Signal Processing 97 (2014) 40–54

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Transcoding resilient video watermarking scheme based on spatio-temporal HVS and DCT

Antonio Cedillo-Hernandez [a], Manuel Cedillo-Hernandez [b], Mireya Garcia-Vazquez [c], Mariko Nakano-Miyatake [a,*], Hector Perez-Meana [a], Alejandro Ramirez-Acosta [d]

[a] Postgraduate Section, Mechanical Electrical Engineering School, National Polytechnic Institute of Mexico, Santa Ana Av. 1000, D.F. 04430, Mexico
[b] Electric Engineering Division, Engineering Faculty, National Autonomous University of Mexico, Universidad Av. 3000, D.F. 04510, Mexico
[c] Research and Development of Digital Technology Center (CITEDI), National Polytechnic Institute of Mexico, Park Av. 1310, Tijuana, B.C. 22510, México
[d] MIRAL. R&D, 1047 Palm Garden, Imperial Beach 91932, USA

ARTICLE INFO

Article history:
Received 3 February 2013
Received in revised form
27 May 2013
Accepted 27 August 2013
Available online 11 September 2013

Keywords:
Video watermarking
Video transcoding
Human Visual System
Motion distortion threshold
Visual attention region

ABSTRACT

Video transcoding is a legitimate operation widely used to modify video format in order to access the video content in the end-user's devices, which may have some limitations in the spatial and temporal resolutions, bit-rate and video coding standards. In many previous watermarking algorithms the embedded watermark is not able to survive video transcoding, because this operation is a combination of some aggressive attacks, especially when lower bit-rate coding is required in the target device. As a consequence of the transcoding operation, the embedded watermark may be lost. This paper proposes a robust video watermarking scheme against video transcoding performed on base-band domain. In order to obtain the watermark robustness against video transcoding, four criteria based on Human Visual System (HVS) are employed to embed a sufficiently robust watermark while preserving its imperceptibility. The quantization index modulation (QIM) algorithm is used to embed and detect the watermark in 2D-Discrete Cosine Transform (2D-DCT) domain. The watermark imperceptibility is evaluated by conventional peak signal to noise ratio (PSNR) and structural similarity index (SSIM), obtaining sufficiently good visual quality. Computer simulation results show the watermark robustness against video transcoding as well as common signal processing operations and intentional attacks for video sequences.

1. Introduction

With the rapid advance of multimedia and networking technologies, multimedia services such as teleconferencing, video on demand and distance learning have become more popular in our daily life. In these applications the video format is often required to be converted in order to adapt to several channel capacities (e.g., network bandwidth) as well as end-user's terminal capabilities (e.g., computing and display capacity) [1]. The transcoding is one of the key technologies to fulfill this challenging task. Using a transcoder we can convert

* Corresponding author. Tel./fax: +52 55 5656 2058.
E-mail addresses: antoniochz@hotmail.com (A. Cedillo-Hernandez), mcedillohdz@hotmail.com (M. Cedillo-Hernandez), freemgraciav@gmail.com (M. Garcia-Vazquez), mnakano@ipn.mx, mariko@infinitum.com.mx (M. Nakano-Miyatake), hmpm@prodigy.net.mx (H. Perez-Meana), ramacos10@hotmail.com (A. Ramirez-Acosta).

0165-1684
http://dx.doi.org/10.1016/j.sigpro.2013.08.019

a previously compressed video bit-stream into another bit-stream with different bit-rates, different spatial resolutions and/or different compression standards, etc. In the copyright protection issue, the transcoding poses new challenges on video watermarking technologies since it performs complex conversion operations that generate problems regarding the preservation of embedded copyright information. Then malicious users can perform the transcoding to obtain copyright-free video sequence with similar quality as the original ones and they can distribute them illegally [2]. Considering the above mentioned situation, watermark robustness against video transcoding must be considered to design an efficient video watermarking algorithm, however in almost all video watermarking techniques proposed in literature, the embedded watermark is not robust against transcoding and therefore the copyright protection is not sufficiently done.

Recently several robust watermarking schemes have been proposed in the literature [3–6], in which the resilience to transcoding is also considered. Lee et al. [3] propose a real-time video watermarking robust against transcoding, in which notable results against spatial reduction are shown, obtaining robustness against conversion of spatial resolution from High-Definition Television (HDTV) to Quarter Video Graphics Array (QVGA). This scheme is performed on MPEG-2 video bit-stream directly in order to satisfy the real-time requirements, however it generates vulnerability against the conversion of other video compression standards with low bit rates. Chen et al. [4] propose a robust video watermarking algorithm using the singular value decomposition (SVD) and slope-based embedding technique, in which synchronization information, together with the watermark sequence, is embedded to combat frame attacks, however re-synchronization mechanism of this scheme is not sufficient for the frame rate reduction caused by some aggressive transcoding. In [5], the Human Visual System (HVS) is used to adapt the watermarking energy of the quantization based video watermarking scheme in DWT domain. This scheme shows watermark robustness to some signal processing attacks; however combined attacks caused by common transcoding tasks remove the embedded watermark sequence. Ling et al. [6] propose a video watermarking algorithm robust mainly to geometrical distortions using Harris-Affine interest point detector. The watermark robustness of this scheme strongly depends on an accurate detection of interest points and generally an aggressive transcoding causes an inaccurate detection of many interest points, reducing the performance of this scheme.

The watermark embedding domain is an important aspect to design a video watermarking scheme robust against video transcoding. Video watermarking algorithms proposed in the literature can be classified into three main categories from embedding domains points of view: base-band domain algorithms [7,8], watermarking during video coding process [9,10] and watermark embedding directly on the encoded video sequence [11,12]. We consider that the base-band domain technique is more suitable for a watermarking scheme robust against aggressive video transcoding, because it is not focused on any video compression standard and also in this domain the watermark embedding energy can be adjusted easily according to its robustness and imperceptivity requirements, since the whole spatial information is available.

In this paper we propose a video watermarking scheme robust against video transcoding which performs in base-band domain using Quantization Index Modulation (QIM) algorithm [13]. To design a video watermarking scheme robust against an aggressive video transcoding task, first the key aspects of video transcoding, such as quality degradation caused by low bit-rate coding, similarities and difference among video compression standards, and effects of the temporal/spatial resolution change, are analyzed in detail. To embed a watermark sequence as robust as possible keeping the watermark imperceptibility, in the proposed scheme the quantization step size of the QIM algorithm is adaptively calculated using spatial and temporal HVS criteria. In the image watermarking techniques, QIM algorithms with adaptive quantization step size based on the spatial HVS properties have been proposed in order to obtain watermark imperceptibility and robustness simultaneously [14,15]. The proposed scheme also considers and exploits temporal information in order to get advantage on deficiency of the HVS to follow regions with high motion speed, using the spatio-temporal contrast sensitivity function and influence of eye movement. Additionally in the proposed scheme, the visual attention region is segmented in each video frame using Information Maximization to obtain more adequate quantification step size. So the watermark embedding process is performed combining four HVS criteria; texture and luminance sensitivity, a motion distortion threshold and visual attention region. The performance of the proposed scheme is compared with four recently reported robust video watermarking schemes [3–6], showing a better performance of the proposed scheme, especially in robustness against video transcoding. We consider that the main contribution of the proposed scheme is robustness to transcoding which is obtained by a detailed analysis of transcoding task and exploiting the spatio-temporal HVS properties in order to obtain adaptively the quantization step size of the QIM algorithm. In our best knowledge there is no another video watermarking scheme that exploits the spatio-temporal HVS criteria and visual attention region estimation to improve watermark robustness against aggressive attacks.

The rest of the paper is organized as follows: In Section 2, we analyze key aspects of video transcoding process to design an efficient watermarking scheme. Section 3 provides a detailed explanation of the watermark energy adaptation based on spatio-temporal HVS-based criteria and the visual attention region segmentation. In Section 4 the proposed scheme is described in detail. The evaluation results of the proposed scheme are compared with four recently reported video watermarking schemes in Section 5 and finally Section 6 provides the conclusions of this research.

