objective observations. In [23], a robust low complexity DCT-based scheme in the H.264 compressed domain is proposed. The watermark embedding is performed based on a spatiotemporal analysis utilizing the useful available information in syntactic elements of the H.264 stream. This scheme performs well in preserving the coding efficiency, but the reported results show relatively low payload capacity.

To circumvent the previous drawbacks, an improved low complexity content-based hard authentication scheme which can detect content-preserving attacks and/or content-changing attacks for H.264/AVC compressed domain is proposed. The concept of the proposed scheme is to extract fragile features such as intra prediction modes of intra 4×4 luminance sub-blocks (I4-block) and 16×16 blocks (I16-block) of I-frames, then generate a content-based Message Authentication Code (MAC) to be encrypted using a content-based key then embedded into the last nonzero quantized residuals of selected luma intra predicted I4-blocks of I-frames based on an efficient spatiotemporal analysis in a GOP-based fashion. The content-based key derived by some features and a secret symmetric key known only to the H.264 stream owner. Accordingly, the embedded watermark protects all kinds of MBs and frames. The authentication information can be detected and verified blindly from the encoded bitstream without the need of the original host video. Figure 1 illustrates the schematic block diagram of the proposed scheme.

The rest of the paper is organized as follows. In Section 2.1, we discuss in details the fragility problem and perform a spatial analysis to enhance the fragility of the watermarking algorithm. Section 2.2 explains the proposed low complexity temporal method. The fragile watermark generation is presented in Section 2.3. Sections 2.4–2.5 present the watermark embedding, watermark extraction and verification. Then, Section 3 illustrates the experimental results and discussions. Finally, conclusions and future work are drawn in Section 4.

2 The proposed watermarking scheme

Fragile watermarking is a popular technique for digital multimedia content-based authentication. The essential requirement of a fragile watermarking is to detect any content-preserving manipulations or content-changing manipulations. This is easily achieved by utilizing a hashed digest of the original signal to determine the authenticity of the content.

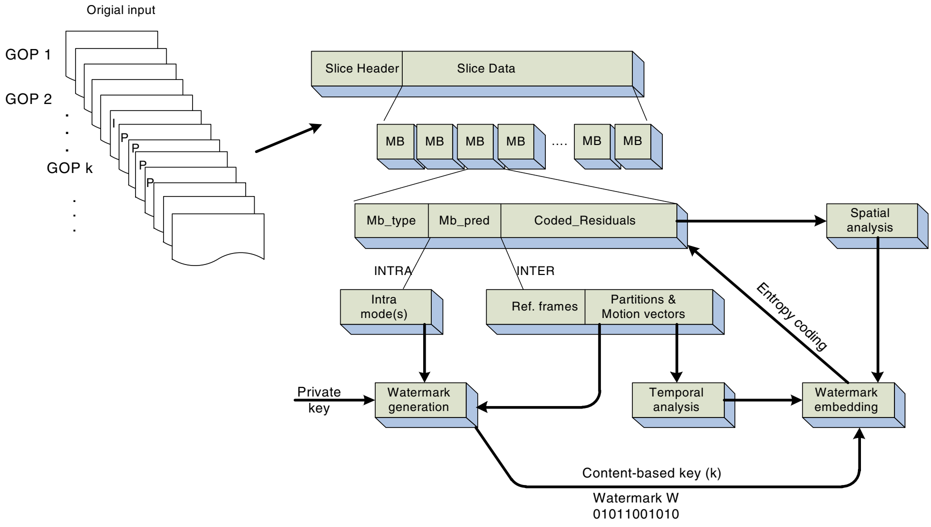

Fig. 1 Schematic block diagram of the proposed scheme

Message digest generation in which content-based bits are extracted from the structural information of the video content is used to authenticate video streams. In this paper, a fragile watermarking scheme is proposed for the well-established H.264/AVC bitstreams to verify whether video data are authentic or not.

In the proposed watermarking scheme, an encrypted content-based Message Authentication Code (MAC) is embedded/extracted in a GOP-based fashion using the syntactic elements of the compressed bitstream. The content-based MAC consists of a vector of hashed and encrypted intra/inter prediction modes of the luma components of the intra/inter MBs. This vector acts as the authentication information to be embedded into last nonzero quantized residuals of selected I4-blocks in I-frames based on spatiotemporal analysis. The majority of previous works in the field of H.264/AVC watermarking embed the watermark information into I-frames because any tampering with these frames would lead to immediate effect on the subsequent P- and B-frames in terms of perceptual quality. Conversely, P- and B-frames are highly compressed by motion compensation, and thus normally they have less capacity to embed additional information.

Embedding watermark into last nonzero quantized ac residuals of selected I4-blocks has three benefits:

1) Perceptual quality: Since the Discrete Cosine Transform (DCT) employed in H.264/AVC coding has a strong “energy compaction” property: i.e. most of the signal information tends to be concentrated mainly in the low-frequency part of the DCT spectrum. Consequently, modifying the last nonzero residuals reduces the perceptual distortion.
2) Fragility or sensitivity to tampering: The predefined zigzag scan order reorders the quantized ac residuals from low to high frequency. Thus, modifying the residuals in high frequency, i.e. last nonzero residuals, enjoys the benefit of higher sensitivity against re-encoding and signal processing manipulations.
3) Bit-rate control: Since the modification does not change the order of residuals, this will not affect each run-length after zigzag scan. In other words, keeping nonzero residuals at the beginning and long runs of zeros at the end of the data stream, makes the run-length coding very efficient.

In this study, the coefficient scanning order scheme for directional spatial prediction-based techniques proposed in [34] is adopted to further improve the bit-rate of the final watermarked video streams. The authors show that the probability distribution of the ac coefficient values is related to the selected intra prediction mode. Based on the statistics and mathematical analysis, they proposed two new coefficient scanning schemes based on the selected intra prediction mode. In this paper, these two scanning schemes (vertical and horizontal) are implemented to reduce the bit-rates of the watermarked video streams.

To combat intra-collusion attack [8], in which a unique key is used to embed the same watermark in all frames, the structural information of the H.264 stream is utilized. The content-based public key (*K*) is derived from 16 intra prediction modes of 16 I4-blocks for a specific MB in an I-frame. The public key is then scrambled using a private symmetric key to generate the resultant key which is used for two purposes: 1) To generate the fragile watermark for each GOP. 2) To select the specified I4-block for watermark embedding. The I4-blocks are selected based on this key and a spatiotemporal analysis to further enhance fragility, imperceptibility, and security demands. As we will see later in this section, the intra prediction modes are prone to change when re-encoding is applied; this fact grants the generated content-based key a higher sensitivity, which will cause the decoder to lose synchronization. Moreover, changing some of the intra prediction modes of some blocks leads to different residuals, and hence makes the embedded watermark unachievable. The watermark embedding in I4-blocks meets the demand of Human Visual System (HVS), in which human eyes are less sensitive to noises in edge and detail regions rather than in smooth areas. The security of the algorithm is granted by using random I4-block selection based on the generated content-based key for each MB.

Unlike some previous video coding standards (namely Motion JPEG and MPEG-2), the intra prediction technique is employed in H.264/AVC video coding. Intra prediction is performed within each I-frame with blocks of two different sizes: 16×16 denoted as I16-block and 4×4 denoted as I4-block. The I4-block is best suited for coding picture regions with significant details (textures), while the I16-block is more suited for coding smooth regions. In an I4-block, a prediction is based on the surrounding, previously coded and reconstructed blocks. The I4-block has nine prediction modes, eight are directional (mode=0, 1, 3, .., 8), and one is directionless (DC mode=2). Similarly, the I16-block has four prediction modes, three of them are directional (mode=0, 1, 3), and one is directionless (DC mode=2) [29]. For Inter MBs, on the other hand, variable block-size motion compensation is used to obtain residual information. The supported sizes include 16×16, 16×8, 8×16, and 8×8, in which the 8×8 partition can be further divided into 8×4, 4×8, or 4×4 blocks.

The H.264/AVC standard is a lossy compression, and thus the process of re-encoding a video sequence produces another video sequence which is similar to the original one but not exactly identical. Specifically, the intra prediction modes are vulnerable to change when re-encoding is applied. We take advantage from this point to estimate the I4-blocks with the highest possibility that an intra prediction mode change may happen. Thus, exploiting these I4-blocks for watermark embedding can yield higher sensitivity to aware attacks. To achieve high sensitivity while preserving the essential authentication demands such as fragility, imperceptibility and bit-rate control, we analyze the syntactic elements of the H.264 bitstreams as detailed below.

2.1 Spatial analysis

By utilizing the advantages of the compressed domain watermarking, we analyze the syntactic elements of the Network Application Layer (NAL) units. Watermark embedding must be applied to blocks with high details (texture) because the human eyes are less sensitive to noise in edge and detail regions rather than smooth areas. In [30], the authors utilized the number of nonzero quantized ac coefficients in H.264 to estimate the spatial activity for a given block. Specifically, the more nonzero quantized coefficients indicate the higher possibility of spatial details in the corresponding block.

As mentioned above, the luma intra prediction modes are vulnerable to change when re-encoding is applied. To demonstrate this effect, we applied re-encoding to several standard non-watermarked video sequences, and then estimated the rate of changes of intra prediction modes of I4-blocks and I16-blocks with different numbers of nonzero quantized residuals. Figure 2 shows the results for rate of changes drawn from 100 frames using six different sequences: (*Mobile*, *Silent*, *Tempete*, *Table*, *Container*, and *Salesman*) which were compressed using *QP*=18 and *QP*=28.

As it can be seen, for blocks containing the less Number of NonZero (NNZ) quantized residuals, yields more changes in prediction mode. Hence, embedding watermark in MBs with a lower value of NNZ normally can ensure more probability that a de-synchronization of the watermark decoding may happen. Yet, the number of NNZ quantized residuals must be constrained to a lower bound to maintain the imperceptibility demand as is discussed later in the paper.

Knowing that the numbers of NNZ quantized residuals vary across video sequences based on the spatial characteristics, in this case a threshold namely ω is used to select the more suitable sets of blocks for watermark embedding. It is inappropriate to select a constant threshold since the distribution of NNZ coefficients varies from one sequence to sequence. Figure 3 shows the distributions of NNZ (within I-frame) for different sequences with *QP*=28. It is clear that sequences of more details and textures, such as *Mobile* and *Tempete* where the MBs contain greater values of NNZ, the corresponding distributions are skewed to the right, while for a smooth sequence, such as *Silent*, the maximum numbers of I4-blocks are lying within the area with lower number of nonzero quantized residuals. Therefore, the appropriate threshold should be selected according to the spatial activity of each sequence.

Concerning the watermarking for authentication demands such as fragility, perceptual quality, and capacity, a percentage of I4-blocks, namely σ, containing a required value of NNZ is selected for watermark embedding. A higher value in σ, leads to an increase in embedding capacity while

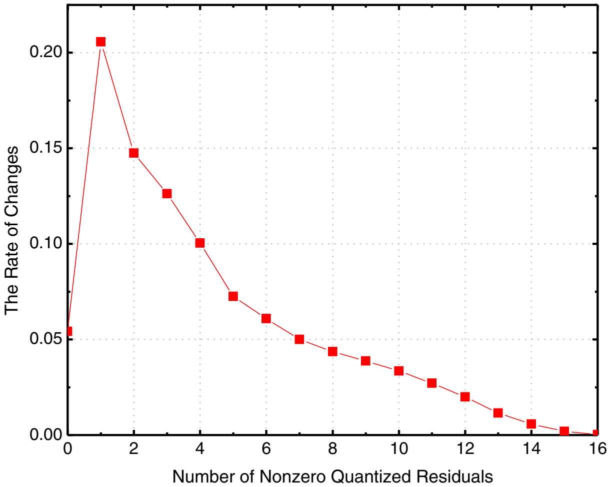

Fig. 2 The rate of changes of I4-block and I16-block intra prediction modes of different sequences with *QP*=18 and 28

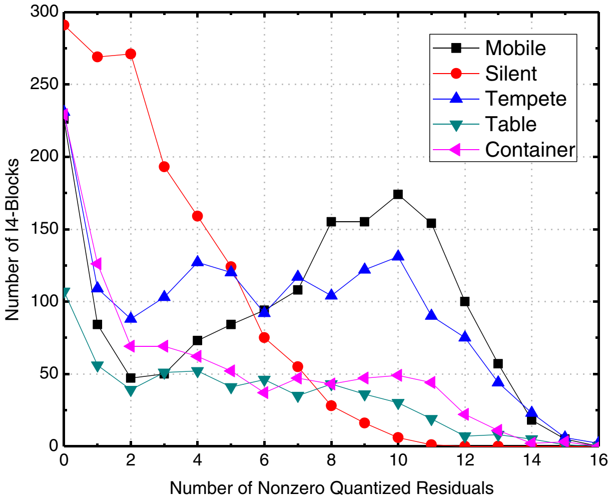

Fig. 3 The distribution of I4-blocks for different nonzero values in first I-frame of different sequences with *QP*=28

the perceptual quality and fragility will decrease. To obtain a target value of ω, the Cumulative Distribution Function (CDF) of NNZ distribution F_{NNZ} is utilized as follows:

$$F_{\mathrm{NNZ}}(\omega)\leq\sigma \tag{1}$$

The CDF of NNZ distribution F_{NNZ}, is defined as below:

$$F_{\mathrm{NNZ}}(\omega) = P(\mathrm{NNZ}\leq\omega) = \sum_{\mathrm{NNZ}\leq\omega} p(\mathrm{NNZ}) \tag{2}$$

Figure 4 depicts the F_{NNZ} of the five different sequences with *QP*=28. As it can be seen, for the same σ, different values of ω are achieved depending on the spatial activity of the

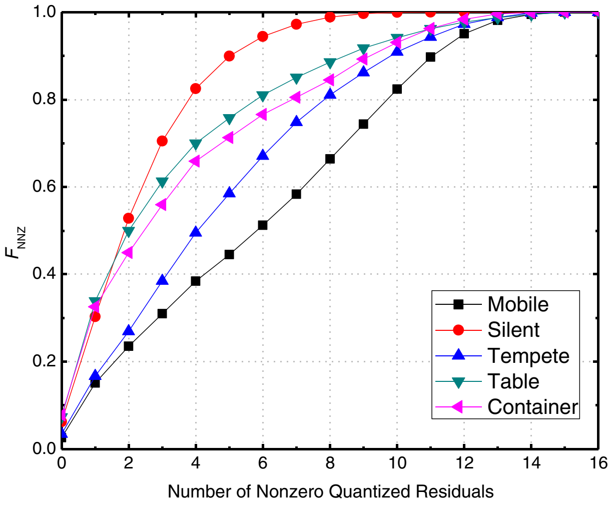

Fig. 4 The F_{NNZ} for different nonzero values of different sequences with *QP*=28

sequences. For instance, when σ=0.5, the value of ω for *Mobile* and *Table* is 6 and 2, respectively. Thus, for each sequence the selected value of ω based on σ, ensures the best number of I4-blocks required for watermark embedding with the highest sensitivity while maintaining better perceptual quality.

In the temporal domain, another perceptual quality enhancement can be explored, since embedding watermarks in flat moving objects may induce some artifacts. To preserve a better perceptual quality, low complexity motion estimation is adopted for MBs according to the corresponding Motion Vectors (MV) as detailed in the following section.

2.2 Temporal analysis

We estimate the motion activity of a video sequence by exploiting the motion vectors of inter MBs for two reasons: 1) In the compressed domain watermarking, the motion vectors information is readily available and easy to decode from the bitstream and there is no need for further decoding and re-encoding. 2) The motion vectors represent the essential motion information for each I4-block. As a result, accessing the motion vectors provides representative information about the motion activity for the smallest block (I4-block) of interest. Hence, a low complexity and accurate temporal analysis can be achieved.

To extract the motion activity of a video sequence, the temporal activity represented by the motion activity of each I4-block is estimated by computing the mean of the corresponding motion vectors in P-frames of the previous GOP. In this case, an array called I4-block Motion (I4M) is extracted by evaluating the average of the motion vectors of the corresponding blocks in P-frames of the previous GOP. This information can be calculated by:

$$\mathrm{I4M} = \frac{\sum\limits_{k\in \mathrm{P}reviousGOP} Mmv_i}{Num(mv_i)} \tag{3}$$

where

$$Mmv_i = \sqrt{mvh_i{}^2 + mvv_i{}^2}$$

where mvh_i and mvv_i denote the horizontal and vertical components of the motion vector of i-th I4-block, respectively; $Num(mv_i)$ denotes the number of the motion vectors considered. For each P-frame of size of 176×144, another matrix called Frame Motion Activity (FMA) of size of 44×36 is achieved, in which each unit shows the value of I4M for each I4-block in that frame.