2. Video transcoding

Digital video can be dynamically adapted according to the available resources of the end-user's devices, such as computing power and display capability, as well as the channel capacity for transmission of the video sequence, such as channel bandwidth. One common example is the delivery of a high-quality multimedia source, such as a Digital Versatile Disc (DVD) or High Definition TV (HDTV), to a receiver with lower resources, such as Smart Phone, Tablets and Personal

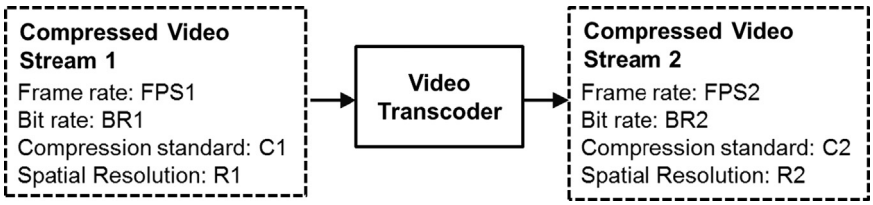

Fig. 1. Compressed video stream conversion using a video transcoder.

Computer. The video transcoding is one of the core technologies to satisfy this challenging task, creating adequate bit-stream directly from the original video sequence without user's conscious about decoding and re-encoding process [16]. The transcoding is defined as the operation of converting a video encoded in some format to another one with different characteristics [17]. In practice, the main characteristics of a digital video which can be transformed by a transcoder are the frame rate, bit-rate, video compression standard and spatial resolution as shown in Fig. 1. The transcoding task can be classified as homogeneous and heterogeneous. A homogeneous transcoder performs the conversion between video bit-streams of the same video compression standard, while a heterogeneous transcoder provides conversions between video bit-streams with different video compression standards.

Frame rate reduction, also known as temporal resolution reduction, is necessary when the user's end-system has less processing capability that cannot support the same frame rate as that of the original video. There are several strategies to avoid the loss of the embedded watermark signal due to this operation. One possible solution is to limit the watermark embedding process to intra frames (I-frames), since the transcoder discards inter-frames (P or B-frames) mainly to obtain frame rate reduction. However, in a blind watermark detection scheme, the parameters such as the group of pictures (GOP) or frame rate are unknown, and then the I-frame cannot be identified in the detection stage. Although a redundant embedding of the watermark signal in every frame of video sequence can solve this problem regardless of the frame rates of the target video, this method becomes vulnerable to intra-video collusion attacks, which can obtain the watermark sequence from several frames with different scenes [2].

According to the applications, several video compression standards, such as MPEG-2, Motion Picture Expert Group 4 Part 2 (MPEG-4), Windows Media Video 9 Codec (VC-1), Motion Picture Expert Group 4 Part 10 (H.264 AVC) and On2 True-Motion VP6 (VP6) are employed [16,17]. The necessity of the conversions between different video compression standards is generally associated to obtain acceptable visual quality with lower bit rate. In the heterogeneous transcoding the change of the frame format, as a result of change of the video compression standard, can be considered as an aggressive operation for many watermarking schemes, because a significant loss of information caused by quantization process may damage the integrity of the embedded watermark signal. The required bit-rate of the target video determines the quantization parameters and it is responsible for maintaining video quality while satisfying bandwidth, delay and memory space constraints [17,18].

Spatial resolution reduction is required when the display capability of the end-user's device is lower than that of the original one. For example, generally a teleconferencing system uses a Common Intermediate Format (CIF) with a resolution of 352×288 pixels. This resolution must be downscaled if the receiver system has a lower display capability, such as smartphone with Quarter CIF (QCIF) format (176×144 pixels). In this situation, a transcoder can carry out the spatial resolution reduction using different techniques [16].

Analyzing all processes performed by a heterogeneous transcoder, we conclude that in order to obtain an efficient watermarking scheme, the following two parameters must be considered: (a) the 2D-DCT coefficients that result more resilient to the quantization process, in which the watermark is embedded and (b) the maximum watermark energy that allows the embedded watermark to be robust against all attacks mentioned in this section, while preserving the watermark imperceptibility by the HVS.

3. HVS-based watermark energy computation

As mentioned in Section 2, in order to obtain the robustness against aggressive heterogeneous transcoding, we need to embed a sufficiently strong watermark signal, considering the trade-off between watermark robustness and imperceptibility. To achieve this difficult task, the watermark energy is adapted using four HVS criteria, which takes an advantage of video watermarking schemes performed in base-band domain. The first two criteria are based on the relationship between the sensibility of the HVS to the visual quality degradation and spatial characteristics of each frame of the video sequence. The watermarking energy is calculated using the texture and luminance masking [19]. The third criterion is based on the failure of the HVS to follow regions with high motion speed. We use a concept based on the just noticeable distortion (JND) adapted for video as a model of the observer's eye movement [20]. As the fourth HVS criterion, the visual attention region is obtained for each video frame in order to generate less distortion in regions where observer's attention is attracted [21].

3.1. Texture and luminance masking

The most well-know HVS properties used in image coding fields are frequency sensitivity, texture and luminance masking. The texture masking property suggests that the human eye's sensitivity to error is low in the highly textured image areas, while the luminance masking property suggests that in the bright and dark image regions, the error sensitivity by human eye is low [19]. The luminance space of each frame of a video sequence is segmented into non-overlapped

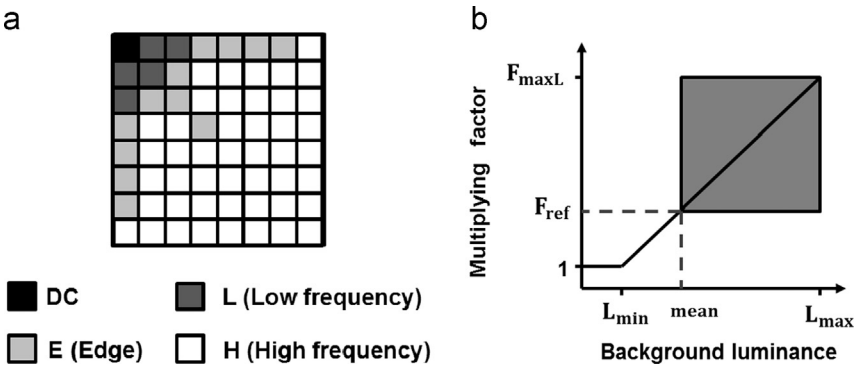

Fig. 2. (a) Segmentation of 2D-DCT block and (b) luminance masking linear modeling [9].

blocks of 8×8 pixels, in which the texture and luminance masking are calculated using the algorithm proposed by [19]. Both criteria are combined to obtain a maximum imperceptible distortion $m(n,k)$ by the HVS for k-th block of n-th video frame, which is given by

$$m(n,k) = TexMask(n,k) \times LumMask(n,k) \tag{1}$$

where $TexMask(n,k)$ and $LumMask(n,k)$ are the texture and luminance masking, respectively. To calculate the texture masking $TexMask(n,k)$, first each block of 8×8 pixels is classified into texture, plain and edge blocks according to their spatial characteristics. To perform this classification, each block is transformed by 2D-DCT and the block in DCT domain is furthermore segmented into four areas as shown in Fig. 2(a), where the absolute sum of coefficients values of each area is denoted by DC, L, E and H, respectively [19]. The classification process can be summarized by the following conditions: a block is classified as plain block if $E+H\le\mu_1$ ($\mu_1=125$), otherwise if $(L+E)/H>\gamma$ ($\gamma=4$) then the block is considered as edge block and it is classified as texture block if $E+H>\kappa$ ($\kappa=290$). Subsequently the texture masking $TexMask(n,k)$ is assigned according to the type of k-th block. If the block is plain then $TexMask(n,k)=1$, considering that the error in the plain block is most noticeable by the HVS. If the block belongs to the edge block and $L+E\le 400$ then $TexMask(n,k)=1.125$, otherwise $TexMask(n,k)=1.25$ [22]. Finally, for texture blocks the texture masking is obtained by

$$TexMask(n,k) = (F_{\max T}-1)\times\frac{TexE(n,k)-Min}{Max-Min}+1 \tag{2}$$

where $TexE(n,k)=E+H$ is the energy value for the k-th texture block of the n-th video frame; Max and Min represent the maximum and minimum energy for texture blocks, whose values are 1800 and 290, respectively [19], and F_{maxT} is the maximum elevation value used for fine adjustment of the model whose value is set to 2.25.