To estimate the I4-block Normalized Motion Activity (I4NMA), the I4M and FMA information related to the previous GOP frames are exploited to construct the GOP Motion Activity (GMA), and which is then used to construct the I4NMA, as defined below:

$$\mathrm{I4NMA} = \frac{\mathrm{I4M}}{\mathrm{GMA}} \tag{4}$$

where

$$\mathrm{GMA} = \bigcup_{k\in \mathrm{P}reviousGOP} \mathrm{mean}(\mathrm{FMA}_k)$$

In this equation, k denotes the index of a P-frame in the previous GOP and mean (FMA$_k$) is a scalar value which represents the average of the FMA matrix entries for all frames. For a given I4-block, it is considered as Fast Moving Blocks (FMB) if the corresponding I4NMA obtained in Eq. (4) is greater than one. As a result, such FMBs are considered for watermark embedding, since FMBs are good candidates for yielding less visual artifact when I4NMA is employed as a qualifying factor. Thus, some blocks are skipped for embedding if the corresponding I4NMA values are less than one since embedding the watermark into non-active areas leads to noticeable temporal artifacts.

2.3 Fragile watermark generation

To authenticate every GOP in the H.264/AVC stream, each GOP is processed individually. This approach enables us to detect the attacked scenes more precisely. To that end, the encrypted MAC value is embedded in each GOP. In the proposed GOP-based authentication scheme, two order sequences are introduced. The GOP Order Sequence (GOS) represents the GOP order sequence in the watermarked sequence, and the Frame Order Sequence (FOS) represents the current frame order in the watermarked sequence. The GOS and FOS are responsible for detecting GOP-based and Frame-based attacks, respectively.

The watermark generation process decodes the NAL syntactic elements to extract the intra/inter MBs and collect the intra I4-block and I16-block prediction modes (IV1), inter prediction modes (IV2), FOS, and GOS into separate buffers, then the collected buffers are XOR-ed and the product is then treated by a one-way hash cryptographic function PJW Hash [3] denoted as H(•). To enhance the security, the generated digest is then encrypted using the aforementioned content-based key (*K*) in the form of

$$W_G = \mathrm{E}(\mathrm{H}(IV1 \oplus IV2 \oplus FOS \oplus GOS), K) \tag{5}$$

where E(•) denotes a low complexity encryption function which scrambles its inputs based on the content-based key *(K)*, ⊕ denotes the XOR logical operator. Finally, this digest acts as the fragile watermark to be embedded into a set of selected intra luminance I4-blocks. The details of watermark embedding are given in the following section.

2.4 Watermark embedding

In the literature, several embedding techniques are found, for example, Spread Spectrum (SS) [7], Least Significant Bits (LSB) [10] and Quantization Index Modulation (QIM) [6]. Spread Spectrum watermarking and Quantization Index Modulation techniques usually support robust watermarking schemes, while the LSB modulation techniques are better suited for the case of authentication schemes. On one hand, LSB is cheaper in terms of computational cost; on the other hand, it is better to aware attacks, and thus it is adopted for watermark embedding in this study.

The watermark payload is the secret Message Authentication Code (MAC) created for each GOP which was generated from the previous section in the form of binary sequence denoted as $W_G=\{w_i \mid \mathrm{i}=0, 1, \ldots, \mathrm{M}, w_i \quad \{0,1\}\}$, where M is the watermark length. The cover payload is the nonzero quantized ac residuals of I4-blocks of I-frames in each GOP. A Host Block (*HB*) denotes the I4-block to embed watermark w_i, the I4-block (*X*) must be a Candidate Block (*CB*), which means the I4-block (*X*) belongs to a set *S* of pseudo randomly selected blocks based on the generated content-based key (*K*).

Regarding the watermark embedding process demands, two constrains, fragility threshold Tr_f *and* quality threshold Tr_q, are established. To prevent quality degradation, the embedding is applied on the I4-blocks which meet the following condition:

$$\begin{array}{l} \text{if } \ Tr_q \leq \mathrm{NNZ}(X) \leq Tr_f \\ \text{where } \ Tr_f = \omega + \phi \end{array} \tag{6}$$

where NNZ(•) indicates the number of nonzero quantized ac coefficients in a selected I4-block, and the value of the threshold Tr_q is application-dependant. The threshold Tr_f is introduced to enhance the watermark fragility. The parameter ω is derived based on the spatial activity of the sequence according to (1) and (2), and the parameter ϕ has to be selected in such a way that the fragility threshold Tr_f does not exceed the maximum number of nonzero residuals. Consequently, the embedding is restricted to a set of I4-blocks which are more sensitive to re-encoding and other signal processing manipulations while maintaining high perceptual quality. The watermark embedding algorithm is organized as below:

Step 1: Parse the current H.264/AVC stream to construct the MBs structure for the first GOP (G_i).
Step 2: Apply the spatiotemporal analysis for G_i.
Step 3: Call the fragile watermark generation algorithm to construct the watermark information W_G for G_i.
Step 4: If the current frame is I-frame, for each MB, if the current I4-block (*X*) is a *CB* and which meets the condition in Eq. 6, then Eq. 7 is applied according to the watermark bit w_i to modulate the last nonzero quantized ac residual.

$$ac'_i \begin{cases} ac_i & \text{if } w_i = 1 \text{ and } |ac_i| \bmod 2 = 1 \\ ac_i - 1 & \text{if } w_i = 1 \text{ and } |ac_i| \bmod 2 = 0 \\ ac_i + 1 & \text{if } w_i = 0 \text{ and } |ac_i| \bmod 2 = 1 \\ ac_i & \text{if } w_i = 0 \text{ and } |ac_i| \bmod 2 = 0 \end{cases} \tag{7}$$

where | • | denotes the absolute value function; ac_i and ac'_i denote the original and watermarked nonzero quantized ac residuals, respectively.
Step 5: Entropy re-encode the modified MBs and record them back to the slice unit.
Step 6: Repeat Steps 1–5 until all GOPs are watermarked.

From Eq. 7, the maximum change made to the selected quantized residuals for watermark embedding is equal to one. If the watermark information *W*and the quantized residuals are uniformly distributed, then 50 % of the embedded bits will not affect the quantized residuals. Thus, the amount of distortion incurred is minuscule and does not degrade the visual quality of the frame.

2.5 Watermark extraction and verification

If a H.264/AVC stream receiver suspects that the video stream was tampered with or intentionally modified for any reason, the watermark extraction and verification algorithm can be applied to confirm the authenticity and integrity. Watermark extraction is performed after entropy decoding. The extraction process is the inverse of the embedding process. The main steps of the watermark extraction and verification are as follows:

Step 1: Partially decode the watermarked H.264/AVC video stream to construct the GOP structures including intra/inter prediction modes, motion vectors, and quantized residuals of all of the I4-blocks.

Step 2: Apply the spatiotemporal analysis for G_i.

Step 3: Call the fragile watermark generation to construct the encrypted and hashed embedded watermark information W_G for the current GOP.

Step 4: If the current frame is I-frame and current I4-block is a *CB* which meets the condition defined in Eq. 6, then extract the embedded watermark from the last nonzero quantized ac residual to construct the extracted watermark information W'_G. The watermark bit w_i' is determined as

$$w_i' = \begin{cases} 1, & \text{if } \left|ac'_i\right| \bmod 2 = 1 \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

Step 5: Compare the two sets of extracted watermarks W_G and W'_G. If they are identical, then G_i is verified and authenticated.

Step 6: Repeat Steps 2–5 for all GOPs. If all of the extracted watermarks, W_G and W'_G are identical, then the H.264/AVC stream is verified and authenticated.

Evidently, the extraction process is simple and fast, because the hidden authentication information can be detected solely from the last nonzero ac residuals and the easily accessible intra/inter prediction modes, and the motion vectors. Consequently, partial decoding is available with the proposed scheme which yields an advantage in fast video authentication scenario. In particular, the video data are normally bulky generated from most application such as video surveillance system.

3 Experimental results and discussions

The proposed watermarking scheme was implemented using the H.264/AVC JM10.2 of the reference software [28]. To confirm the effectiveness of the proposed watermarking scheme, different standard video sequences (*Foreman*, *Grandma*, *Carphone*, *Container*, *Claire*, *Mother*, *Bus, Salesman, Table*, *Soccer*, *Tempete*, *Akiyo*, *Stefan*, *Silent*, *News*, and *Mobile*) of QCIF format (176×144) at rate 30 frames/s were tested. The selected video sequences include the low to high spatial detail and low to high amount of movement activities. Specifically, the QCIF (YUV 4:2:0) was selected for its common resolution in mobile and low bit-rate applications. The GOP structure consists of an I-frame followed by 4 P-frames in the Main profile and the CABAC entropy coding.

In this study, the variation on bit-rate (VAR_{RATE}) and *PSNR* (VAR_{YPSNR}) as defined in Eqs. 9 and 10 are employed for objective comparisons.

$$VAR_{RATE} = \frac{R' - R}{R} \times 100 \tag{9}$$

where *R and R'* denote the bit-rate of the original and watermarked bitstreams, respectively.

$$VAR_{YPSNR} = YPSNR - YPSNR' \tag{10}$$

where *YPSNR* and *YPSNR'* denote the average *PSNR* of the luma (Y) samples in all frames of the original and watermarked bitstreams, respectively.

Since *PSNR* does not consider the temporal activity of the encoded bitstreams, the Video Quality Metric (VQM) [4] is employed in this study. The value of the VQM lies in between zero and one, where one and zero indicate maximum impairment and the best quality scenario, respectively.

Since the bit-rate depends on the embedded capacity, in order to perform fair comparisons, the watermark cost (δ) defined in [25] is employed to denote the increase in the number of bits used to encode the watermarked video per watermark bit:

$$\delta = \frac{R' - R}{Cp} \tag{11}$$

where Cp denotes the payload capacity.

Regarding the relationships among fragility, imperceptibility, bit-rate, and payload capacity, where a higher payload capacity normally implies higher visual quality degradation and higher bit-rate increment, thus a tradeoff between these paradoxical factors was adapted in this study. Subsequently, extensive experiments have been conducted to determine the optimum values of the threshold parameters σ, Tr_q, and Tr_f. On one hand, increasing the value of σ will increase the embedded payload on the cost of inducing distortion and the increase in bit-rate. On the other hand, increasing Tr_q leads to positive effect on visual quality while decreasing the embedded payload. To efficiently estimate the control parameters σ, Tr_f, and Tr_q, several experiments were conducted over 100 frames of five standard sequences, including *Mobile*, *Container*, *Table*, *Tempete*, and *Silent*, using the proposed method by setting QP=28. In the experiments, four values were tested for the quality threshold Tr_q ranging from 1 to 4. This range has been carefully selected, since in the H.264/AVC coding, statistics shows that on average 60 % of transform coefficients of the prediction residue in one MB are quantized to zero. In addition, three values of σ were tested, including 0.7, 0.6, and 0.5. For simplicity, the fragility threshold Tr_f was fixed at ω+1 by setting ϕ=1. Thus, the averaged results of 60 experiments are reported in Fig. 5. As it is shown in Fig. 5a, a greater value in σ, the higher capacity can be obtained, and correspondingly the higher distortion induced as in Fig. 5c; similarly, higher fragility—lower Normalized Correlation Coefficient (*NCC*) was obtained as depicted in Fig. 5d. In the meantime, Fig. 5b shows an exponential trend, where a lower watermark cost (δ) is achieved when Tr_q=4 for all values of σ. Hence, by setting σ to 0.6 and Tr_q to 4, a tradeoff between these conflicting factors is obtained. Thus, the lowest effect on video coding efficiency, and the highest fragility are assured.

3.1 Imperceptibility test

To evaluate the imperceptibility of the proposed scheme, a series of experiments have been performed. Fig. 6a, e and b, f illustrate the original and watermarked second intra coded frame of *Tempete* and *Foreman* sequences, respectively. Similarly, Fig. 6c, g and d, h show the original and watermarked first inter coded frame (in the 2nd GOP) of *Tempete* and *Foreman* sequences, respectively. It is clear that no significant difference of subjective visual quality is found between the original and watermarked frames. Moreover, in the carried out experiments, no visible artifacts can be observed in all of the test video sequences. This can be clearly noticed from the subjective evaluation of the subsequent P-frames in Fig. 6c, g and d, h since no propagated flickering is noticed at all.

For objective evaluation, Fig. 7 shows the frame-by-frame *YPSNR* of the luma samples of the original and watermarked *Tempete* and *Table* sequences. As it can be seen, the VAR_{YPSNR} does not exceed 0.05 (dB) in average for all of the intra frames, which proves that the proposed scheme can maintain the visual quality of the watermarked bitstreams.

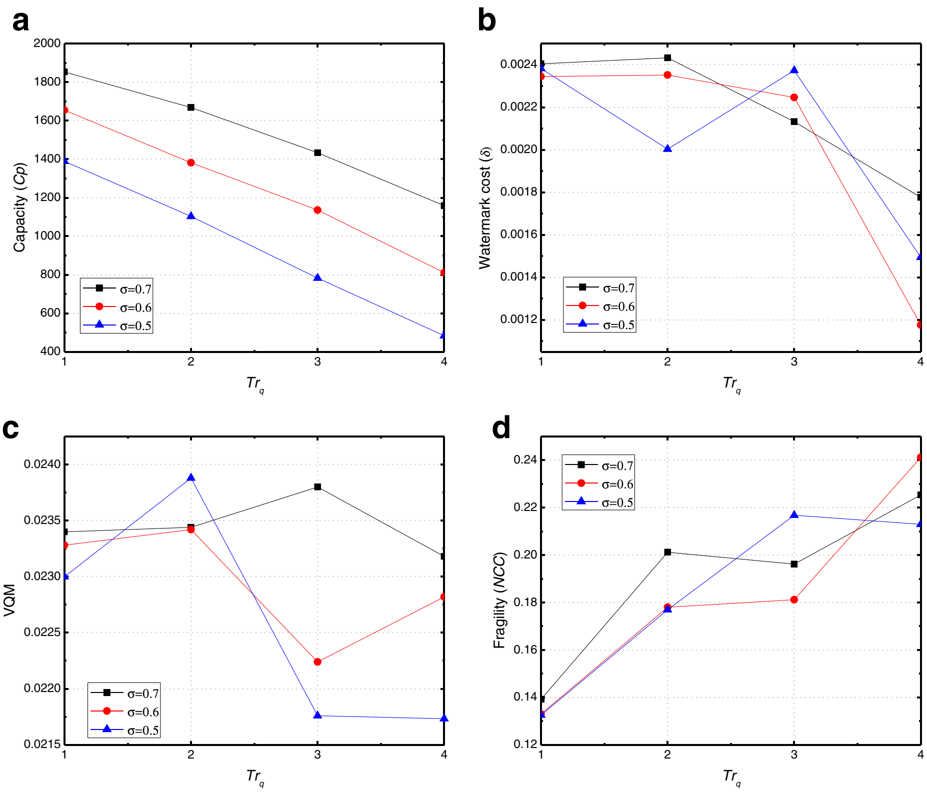

Fig. 5 The results of average variation on **a** Capacity (*Cp*), **b** Watermark cost (*δ*), **c** Video Quality Metric (VQM), and **d** Fragility (*NCC*) of different sequences with multiple values of *σ* (0.7, 0.6. and 0.5) and Tr_q (1, 2, 3 and 4)

Figure 8 shows the *YPSNR* using the proposed watermarking scheme at several values of *QP* using multiple values of Tr_q ranging from 1 to 4. Apparently, we can see almost no difference between the original and watermarked sequence, thus, the proposed scheme is suitable for applications with different fidelities. Notably, the lowest difference is obtained when Tr_q=4.

Figure 9 shows the Rate-distortion curve for the proposed scheme at multiple values of Tr_q ranging from 1 to 4. It appears that the proposed scheme does not reduce the perceptual quality but kept almost the same of the original codec. Thus, the proposed scheme can maintain the perceptual quality under various bit-rates.

Table 1 presents a comparison of the visual quality variation VAR_{YPSNR} (dB) using different video sequences of length 120 frames between the proposed scheme and the former method [32] with *QP*=28. The results show that the proposed scheme outperforms the former scheme for all video sequences. This inconsiderable perceptual distortion of the proposed method can be explained due to the watermark embedding in the last nonzero ac residuals, i.e. the high frequency region of the DCT spectrum.

Here below, the Structural SIMilarity index (SSIM) [31] is employed to compare the proposed method in terms of temporal perceptual quality with the former semi-fragile method [36]. The value of the SSIM lies in between zero and one, where zero and one indicate maximum impairment and the best quality scenario, respectively. In