In the case of luminance masking, the method is divided into two parts: a linear model for middle and high luminance which is based on the Weber's law and a nonlinear model for low luminance where Weber's law is invalid [19]. Fig. 2(b) shows an approximation of the Weber's luminance law, where L_{min} and L_{max} values denote the luminance range for the linear model, which are determined as 90 and 255, respectively [19]. F_{maxL} represents the maximum luminance factor whose value is set to 2 [19]. The luminance masking $LumMask(n,k)$ of the k-th block of the n-th video frame is given by

$$LumMask(n,k) = \begin{cases} (F_{\max L}-F_{ref})\times\frac{DC(n,k)-mean}{L_{\max}-mean}+1 & \text{if } DC(n,k)>mean \\ 1 & \text{if } 25\le DC(n,k)\le mean \\ 1.125 & \text{if } 15\le DC(n,k)<25 \\ 1.125 & \text{if } DC(n,k)<15 \end{cases} \tag{3}$$

where $DC(n,k)$ is the DC coefficient of k-th block of n-th frame, $mean$ is mean luminance value of n-th frame and F_{ref} is the reference factor corresponding to $mean$ luminance value in linear model. Fig. 3 shows the first frame from "Foreman" video sequences and its texture and luminance masking representation. In the texture masking (Fig. 3(b)), "black" represents plain blocks, where the HVS sensitivity to distortion is larger, while "white" and "gray" represent edge and texture blocks, respectively. In the luminance masking (Fig. 3(c)), regions with higher and lower brightness (darkness) suggest a lower sensitivity to distortion by the HVS than other regions.

3.2. Motion distortion threshold

The third criterion is based on the deficiency of the HVS to follow regions with high motion speed. Taking in account the perceptual motion speed of each 2D-DCT block, the motion distortion threshold $T(n,k,i,j)$ can be computed. This threshold value indicates the maximum distortion that can be tolerated in the (i,j)-th sub-band of the k-th 2D-DCT block of the n-th video frame, because a distortion with smaller value than this threshold cannot be perceived by the HVS. To compute this threshold, we

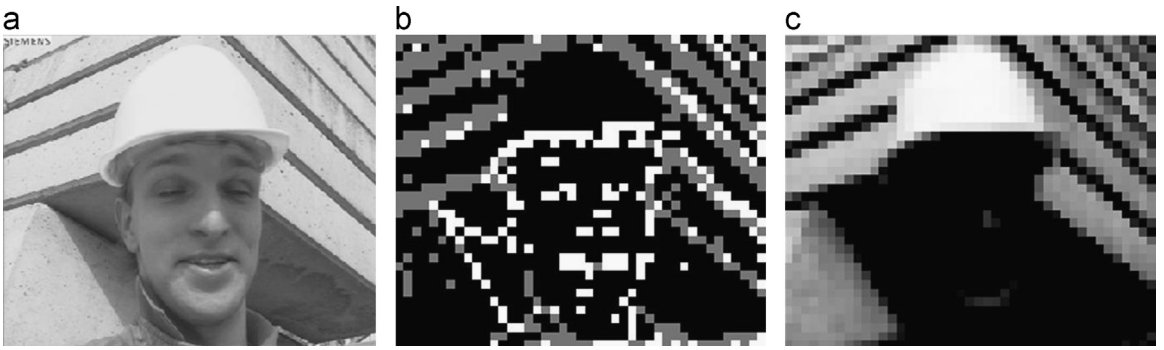

Fig. 3. First frame of "Foreman" video sequence: (a) original video frame, (b) texture masking and (c) luminance masking.

use a DCT-domain JND-based method for video sequence proposed in [20], which is given by

$$T(n,k,i,j)=\frac{1}{G(n,k,i,j)}\cdot\frac{M}{\varphi_i\varphi_j(L_{\max}-L_{\min})}\cdot \times\frac{1}{0.6+0.4\cos^2\theta_{i,j}},\quad i,j=0,1,\ldots L-1 \tag{4}$$

where L is the sub-block size, $G(n,k,i,j)$ is the spatio-temporal contrast sensitivity function (CSF), which is described later, $L_{\max}$ and $L_{\min}$ represent the maximum and minimum display luminance values while M is the number of gray levels, which is set to 256, ϕ_i and ϕ_j are 2D-DCT normalization factors, which are $\varphi_u=\sqrt{1/L}$ for $u=0$ and $\varphi_u=\sqrt{2/L}$ for $u\neq 0$; $\theta_{i,j}=\arcsin(2\rho_{i,0}\rho_{0,j}/\rho_{i,j}^2)$ is the visual angle, where $\rho_{i,j}=\sqrt{(i/\omega_x)^2(j/\omega_y)^2/2}$ is the spatial sub-band frequency, where $\omega_\hbar=2\cdot\arctan(\Lambda_\hbar/2l)$, $\hbar=x,y$ denotes either the horizontal or vertical sizes of a pixel in degrees of visual angle, which are represented by the viewing distance l and display size $\Lambda_x\times\Lambda_y$ [20].

The spatio-temporal CSF $G(n,k,i,j)$ in (4) is defined by [20]

$$G(n,k,i,j)=c_0\left(k_1+k_2\left|\log\left(\varepsilon\cdot\frac{v(n,k)}{3}\right)\right|^3\right)\cdot v(n,k)\cdot(2\pi\rho_{i,j})^2 \cdot\exp(-2\pi\rho_{i,j}c_1(\varepsilon\cdot v(n,k)+2)/k_3) \tag{5}$$

where k_1,k_2,k_3 and ε are constants which are empirically set to 6.1,7.3, 23 and 1.7, respectively [20], and c_0 and c_1 are constant values that control the magnitude and the bandwidth of a CSF whose best-fitted values are 7.126 and 0.565, respectively [20]. $v(n,k)$ is the perceptual motion speed given by $v(n,k)=v_I(n,k)-v_E(n,k)$, which corresponds to the motion speed without eye movement, v_I, compensated by eye movement speed, v_E, where

$$v_I(n,k)=f\cdot\sqrt{(MV_x(n,t)\cdot\omega_x)^2+(MV_y(n,t)\cdot\omega_y)^2} \tag{6}$$

represents the motion speed of k-th block without eye movement, f is the video frame rate and $(MV_x(n,t), MV_y(n,t))$ is motion vector, while $v_E(n,k)=\min[g\cdot v_I(n,t)+v_{\min}, v_{\max}]$ is eye movement speed [20], where g is the gain factor related to the object tracking efficiency of eye, $v_{\min}$ and $v_{\max}$ are minimum and maximum eye movement speed, whose values are set to 0.92, 0.15 and 80.0 deg/s, respectively [20].

3.3. Visual attention region

Visual attention is a mechanism that filters out redundant visual information and detects the most relevant parts of our visual field. It is one of the fundamental properties of the HVS that can be used in several applications, such as image and video coding, and computer vision. Due to that, many research groups are currently investigating computational modeling of the visual attention system [21,22]. In the watermarking schemes, especially in video watermarking schemes where the duration of observation of each frame is short, the detection of the visual attention regions allows us to determine an adequate watermarking energy, strong enough to survive transcoding attacks, while keeping a minimum perceptual distortion.

The visual Attention based on Information Maximization (AIM) [21] is one of the methods most supported by the mechanism of the HVS, in which some orthogonal basis functions are generated using Independent Component Analysis (ICA) from a great amount of visual patches of $T\times T$ pixels randomly obtained from a considerably wide range of natural images. After generation of the basis functions, which are matrices with $T\times T$ elements, their pseudo-inverse matrices are calculated. Each patch $\mathbf{C}_{i,j}$ of $T\times T$ pixels with a center (i, j) of the input frame is analyzed in terms of similarity $\mathbf{S}_{i,j,r}$ with the r-th basis function, where $r=1,\ldots,R$.

$$\mathbf{S}_{i,j,r}=\sum_{p=0}^{T-1}\sum_{q=0}^{T-1}\mathbf{C}_{i,j}(u,v)\cdot\Psi_r(p,q) \tag{7}$$

where $\mathbf{\Psi}_r$ is inverse matrix of r-th basis function, $u=i-\lfloor(T-1)/2\rfloor+p, v=j-\lfloor(T-1)/2\rfloor+q$ and R is the number of basis functions. Then $\mathbf{S}_{i,j,r}$ is normalized in order that the range of normalized $\overline{\mathbf{S}}_{i,j,r}$ is [0, 1], where the highest and lowest $\overline{\mathbf{S}}_{i,j,r}$ are 1 and 0, respectively. The higher similarity between the (i,j)-th patch $\mathbf{C}_{i,j}$ and r-th basis function produces the higher value of $\overline{\mathbf{S}}_{i,j,r}$. Considering that the observer's attention is attracted in the regions with unexpected objects or patterns compared with their surround, and then to obtain this unexpectedness, the self-information of the (i,j)-th patch $\mathbf{C}_{i,j}$ for r-th basis function is calculated, which is $1/\log(\Pr_{i,j}(r))=-\log(Pr_{i,j}(r))$, where $\Pr_{i,j}(r)$ is the probability density function of the normalized similarity $\overline{\mathbf{S}}_{i,j,r}$. For example, if the similarity $\overline{\mathbf{S}}_{i_k,j_k,r}$ between a patch $\mathbf{C}_{i_k,j_k}$ and r-th basis function is high and also many other patches of the frame have same high similarities with the r-th basis

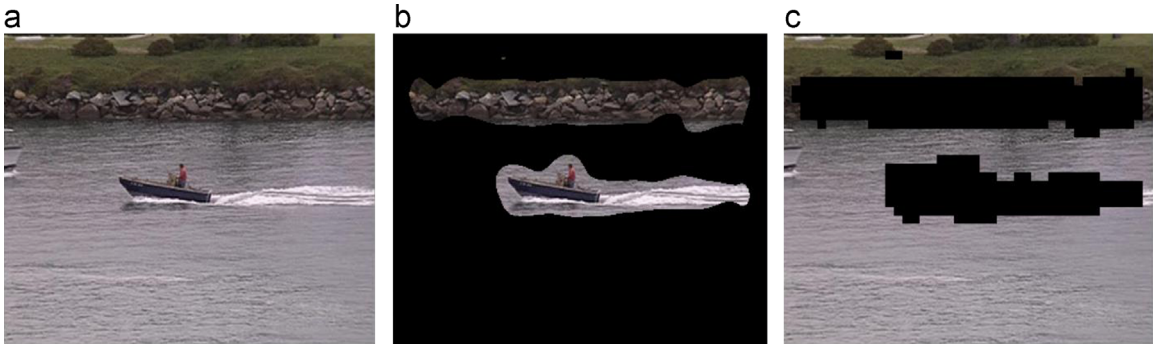

Fig. 4. First frame of "Coastguard" video sequence: (a) original raw video frame, (b) visual attention region and (c) RONI defined as the opposite region to visual attention represented by black blocks.

function, then the $\Pr_{i_k j_k}(r)$ is high meaning that the patch $\mathbf{C}_{i_k j_k}$ is a common pattern in the frame, so its self-information is low. The self-information of the patch $\mathbf{C}_{i,j}$ for all basis functions is a sum of the self-information of each basis function, which is

$$SMap(i,j) = -\sum_{r=1}^{R} \log\left(\Pr_{i,j}(r)\right) \tag{8}$$

The resultant *SMap* is visual attention map, whose values indicate the unexpectedness grades and then using these values the visual attention regions can be segmented from the rest of the frame. Fig. 4 shows the visual attention region obtained applying the AIM-based method [21] to the first frame of "Coastguard" video sequence. Once the visual attention regions are obtained, the rest of image is segmented by blocks of 8×8 pixels, denominated as Region of No Interest (RONI) blocks.

4. Proposed method

As we discussed in Section 2, each of the transcoding tasks performs complex conversion operations which generate problems related to the preservation of the embedded copyright information. In the proposed video watermark scheme, the particular tasks of all processes performed in the video transcoding, such as changes of frame rate, bit-rate, compression standard and spatial resolution as shown in Fig. 1, are considered to determine an adequate watermark embedding energy avoiding loss of watermark signal, while keeping high quality of watermarked video sequences.

In order to get robustness against frame rate reduction, we use two strategies: (a) The detector of video scene change is introduced that allows embedding watermark signals generated by different keys in frames of different scenes, which allows the scheme to be robust against frame-based attacks, whose objective is the video temporal de-synchronization, such as frame dropping, frame averaging and frame swapping; (b) the adaptive watermark energy calculation based on the HVS is introduced to avoid the intra-video collusion attack, because the watermark sequence is also different among frames of the same scene.

Since in proposed scheme, robustness against heterogeneous transcoding is considered, the affinities of different video compression standards must be analyzed to obtain robustness against this attack. Some of the most significant similarities of the video compression standards are the macroblock as the basic processing unit for motion estimation, which is used to determine the motion distortion threshold described in Section 3.2, the YCbCr color space usage, the 2D-DCT transform and the Sub Quarter CIF (SQCIF) format (128×96 pixels) as the minimum unit of spatial resolution.

In the proposed scheme we embed the watermark signal into non-overlapped 2D-DCT blocks of 8×8 pixels in luminance space. Considering standard color space qof video and low correlation among three color components, YCbCr color space is used in the proposed video watermarking scheme. To obtain a robust video watermarking scheme against aggressive quantification process, which is used to reduce bit-rate, first we analyzed video quality degradation caused by the quantization process with different bit-rates in several compression standards. We analyzed 10 video sequences with 150 frames, CIF format and 30 FPS each one, under six different bit-rates, which are 4 Mbps, 2 Mbps, 1 Mbps, 512 Kbps, 256 Kbps and 128 Kbps, with five video compression standards, which are MPEG-2, MPEG-4, VC-1, H.264 AVC and VP6. From the results shown in Fig. 5, we can determine that the average video quality obtained from H.264 AVC with 4 Mbps bit-rate causes the lowest degradation of the video quality (PSNR = 37.02, SSIM = 0.951), while the VC-1 with 128 Kbps bit-rate causes the highest degradation of the video quality (PSNR = 29.52, SSIM = 0.783). The visual quality of both compressed video can be observed in Fig. 6.

Taking in account that VC-1 compression standard with 128 Kbps bit-rate causes the highest distortion of video quality, the goal of proposed scheme can be considered to obtain the watermark robustness against this processing. To determine the DCT coefficients more resilient to the quantization process caused by the VC-1 video compression standard with bit-rate of 128 Kbps, the corresponding quantization process is applied to DCT blocks of 8×8 coefficients and the number of non-zero AC coefficients is computed.

Fig. 7 shows the percentage of AC coefficients which satisfy $|AC_i| > 0$ ($i \in \{1 \ldots 63\}$), after quantization process, where 10 video sequences are used. From the figure, two AC coefficients with lowest frequency, AC_1 ($C_{1,2}$) and AC_2 ($C_{2,1}$) are considered as adequate coefficients where the watermark can be embedded in a robust manner against most aggressive case, i.e. the VC-1 video compression

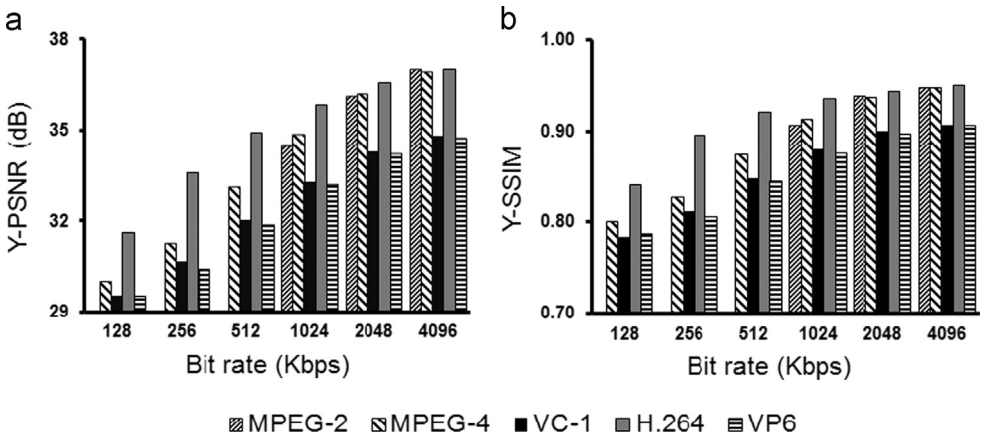

Fig. 5. Visual quality degradation measured by (a) PSNR and (b) SSIM for 10 video sequences encoding over MPEG-2, MPEG-4, VC-1, H.264 AVC and VP6 video compression standards with bit rate from 128 Kbps to 4 Mbps.