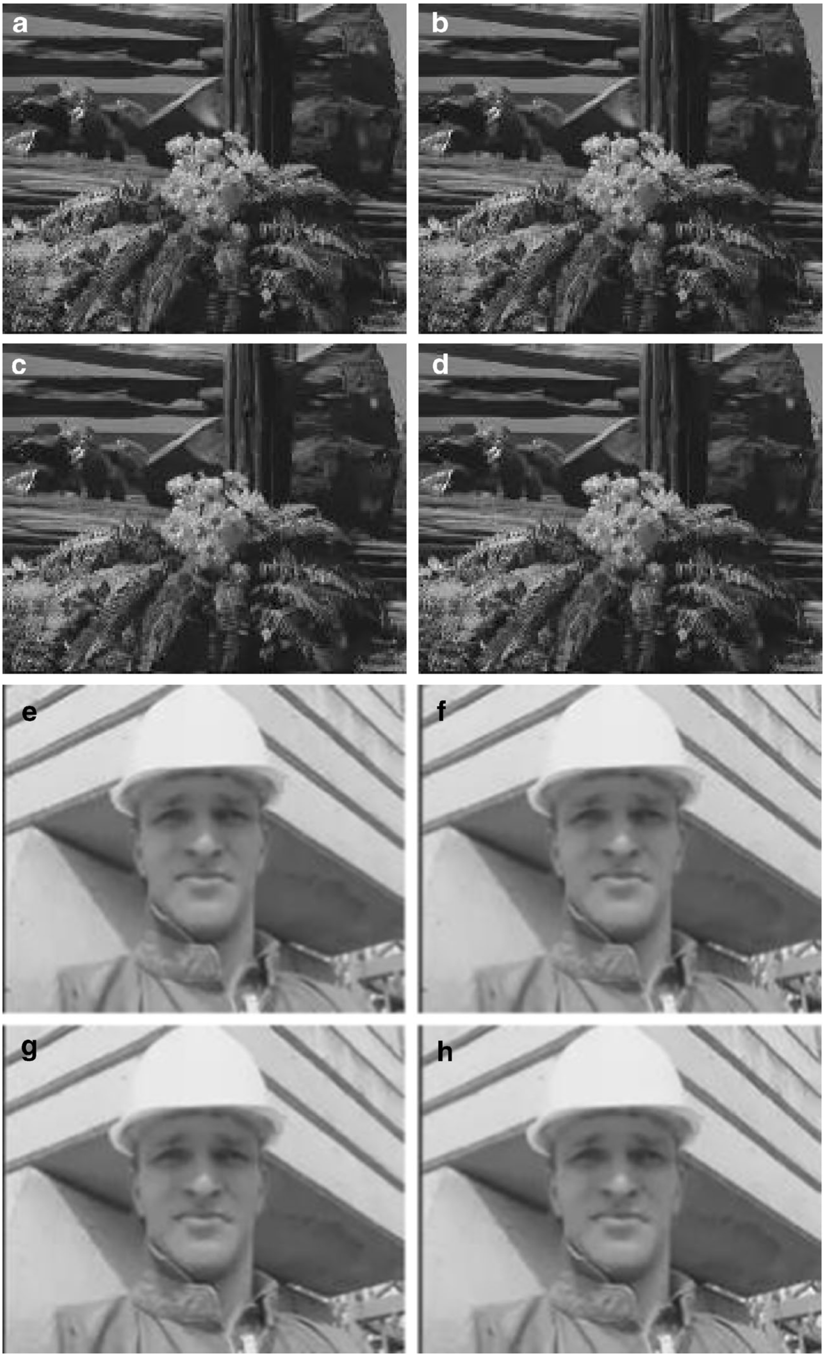

Fig. 6 Visual quality evaluation of the proposed scheme for *Tempete* and *Foreman* with *QP*=28 **a** Original I-frame (*YPSNR*=35.25 dB) and **b** Watermarked I-frame (*YPSNR'*=35.28 dB) **c** Original P-frame (*YPSNR*=33.82 dB) and **d** Watermarked P-frame (*YPSNR'*=33.78 dB) **e** Original I-frame (*YPSNR*=36.78 dB) and **f** Watermarked I-frame (*YPSNR'*=36.77 dB) **g** Original P-frame (*YPSNR*=36.17 dB) and **h** Watermarked P-frame (*YPSNR'*=36.15 dB)

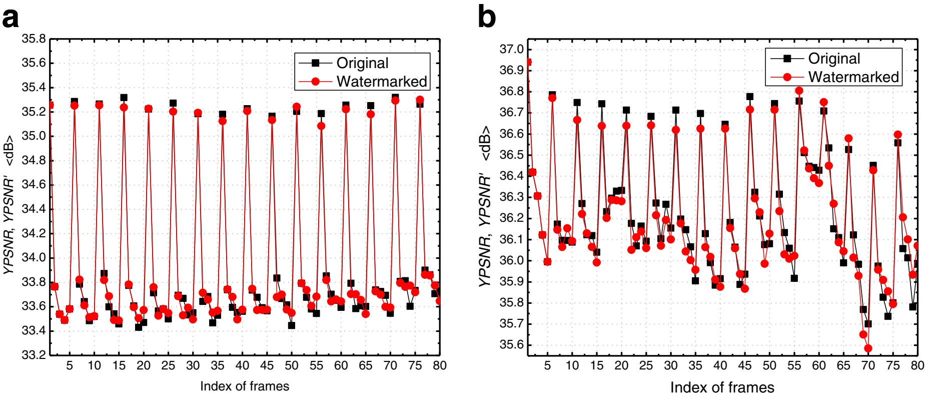

Fig. 7 Frame-by-frame *YPSNR, YPSNR′* of the original and watermarked **a** *Tempete* and **b** *Foreman* with *QP*=28

this test, four QCIF videos of length 150 frames were used. The test was performed using the GOP structure “IPPPPPPPPPI” and with *QP*=28. Figure 10 shows the corresponding SSIM results. Obviously, the proposed method outperforms the other method in most cases, and the average SSIM of the proposed method is above 0.98 for all the tested sequences. The prominent improvement is achieved by the proposed spatiotemporal analysis.

Finally, the VQM is employed to compare the proposed method in terms of temporal perceptual quality with the former methods [24, 33] and [25]. In this test, eight QCIF videos were used. The test was performed using the GOP structure “IBPBPBI” and with *QP*=28. Table 2 shows the corresponding VQM of these methods when only I-frames are watermarked. Obviously, the proposed method outperforms the other methods, and the average VQM of the proposed method is about five times less than the other methods. The prominent improvement is achieved by the proposed spatiotemporal analysis.

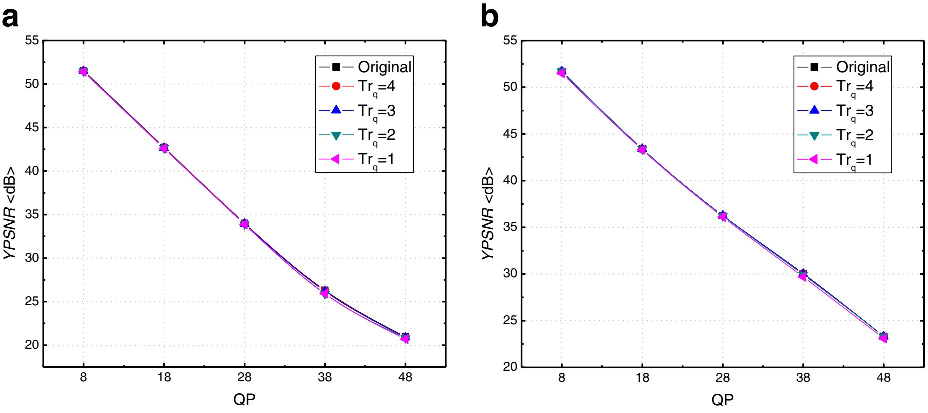

Fig. 8 **a** *Tempete* and **b** *Foreman YPSNR curves* at constant *QP* with Tr_q=(1, 2, 3 and 4)

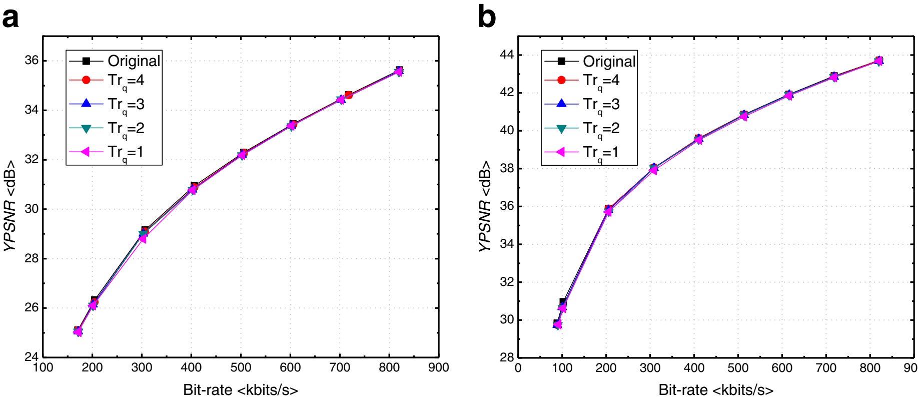

Fig. 9 Rate-distortion curve for **a** *Tempete* and **b** *Foreman* with Tr_q=(1, 2, 3 and 4)

3.2 Bit-rate and payload capacity test

To show the limited bit-rate increment incurred using the proposed watermarking scheme, several experiments have been performed using various video sequences with *QP*=28.

Table 3 shows the minuscule effect of the proposed watermarking scheme on bit-rate. Obviously, we can see the acceptable bit-rate increases with reasonable payload capacities. The tiny bit-rate increase (0.50 in average) is achieved because the embedding is limited on the last nonzero quantized ac residuals.