Fig. 6. (a) Close-up of the first frame of "Paris" video sequence, (b) encoded over H.264 AVC video compression standard with a bit rate of 4 Mbps, and (c) encoded over VC-1 video compression standard with a bit rate of 128 Kbps.

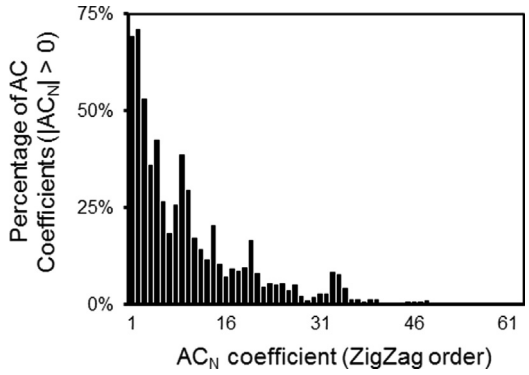

Fig. 7. Analysis of AC coefficients after VC-1 compression standard with 128 Kbps bit rate video coding.

standard with bit-rate of 128 Kbps. Considering the above results, in the proposed scheme one-bit of the watermark sequence is embedded into AC_2 coefficient of each non-overlapping 2D-DCT block of 8×8 coefficients. Finally, in order to obtain robustness against spatial resolution changes, which causes spatial de-synchronization between watermark embedding and detection processes, in the proposed watermark detection process first input video sequence is resized to obtain a standard spatial resolution format, and then the detection process is done.

4.1. Watermark embedding process

This section details the proposed watermark embedding process. Fig. 8 shows the watermark embedding process, which is described as follows: (1) Divide the video sequence into scenes, which will allow embedding watermark signals generated by different keys for each scene. (2) Obtain the luminance space of the original video frame F_n, where $n = 1, 2, \ldots, N$ and N is the total number of video frames in a scene. (3) Segment the luminance space into non-overlapping blocks of size 8×8 pixels and the 2D-DCT transform is applied to each block. One bit of the watermark signal is embedded into each block. (4) Calculate the visual attention region by performing the process described in Section 3.3. Each k-th block of the n-th video frame is classified as RONI or ROI block. (5) Only for RONI blocks, compute the values of $m(n, k)$ given by (1) and $T(n,k,2,1)$ given by (4). (6) Generate a watermark vector $W = \{w_1, w_2, \ldots, w_K\}$ using a user's secret key for a given scene, where $w_k \in \{0, 1\}, \quad k = 1, 2, \ldots, K$. (7) The QIM embedding process is applied to embed a watermark bit into AC_2 coefficient of each 2D-DCT block, according to the following equation:

$$C'_{2,1} = \begin{cases} sign(C_{2,1}) \times \lfloor \frac{|C_{2,1}|}{2S_{n,k}} \rfloor \times 2S_{n,k}, & w_k = 0 \\ sign(C_{2,1}) \times \left(\lfloor \frac{|C_{2,1}|}{2S_{n,k}} \rfloor \times 2S_{n,k} + S_{n,k} \right), & w_k = 1 \end{cases} \quad (9)$$

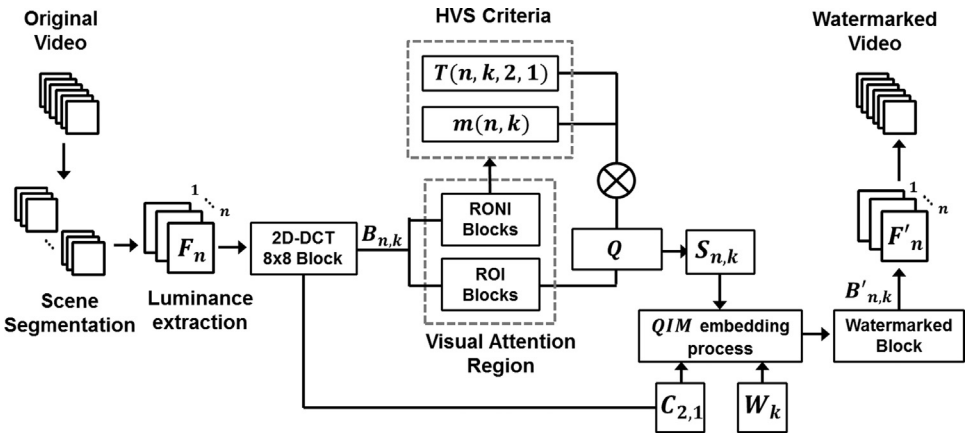

Fig. 8. Proposed watermark embedding process.

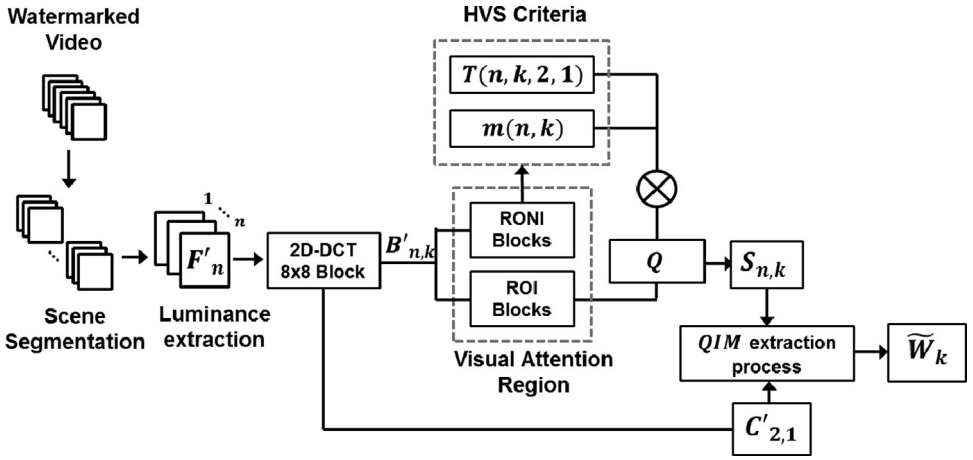

Fig. 9. Proposed watermark extraction process.

where $C_{2,1}$ and $C'_{2,1}$ are the original and the watermarked AC_2 coefficients, respectively, $S_{n,k}$ represents the dynamic quantization step size for the k-th block and its value is determined as follows:

$$S_{n,k} = \begin{cases} Q \cdot m(n,k) \cdot T(n,k,2,1), & B_{n,k} \in RONI \\ Q, & B_{n,k} \notin RONI \end{cases} \quad (10)$$

where Q represents the static quantization step size (Q-step) determined through the trade-off between robustness and imperceptibility, which is described later.

4.2. Watermark extraction process

The watermark extraction process is shown in Fig. 9 and is described as follows: (1) Divide the video sequence into scenes. (2) Obtain the luminance space of the watermarked video frameF'_n. (3) Segment the luminance space into non-overlapping blocks of size 8×8 pixels and apply the 2D-DCT transform to each block. (4) Extract the watermark vector from luminance space of the watermarked frame using the QIM extraction process:

$$\tilde{w}_k = \begin{cases} 0 & \text{if } round(C'_{2,1}/S_{n,k}) = even \\ 1 & \text{if } round(C'_{2,1}/S_{n,k}) = odd \end{cases} \quad (11)$$

where $C'_{2,1}$ are the watermarked AC_2 coefficient and $S_{n,k}$ represents the dynamic quantization step size for the k-th block of the n-th video frame.

5. Experimental results

To evaluate the performance of the proposed scheme, we used 10 video sequences with CIF format and 30 FPS. All video sequences have at least 150 frames which are available in [23]. Fig. 10 shows video sequences used for evaluation of the proposed scheme. The proposed scheme is evaluated from the watermark imperceptibility and robustness points of view.