To demonstrate that the proposed method preserves the coding efficiency, a comparison in term of VAR_{RATE} with the method in [36] is performed. Figure 11 shows the minuscule effect of the proposed watermarking scheme on bit-rate. Obviously, we can see a significant reduction on the final bit-rate of 25 times lower in average. This can be explained by the novel anticipated watermark embedding in which only the last nonzero quantized ac residuals are modified.

A further comparison with the DCT-based methods [24, 33], and [25] have been performed based on the watermark cost δ. This test was performed using the GOP structure "IBPBPBI" and *QP*=28. Table 4 shows the comparison in terms of the watermark cost δ using various video sequences between the proposed method and the former methods [24, 33], and [25]. Obviously, the proposed method is superior to these schemes with the test video sequences.

3.3 Fragility to tampering test

As mentioned above, the objective of this study is to propose a hard authentication scheme which is able to detect any content-preserving and/or content-changing attacks. In general, in a fragile watermarking, an attacker would preferably want to attack the watermarked video

Table 1 Performance comparison in terms of visual quality variation VAR_{YPSNR} (dB) with former method [32]

	Foreman	*Akiyo*	*Stefan*	*Bus*	*Tempete*	*Mobile*	(Average)
Intra watermarking method [32]	−2.03	−2	−1.61	−1.53	−1.04	−0.64	−1.475
Proposed watermarking scheme	−0.01	0.01	0	−0.02	0	−0.01	−0.005

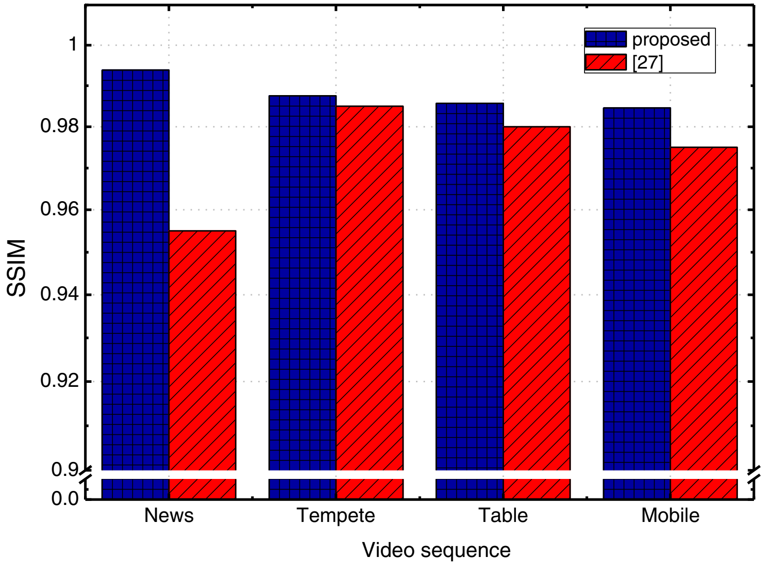

Fig. 10 SSIM for the proposed method and the method in [36]

stream in such a way that the watermark is completely destroyed, yet no perceivable degradation in quality is found in the attacked video. To evaluate the fragility sensitivity of the proposed scheme against such attacks, the watermarked streams, *Foreman*, *Grandma*, *Carphone*, *Container*, *Tempete*, and *Mobile*, were subjected to a three groups of simulated attacks: 1) Group 1 belongs to the content-preserving attacks, including re-encoding, rate control (50 Kbps) and transcoding (*QP*=32). 2) Group 2 belongs to the common signal processing attacks, including median filtering (3×3), gaussian blurring (2.5×2.5), cropping (170×140) from bottom-right, and rotation (1°) attacks. 3) Group 3 includes conventional GOP-based and Frame-based attacks. The simulated attacks in Group 1, 2 and 3 vary in strength from weak, medium, to strong, respectively.

Table 2 VQM for the proposed watermarking method and the methods in [24, 33] and [25] when watermark is embedded in I-frames

Video sequence	No. of frames	VQM Intra watermarking			
		[33]	[24]	[25]	Proposed method
Carphone	150	0.120	0.130	0.130	0.019
Claire	200	0.100	0.100	0.100	0.016
Mobile	100	0.050	0.050	0.060	0.017
Mother	100	0.150	0.150	0.140	0.021
Salesman	150	0.140	0.140	0.150	0.025
Soccer	60	0.080	0.080	0.090	0.027
Table	100	0.150	0.160	0.160	0.028
Tempete	60	0.080	0.070	0.080	0.019
(Average)	–	0.110	0.110	0.110	0.022

Table 3 Payload capacity (*CP*) and bit-rate variation (VAR_{RATE})

Video sequence	No. of frames	VAR_{RATE}	Payload capacity (*Cp*)
Carphone	150	1.03	973
Claire	200	0.79	499
Mobile	100	0.49	1,041
Mother	100	0.13	549
Salesman	150	0.53	852
Soccer	60	0.55	390
Table	100	0.13	371
Tempete	60	0.34	432
(Average)	–	0.50	638

Regarding the content-preserving attacks in Group 1, they are considered as the most common attacks in the H.264/AVC bitstream domain. The attacker aims to produce another version similar but not identical to the encoded sequence, aiming to remove the embedded watermark without damaging the video stream. While the median filtering, gaussian blurring, and geometric attacks are designed to test the capability of detecting unintentional attacks such as noisy channels causing high bit-error rates. The GOP-based and Frame-based attacks represent a set of strong attacks to simulate intentional ruin attacks.

In the experiments, the Normalized Correlation Coefficient (*NCC*) is employed to measure the similarity between the embedded and detected watermark bits. *NCC* ranges from 1 to −1, where 1 indicate perfect watermark match while −1 indicates totally watermark mismatch. Table 5 shows the performance of the fragile watermarking scheme under these types of attacks in terms of the *YPSNR′* using the attacked 100 frames of *Tempete* sequence. From the detected *NCC* and *YPSNR′* metrics of the simulated attacks, we can observe that

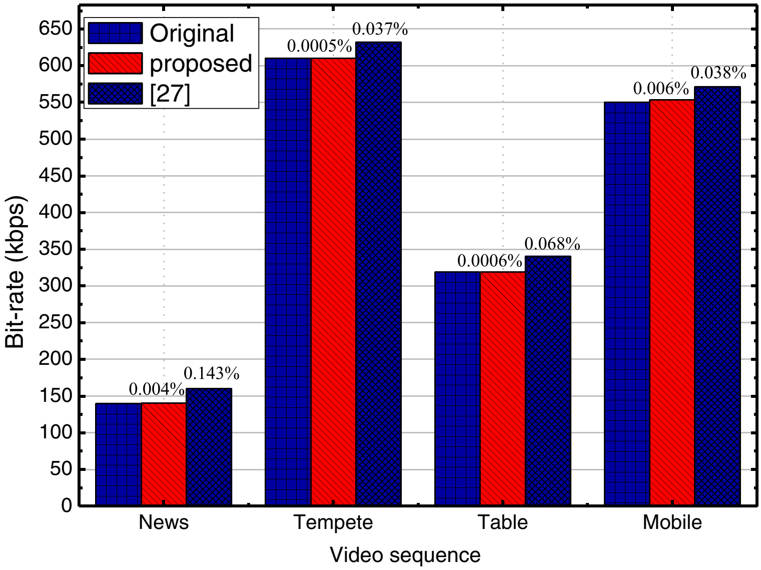

Fig. 11 The VAR_{RATE} for the proposed watermarking method and the method in [36]

Table 4 Comparison of the proposed watermarking method with methods in [24, 33] and [25] in terms of the δ when watermark is embedded in I-frames

Video sequence	No. of frames	Watermark cost (δ)			
		[33]	[24]	[25]	Proposed method
Carphone	150	1.28	1.56	0.41	1.60×10^{-3}
Claire	200	2.62	2.70	0.60	1.22×10^{-3}
Mobile	100	0.77	1.08	0.43	2.47×10^{-3}
Mother	100	2.31	2.59	0.89	2.55×10^{-4}
Salesman	150	1.00	1.10	0.31	1.13×10^{-3}
Soccer	60	1.34	1.40	0.31	4.18×10^{-3}
Table	100	0.58	0.96	0.21	1.27×10^{-3}
tempete	60	0.99	1.03	0.52	4.21×10^{-3}
(Average)	–	1.36	1.55	0.46	2.04×10^{-3}

the proposed scheme is very sensitive to attacks which vary from small pixel value changes to strong ruin attacks, since the visual quality of the attacked video was seriously degraded after the attacking process while the *NCC* values remained very low. Thus, the scheme is able to authenticate the watermarked H.264 streams effectively.

Figure 11 shows the *NCC* of six video sequences under the aforementioned attacks. It can be seen that the detected *NCC* values are very low. In Fig. 10a, the Rate control attack shows lower *NCC* values since the rate control mechanism dynamically adjusts the *QP* of the video sequence being encoded depending upon the constraint limit set by the encoder, and thus a higher possibility of intra mode change may happen, which cause change in the generated watermark and the ac residuals. Moreover, it is observed that the results of Groups 2 and 3 attacks show low values of *NCC* [0.15 to −0.20], which is caused by the use of FOS and GOS, since these attacks affects them directly leading to watermark de-synchronization, i.e. authentication failure.

Table 5 Fragility performance of the proposed watermarking scheme for *Tempete* sequence

Group No.	Attack name	*NCC*	VAR_{YPSNR} (dB)
Group 1	Re-encoding	0.295	7.49
	Rate control (50 Kbps)	−0.344	−8.19
	Transcoding (*QP*=32)	−0.007	−2.85
Group 2	Median filtering (3×3)	0.024	−3.62
	Gaussian blurring (2.5×2.5)	−0.05	−3.46
	Cropping (170×140, from bottom-left)	−0.061	−4.24
	Rotation (1° to right)	−0.11	−4.18
Group 3	GOP-cutting (GOP No. 3, 5, 8)	−0.24	−2.62
	GOP-replacement (GOP No. 4 with 5)	−0.22	−2.47
	GOP-insertion (GOP No. 2, 5)	−0.35	−2.11
	Frame-cutting (GOP No. 1 [19, 20], GOP No. 2 [1, 11, 38], GOP No. 4 [24, 25, 27])	−0.24	−2.40
	Frame-replacement (Frame No. 16 with 17)	0.15	−1.60
	Frame-insertion (Frame No. 12, 42, 80)	−0.11	1.52

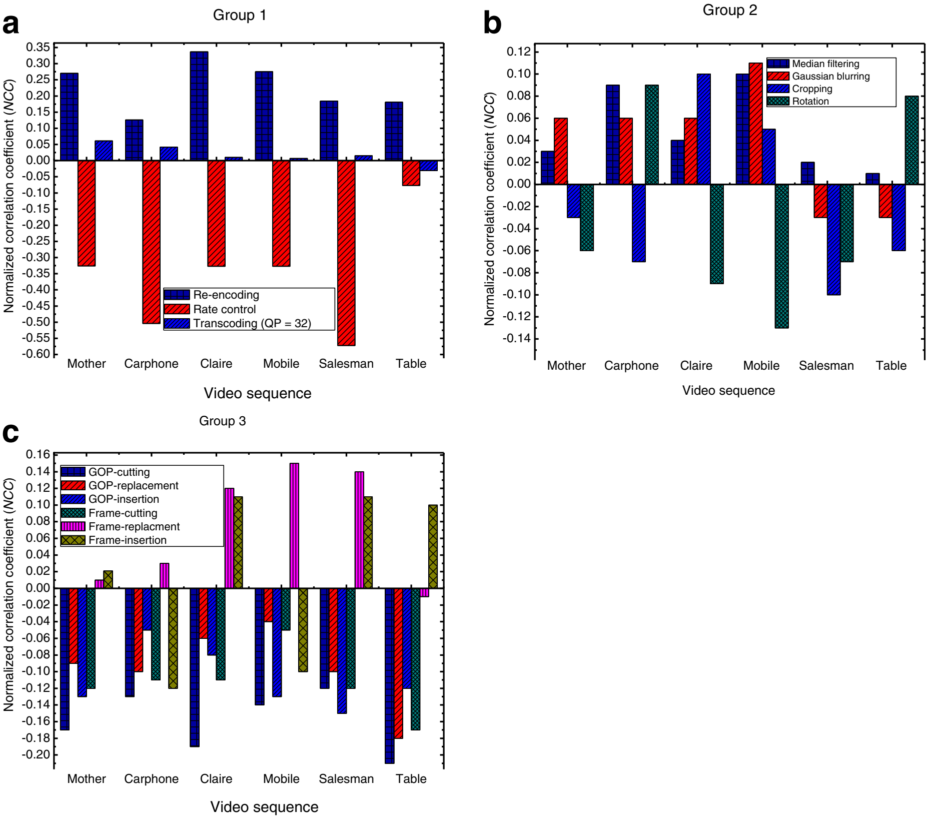

Fig. 12 Fragility performance evaluation under three groups of attacks for different sequences **a** Group 1, content-preserving attacks **b** Group 2, content-changing attacks **c** Group 3, conventional GOP-based and Frame-based attacks

The reason behind authentication failure for the attacks mentioned in this section is that when the frames are tampered with, the hash produced at the decoder is different from that produced by the encoder, as the PJW hash is a one way hash and the probability of yielding the same hash from two different sets of inputs is close to zero. The effectiveness of the proposed method was enhanced by utilizing the sensitivity of the intra/inter prediction modes, and the content-based key (K) which makes the embedded watermark to be collapsed when any manipulations are involved. From the results drawn in Table 5 and Fig. 12, we conclude that the proposed scheme is able to detect any kind of spatial and/or temporal manipulations, meaning that the method is proficient to verify the authenticity of any watermarked H.264 stream.

3.4 Complexity test

The proposed method has a low computational complexity, since the major processes such as the fragile watermark generation, the watermark embedding, and finally the watermark extraction and verification are all composed of simple arithmetic operations. The overhead cost of the watermark embedding is induced by comparing the partial decoding against the partial decoding plus watermark embedding, while the watermark extraction overhead cost is induced by comparing the partial decoding against the partial decoding plus watermark extraction, respectively.

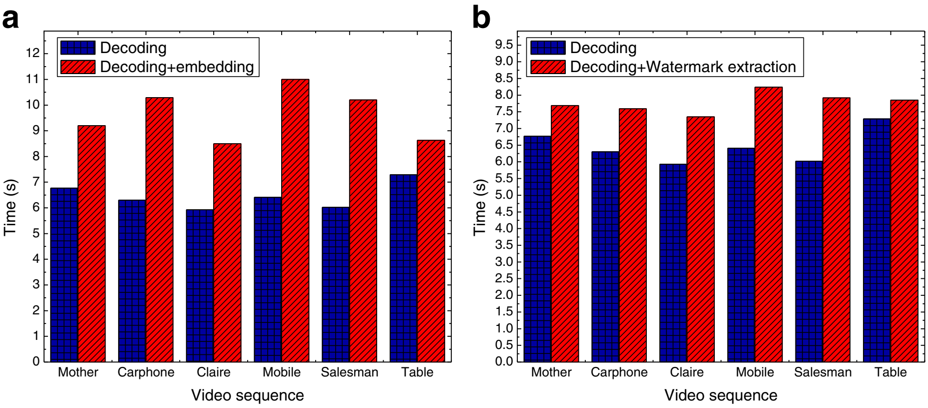

Fig. 13 Results of average time overhead for **a** watermark embedding, and **b** watermark extraction and verification

Figure 13 depicts the average incurred overheads with the modified JM reference software for both encoder and decoder using various QCIF sequences of size 80 frames with QP=28. Apparently, the overhead cost is negligible, since the average delay is no more than 3.18 s. in the embedding, and no more than 1.32 s. in the extraction. Consequently, we conclude that the proposed method can be practically applied to H.264 streams with good efficiency.

4 Conclusions and future works

In this study, a low complexity, blind GOP-based fragile watermarking scheme for authenticating the integrity of H.264/AVC videos was proposed. The use of the digital watermarking to authenticate digital video and the necessity of hard authentication were discussed. The scheme utilizes three major features of the H.264/AVC standard, including intra/inter mode prediction, motion vectors and I4-blocks quantized ac residuals. A low cost spatiotemporal analysis is proposed to lessen the effect of the imposed degradation and bit-rate increase while maintaining a high fragility to tampering. Moreover, the watermark embedding is performed in the compressed domain, and thus no extra overhead is induced.

A content-based key is generated to control fragile watermark generation, watermark embedding, and watermark extraction and verification processes. The secret self-authentication code is embedded into the last nonzero quantized ac residuals of the luma I4-blocks of I-frames in each GOP of the H.264/AVC video. Hidden information can be extracted via partially decoding the watermarked H.264 stream without the need of the original video stream.

Experimental results over several representative video sequences show the scheme has a comparatively high payload capacity with negligible effect on both video quality and coding efficiency. In addition, fragility tests demonstrated high sensitivity against content-preserving and/or content-changing attacks. Finally, the technique exhibits a very low computational complexity as the watermark embedding operation involves simple mathematical operations. This makes it ideal for content-authentication and tamper proofing in low-power handheld devices and real-time mobile applications. The future research can be put of developing a new semi-fragile scheme exploiting the new features of the next generation HEVC standard.