5.1. Parameter setting

Determining an appropriate static quantization step size Q is crucial to obtain a good performance of the proposed scheme from the watermark imperceptibility and robustness points of view. First to evaluate the relationship between static quantization step size Q and watermark imperceptibility, the PSNR and SSIM index are obtained varying this value (Fig. 11(a)). Also the relationship between this factor and watermark robustness is evaluated using the Bit Error Rate (BER) of the extracted watermark bit sequence respect to the embedded one without any attacks (Fig. 11(b)). In

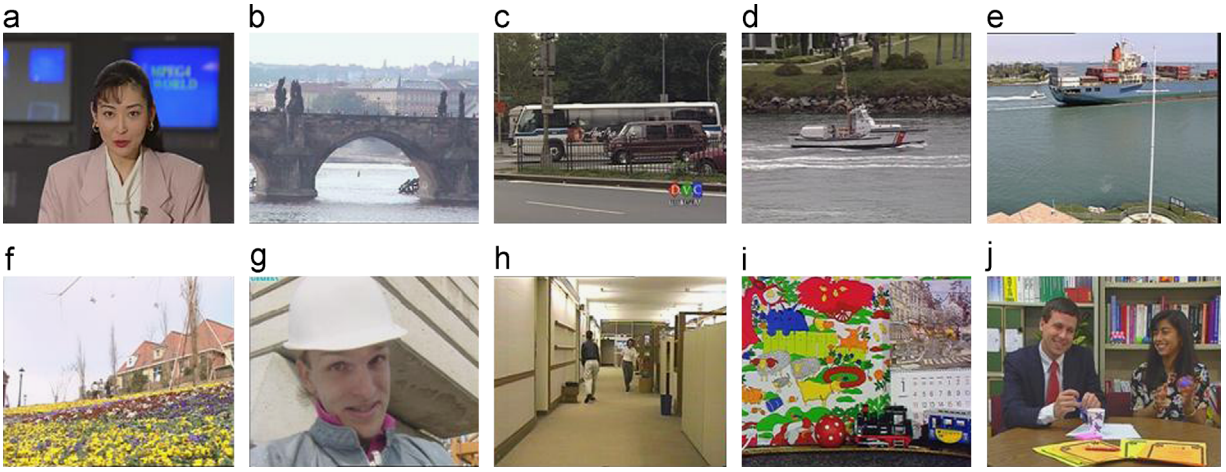

Fig. 10. Video sequences used for the evaluation of the proposed scheme: (a) Akiyo, (b) bridge-close, (c) bus, (d) coastguard, (e) container, (f) flower, (g) foreman, (h) hall, (i) mobile and (j) Paris.

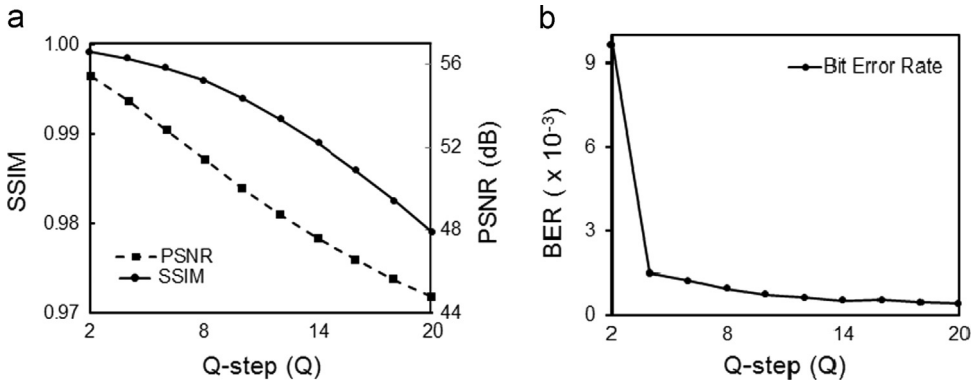

Fig. 11. (a) PSNR and SSIM index values of the watermarked video scheme for different static quantization steps Q, (b) Bit Error Rate (BER) with different static step sizes Q, here any attack is received.

both cases, Q value is increased by 2–20. From both figures, we determined that the most adequate value of Q is equal to 16, because using this value the PSNR and SSIM of the watermarked video sequence respect to the original one is approximately 47 dB and 0.99, respectively, also the BER is approximately 0.0005. It is worth noting that the SSIM provides a perceptual distortion in range of [0.0, 1.0], when both images are numerically same, this value is equal to 1.

5.2. Watermark imperceptibility

We evaluate the watermark imperceptibility of the proposed scheme using PSNR and SSIM which are calculated using the luminance space of watermarked frame and the original one. To evaluate the effect of adaptive watermark embedding energy based on the HVS, the imperceptibility is calculated under two conditions: (a) watermark is embedded without consider HVS criteria, i.e. only constant Q value is applied to embed watermark into 2D-DCT blocks of whole frame and (b) watermark is embedded considering four HVS criteria described in Section 3. Table 1 shows the mean and variance values for dynamic quantization step size $S_{n,k}$, the PSNR and SSIM metrics calculated using the 150 frames of all video sequences under the above mentioned conditions. From Table 1, we can observe that the mean value of dynamic quantization step size $S_{n,k}$ with the four HVS criteria is much larger than that value without the HVS criteria, which means that the watermark robustness is increased considerably when the HVS criteria are used. However, the watermark imperceptibility is not sacrificed, which can be observed specially from the SSIM values of both situations, i.e. with/without HVS. The higher variance values of $S_{n,k}$, PSNR and SSIM with the HVS criteria than these without

Table 1
The watermark imperceptibility of the proposed scheme using mean and variance values for dynamic quantization step $S_{n,k}$, PSNR and SSIM metrics calculated with and without HVS criteria.

Criteria	$S_{n,k}$		PSNR		SSIM	
	Mean	Variance	Mean	Variance	Mean	Variance
Without HVS	16	0	47.76	0.004	0.9993	2.36E−08
With HVS	22.42	0.042	46.81	0.019	0.9978	5.44E−07

Fig. 12. (a) The original 150th frame of coastguard video sequence, (b) watermarked frame without HVS criteria (PSNR = 47.71 dB and SSIM = 0.999) and (c) watermarked frame with four HVS criteria (PSNR = 46.81 dB and SSIM = 0.994).

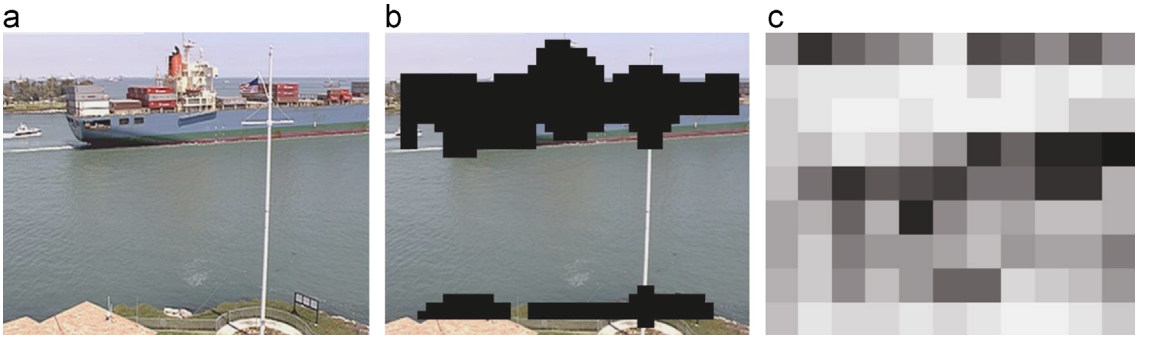

Fig. 13. (a) The original first frame of container video sequence, (b) ROI blocks (black blocks) and RONI (rest of frame) and (c) SSIM map represented by gray scale values, here brighter blocks have the higher SSIM values.

the HVS criteria is the result of the adaptive assignation of $S_{n,k}$ depending on characteristics of each video frame. An example of the visual quality distortion as a result of applying four HVS criteria is shown in Fig. 12.

From Fig. 12, we can observe that four HVS criteria provide an adequate watermark energy to obtain a watermarking scheme robust against aggressive operations without generating a visually quality distortion. According to (10) a higher distortion by watermarking process is caused to blocks belong to RONI, which means that quality distortion is not constant in a frame. To measure partial visual quality distortion, we introduce a SSIM map, in which each frame is divided into 32×32 non-overlapping blocks and then SSIM is computed using the watermarked and original frames. Fig. 13(a) shows the first frame of container video sequence, the blocks belong to the ROI, which are indicated by black blocks and the rest are RONI regions (Fig. 13(b)) and the SSIM map represented by gray scale values where brighter blocks have higher SSIM values (Fig. 13(c)). We can observe that blocks with higher values of SSIM are inside of observer's attention regions (ROI), which do not belong to the RONI.