Acknowledgments This work was supported in part by the National Science Council under contract number NSC101-2221-E-011-063-MY3 and it was also partially supported by the 111 Project under the grant No. 111-2-14, One Hundred Talents Program 2012, Sichuan Province. M. E. Farfoura is currently a system analyst with the Information Technology Center- Royal Scientific Society, Jordan-Amman. The work of X. Wang was supported by the National Science Foundation of China under Grants 60902023 and 61171096, as well as by the Fundamental Research Funds for the Central Universities under the Grant SWJTU12CX093.

References

1. (2003) ISO/IEC 14496–10 and ITU-T Rec. H.264, Advanced video coding
2. Ahmad I, Wei X, Sun Y, Zhang Y (2005) Video transcoding: an overview of various techniques and research issues. IEEE Trans Multimed 7(5):793–804
3. Aho A, Sethi R, Ullman J (1986) Compilers: principles, techniques, and tools. Addison-Wesley, pp. 434–438
4. American National Standard for Telecommunications-Digital Transport of One-Way Video Signals-Parameters for Objective Performance Assessment, Standard T1.801.03-003, Jul. 2003, vol. ANSI
5. Cappellini V, Bartolini F, Caldelli R, De Rosa A, Piva A, Bami A (2001) "Robust frame-based watermarking for digital video," Proceedings 12th International Workshop on Database and Expert Systems Applications, pp. 825–829
6. Chen B, Wornell GW (2001) Quantization index modulation: a class of provably good methods for digital watermarking and information embedding. IEEE Trans Inf Theory 47(4):1423–1443
7. Cox I, Kilian J, Leighton FT, Shamoon T (1997) Secure spread spectrum watermarking for multimedia. IEEE Trans Image Process 6(12):1673–1687
8. Furht B, Marques O (2003) Handbook of video databases: design and applications. CRC Press, Boca Raton
9. Golikeri A, Nasiopoulos P, Wang ZJ (2007) Robust digital video watermarking scheme for H.264 advanced video coding standard. J Electr Imaging 16(4):043008
10. Gonzalez RC, Woods RE (2008) Digital image processing, 3rd edn. Prentice-Hall, Englewood Cliffs
11. Horowitz M, Joch A, Kossentini F, Hallapuro A (2003) H.264/AVC baseline profile decoder complexity analysis. IEEE Trans Circ Syst Video Technol 13(7):704–716
12. Hu Y, Zhang CT, Su YT (2007) "Information hiding based on intra prediction modes for H.264/AVC," IEEE International Conference on Multimedia and Expo (ICME 2007), Beijing, China, pp. 1231–1234
13. Huang SC, Kuo SY (2008) Optimization of hybridized error concealment for H.264. IEEE Trans Broadcast 54(3):499–516
14. Kapotas SK, Skodras AN (2009) Real time data hiding by exploiting the IPCM macroblocks in H.264/AVC streams. J Real-Time Image Proc 4(1):33–41
15. Kim SM, Kim SB, Hong Y, Won CS (2007) "Data hiding on H.264/AVC compressed video," International Conference on Image Analysis and Recognition (ICIAR 2007), Montreal, Canada, LNCS, vol. 4633, 2007, pp. 698–707
16. Kim T, Park K, Hong Y (2012) Video watermarking technique for H.264/AVC. Opt Eng 51(4)
17. Koz A, Alatan AA (2008) Oblivious spatio-temporal watermarking of digital video by exploiting the human visual system. IEEE Trans Circ Syst Video Technol 18(3):326–337
18. Lin ET (2005) Video and image watermark synchronization Purdue University
19. Lin WH, Horng SJ, Kao TW, Chen RJ, Chen YH, Lee CL, Terano T (2009) Image copyright protection with forward error correction. Expert Syst Appl 36(9):11888–11894
20. Lin WH, Horng SJ, Kao TW, Fan P, Lee CL, Pan Y (2008) An efficient watermarking method based on significant difference of wavelet coefficient quantization. IEEE Trans Multimed 10(5):746–757
21. Lin WH, Wang YR, Horng SJ, Pan Y (2009) A blind watermarking method using maximum wavelet coefficient quantization. Expert Syst Appl 36(9):11509–11516
22. Liu CH, Chen OTC (2008) "Data hiding in inter and intra prediction modes of H.264/AVC," IEEE International Symposium on Circuits and Systems (ISCAS 2008), Seattle, Washington, USA, pp. 3025–3028
23. Mansouri A, Aznaveh A, Torkamani-Azar F, Kurugollu F (2010) A low complexity video watermarking in H.264 compressed domain. IEEE Trans Inf Forensic Secur 5(4):649–657
24. Noorkami M, Mersereau RM (2007) A framework for robust watermarking of H.264-encoded video with controllable detection performance. IEEE Trans Inf Forensic Secur 2(1):14–23
25. Noorkami M, Mersereau RM (2008) Digital video watermarking in P-frames with controlled video bit-rate increase. IEEE Trans Inf Forensic Secur 3(3):441–455
26. Proefrock D, Richter H, Schlauweg M, Mueller E (2005) H.264/AVC video authentication using skipped macroblocks for an erasable watermark. Proc SPIE 5960:1480–1489

27. Qiu G, Marziliano P, Ho ATS, He D, Sun Q (2006) "A hybrid watermarking scheme for H.264/AVC video," Proceedings of the 17th International Conference on Pattern Recognition (ICPR 2004), vol. 4, Aug., pp. 2353–2356
28. Reference Software JM, J.V.T., version 14.0, http://iphome.hhi.de/suehring/tml/download/
29. Richardson IE (2004) H.264 and MPEG-4 video compression. Wiley
30. Saadi K, Bouridane A, Guessoum A (2009) "Combined fragile watermark and digital signature for H.264/AVC video authentication," EUSIPCO 2009, Scotland, Glasgow
31. Wang Z, Bovik AC, Sheikh HR, Simoncell EP (2004) Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 13(4):600–612
32. Wang CC, Hsu YC (2010) Fragile watermarking scheme for H.264 video authentication. Opt Eng 49(2)
33. Wolfgang RB, Podilchuk CI, Delp EJ (1999) Perceptual watermarks for digital images and video. Proc SPIE 3567:40–51, San Jose CA
34. Xiaopeng F, Yan L, Wen G (2003) "A novel coefficient scanning scheme for directional spatial prediction-based image compression," Proceedings of the International Conference on Multimedia and Expo, (ICME'03), Jul., pp. II-557–560
35. Xu D, Wang R (2011) Watermarking in H.264/AVC compressed domain using Exp-Golomb code words mapping. Opt Eng 50(9):267–279
36. Xu D, Wang R, Wang J (2011) A novel watermarking scheme for H.264/AVC video authentication. Signal Process Image Commun 26(6)
37. Yang G, Li J, He Y, Kang Z (2011) An information hiding algorithm based on intra-prediction modes and matrix coding for H.264/AVC video stream. Int J Electron Commun 65(4):331–337
38. Zhu BB, Swanson MD, Tewfik AH (2004) When seeing isn't believing. IEEE Signal Process Mag 21(2):40–49

Shi-Jinn Horng received his PhD. degree in computer science from the National Tsing Hua University in 1989. Currently, he is a professor in the Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology. His research interests include multimedia, biometrics, and parallel algorithms.

Mahmoud E. Farfoura is a PhD Student in the Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, Taiwan. His research interests include watermarking, and multimedia.

Pingzhi Fan (FIEE, SMIEEE) received his PhD degree in electronic engineering from Hull University, UK. He is currently a professor and a vice president of Southwest Jiaotong University, Chengdu, China. He is the inventor of 20 patents, and the author of over 300 research papers and 8 books. His research interests include CDMA theory and technology, information theory & coding, information security.

Xian Wang received his Ph.D. degrees in Communication and Information Systems from Southwest Jiaotong University, Chengdu, China in 2008 and was an associate professor at the same university. He is also a postdoctoral researcher with the Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, Taiwan. His research interests include mobility management and performance modeling for personal communications service network.

Tianrui Li is a Professor of School of Information Science and Technology, Southwest Jiaotong University, China. His current research is supported by the National Science Foundation of China (NSFC: 60873108, 60875034). Since 2000, he has published over 50 research papers in refereed journals, and conferences. He has served as ISKE2007, ISKE2008, ISKE2009 program chairs, IEEE GrC 2009 program vice chair and RSKT2008, FLINS2010 organizing chair, etc. and has been a reviewer for several leading academic journals.

Dr. Jing-Ming Guo received his Ph. D. from the Institute of Communication Engineering, National Taiwan University. He is currently working as a professor at the Department of Electrical Engineering, National Taiwan University of Science and Technology. He has published more than 100 papers in journals and conferences. His areas of interest are Multimedia signal processing, digital image processing, digital video processing, computer vision, digital halftoning, digital watermarking, biometric recognition.