5.3. Watermark robustness

The proposed watermarking algorithm was designed to resist legitimate operations, specially transcoding, and malicious attacks. In order to assess the embedded watermark robustness, the watermarked video sequences are attacked using some common signal processing and frame-based attacks. The robustness against video transcoding is evaluated changing all video properties, such as compression standard, bit-rate, spatial/temporal resolution as well as a combination of them. The robustness performance is compared with four recently proposed video watermarking methods [3–6]. These schemes are some of the most robust video watermarking algorithms with similar purpose that the proposed one. In order to do a fair comparison among these video watermarking methods, some criteria are considered: the quality distortion generated by each watermark embedding process is approximately same, which is 47 dB in the PSNR; additionally every scheme is evaluated with the same frame format (CIF) and the same 10 video sequences. The watermark robustness in terms of the BER represents an average performance of 10 video sequences.

5.3.1. Signal processing attacks

The watermarked video sequences are subject to signal processing attacks including noise contamination and Gaussian low-pass filtering and volumetric scaling. Fig. 14(a) and (b) shows the robustness against Gaussian and Impulsive noise contamination, respectively. Lee's scheme [3] shows a better performance than our method against noise contamination attacks, due to the redundant embedding of small amount of watermark bit (23 bits) sequence to all frames of the video sequence, while the performances of Chen's [4] and

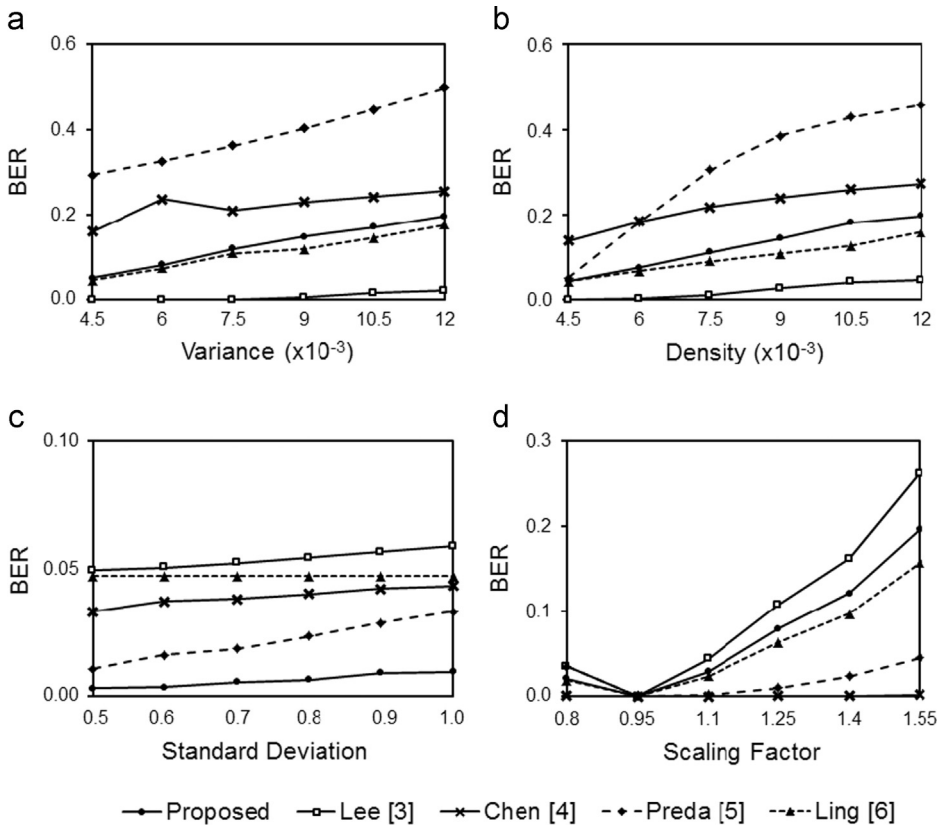

Fig. 14. Watermark robustness of the proposed and four video watermarking schemes against (a) Gaussian noise contamination, (b) impulsive noise contamination, (c) Gaussian low-pass filter and (d) volumetric scaling attacks.

Preda's [5] methods are severely damaged by noise contamination. Ling's method [6] provides a similar performance than the proposed one. In the proposed scheme, the average BER value is 0.197 when the quality degradations by noise contaminations are approximately 23.38 dB and 0.522 in terms of the PSNR and SSIM, respectively. Fig. 14(c) shows the robustness against Gaussian low-pass filtering attack, where the proposed scheme obtains the highest performance compared with those of four schemes [3–6]. The results presented in Fig. 14(d) shows the robustness against volumetric scaling attack, where the scaling factors are varied from 0.5 to 1.55. The highest volumetric scaling factor generates visual quality degradation of 20.75 dB and 0.611, respectively, in the PSNR and SSIM assessments. In the proposed scheme, an average BER value of 0.194 is obtained with the highest volumetric scaling factor.

5.3.2. Frame rate reduction

To evaluate robustness against the frame rate reduction, the frame rate of the watermarked video sequences is changed from 30 to 10 FPS using a transcoder [24]. Subsequently we extract the watermark sequence in order to analyze the robustness against this operation. Fig. 15 shows the watermark robustness against the frame rate reduction of the proposed scheme together with the performance of four previously reported schemes [3–6]. The proposed algorithm embeds the same watermark sequence along the frames in each video scene; so the embedded watermark is inherently robust against these attacks obtaining the BER value equals to 0.0005. Lee's method [3] has similar performance that the proposed one; however in this scheme the same watermark signal with same embedding energy is embedded at the same frequency band of all frames of video sequence [3], making it vulnerable against intra-collusion attacks. Ling's method [6] obtains relatively good performance against frame reduction task since the watermark signal is embedded redundantly in every I-frame; the disadvantage of this type of strategy against intra-collusion attacks has been described in Section 2. The methods [4,5] show vulnerability against this operation, since these methods embed the watermark signal along the temporal video information, becoming very sensible against frame rate reduction when the number of reduced frames is beyond their re-synchronization mechanism.

Additionally we evaluate the proposed scheme against frame-based attacks. A video sequence contains a lot of temporal redundancy, so frame-based attacks, such as frame dropping, frame averaging and frame swapping, are efficiently done to remove the watermark sequence, without causing significant quality degradation. In the proposed scheme, the same watermark sequence is embedded, with dynamic step size, into all frames of each scene, which prevents attackers from removing the watermark by frame dropping and frame swapping attacks. If attackers try to remove the

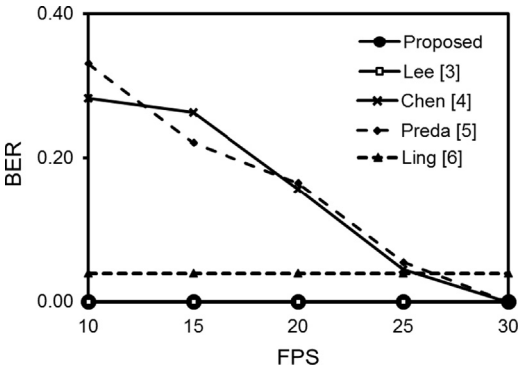

Fig. 15. Watermark robustness of the proposed and comparison schemes against Frame Rate Reduction.

Table 2
Robustness of the proposed scheme against conversion of video compression standard.

Compression standard/video container	Results report				
	Compression rate	Bit rate (Kb/s)	PSNR	SSIM	BER
MPEG-2/MPEG	1:18	4000	33.52	0.840	0.008
MPEG-4/MP4	1:18	3994	33.66	0.845	0.007
VC-1/ASF	1.27	3810	33.04	0.819	0.008
H.264 AVC/MP4	1:19	3810	33.61	0.845	0.009
VP6/FLV	1:27	3921	33.04	0.819	0.006

watermark they need to remove the whole scene causing a severe damage to the video. On the other hand, if an attacker collects a number of watermarked frames, the watermark sequence can be estimated using statistical averaging if the watermark energy is constant, and then it can be remove from the watermarked video [2]. As mentioned in Section 4, in the proposed scheme, the dynamic step size calculation based on the HVS generates different watermarking energies according to the spatio-temporal features of each frame. This adaptive watermark energy assignment makes hard to perform the frame averaging attack. Moreover, embedding different watermarks in each scene can efficiently avoid possible collusion attacks [25].

5.3.3. Change of bit-rate and video compression standard

In order to evaluate the robustness of the proposed scheme against the change of bit-rate and video compression standard, we use a transcoding [24] to modify these properties of the watermarked video sequence. First, to evaluate only robustness against change of the compression standard, the watermarked video sequences are encoded with bit-rate of 4 Mbps in all video compression standards mentioned in Section 2. Table 2 shows the results of this operation, where we can observe the video container, employed codec, the compression rate, the encoding bit-rate, the PSNR, SSIM and BER values of the extracted watermark sequence respect to the embedded one. From Table 2, we can observe that the proposed algorithm has an excellent performance when the watermarked video sequence is encoded in every compres-

Table 3
Robustness of the proposed and four video watermarking schemes against video compression standard and bit rate changing.

Compression standard/video container	Bit rate	BER				
		Proposed	Lee [3]	Chen [4]	Preda [5]	Ling [6]
MPEG-2/MPEG	1 Mbps	0.014	0.024	0.100	0.176	0.155
	2 Mbps	0.009	0.017	0.070	0.139	0.118
	4 Mbps	0.008	0.000	0.060	0.094	0.088
MPEG-4/MP4	128 Kbps	0.188	0.350	0.361	0.404	0.458
	256 Kbps	0.094	0.112	0.321	0.327	0.348
	512 Kbps	0.037	0.069	0.210	0.217	0.246
	1 Mbps	0.009	0.018	0.160	0.126	0.167
	2 Mbps	0.008	0.011	0.070	0.099	0.113
	4 Mbps	0.005	0.000	0.050	0.063	0.080
VC-1/ASF	128 Kbps	0.198	0.367	0.360	0.426	0.447
	256 Kbps	0.124	0.288	0.330	0.339	0.366
	512 Kbps	0.043	0.081	0.251	0.260	0.269
	1 Mbps	0.015	0.054	0.140	0.165	0.194
	2 Mbps	0.009	0.016	0.090	0.108	0.139
	4 Mbps	0.008	0.010	0.070	0.086	0.098
H264-AVC/MP4	128 Kbps	0.183	0.311	0.330	0.359	0.473
	256 Kbps	0.114	0.155	0.310	0.321	0.386
	512 Kbps	0.064	0.066	0.230	0.246	0.271
	1 Mbps	0.024	0.020	0.150	0.163	0.164
	2 Mbps	0.011	0.015	0.090	0.108	0.114
	4 Mbps	0.009	0.000	0.050	0.059	0.073
VP6/FLV	128 Kbps	0.197	0.333	0.361	0.402	0.482
	256 Kbps	0.110	0.247	0.341	0.359	0.357
	512 Kbps	0.040	0.088	0.251	0.259	0.234
	1 Mbps	0.013	0.070	0.181	0.197	0.104
	2 Mbps	0.007	0.014	0.130	0.137	0.086
	4 Mbps	0.006	0.010	0.050	0.096	0.063

sion standards. In all cases, the BER values are sufficiently low, which allows a correct detection of the embedded watermark after conversion of the compression standards listed in Table 2.

Continuously, in order to evaluate the watermark resilience of the proposed method against the change of compression bit-rate, the watermarked video sequences are encoded to lower bit-rates, such as 128 Kbps, 256 Kbps, 512 Kbps, 1 Mbps, 2 Mbps and 4 Mbps, using five compression standards mentioned above. Table 3 presents the performance of the proposed scheme together with these of four video watermarking schemes [3–6]. It is worth noting that the results for MPEG-2 video conversion are reported until 1 Mbps, because MPEG-2 is not optimized for bit-rates lower than 1 Mbps. This table shows a good performance of the proposed method when the watermarked video sequence is encoded with six different bit-rates, obtaining small values of the BER, which guarantees a correct watermark extraction after the compression with lower bit-rates. Ling's method [6] is not robust to the compression with bit-rates lower than 512 Kbps for all compression standards. The principal reason of this low performance is the instability of the relevant points detected by Harris-Affine detector from the quality degraded video caused by compression with lower bit rates. The BER values obtained for Chen's[4] and Preda's [5] methods with low bit rates equal to 256 Kbps and 128 Kbps for all compression standards suggest watermark vulnerability of these schemes. Lee's method [3] shows a good performance with bit rates higher than 512 Kbps, however the performance decreases with 128 Kbps for all compression standards and with 256 Kbps for VC-1 and VP6 compression standards.

5.3.4. Spatial resolution

The robustness of the proposed and four video watermarking schemes [3–6] against the change of spatial resolution is evaluated by converting from CIF (352 × 288 pixels) to several other standard video formats with different spatial resolutions, such as SQCIF (128 × 96 pixels), QCIF (176 × 144 pixels), 4CIF (704 × 576 pixels) and 16CIF (1408 × 1152 pixels) video formats. The evaluation results are shown in Fig. 16, which shows a good performance of the proposed scheme obtaining the BER values lower than 0.2 for change to all spatial resolution formats. The robustness performance of Lee, Chen and Preda's methods [3–5] is affected when the original frame is changed to SQCIF and QCIF, i.e. with a scaling factor less than 1, in which some spatial information is lost and the embedded watermark information cannot be extracted correctly. Lee et al. reported in [3] a good performance in the spatial resolution changes from HDTV to QVGA, however the resolution changes to 4CIF and 16CIF formats, which are smaller than QVGA, cause performance degradation. Ling's method [6] is affected by both downscaling and up-scaling; because the Harris-Affine detector is affected by scaling.

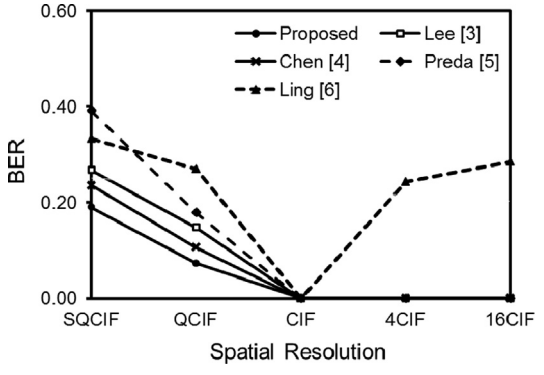

Fig. 16. Robustness against changing spatial resolution.

Table 4
Watermark robustness of the proposed and four video watermarking schemes against homogeneous transcoding.

Frame rate	Spatial resolution	BER				
		Proposed	Lee [3]	Chen [4]	Preda [5]	Ling [6]
30 FPS	640 × 480	0.001	0.000	0.000	0.000	0.040
	480 × 270	0.010	0.020	0.032	0.072	0.073
	320 × 240	0.028	0.047	0.057	0.097	0.105
24 FPS	640 × 480	0.001	0.000	0.072	0.054	0.041
	480 × 270	0.010	0.021	0.104	0.154	0.074
	320 × 240	0.028	0.043	0.149	0.202	0.105
15 FPS	640 × 480	0.001	0.000	0.300	0.271	0.040
	480 × 270	0.010	0.022	0.408	0.325	0.073
	320 × 240	0.028	0.032	0.489	0.416	0.105

5.3.5. Evaluation in practical situations

In practice, video transcoding is processed as a combination of changes of compression standard, bit-rate, spatial and temporal resolution. To measure the robustness of the proposed method in practical situations, we simulated homogeneous and heterogeneous video transcoding by combining changes of compression standard, frame rate, spatial resolution and bit-rate. Table 4 shows the watermark robustness of the proposed and four video watermarking methods [3–6] in the homogeneous transcoding case, in which a frame rate reduction and spatial resolution changes occur simultaneously, from the watermarked CIF video with the 30 FPS. The compression standard for all video sequences is unchanged, which is MPEG-4. The spatial resolutions listed in Table 4 are commonly used in PC monitor and some cellular phone touch screens. From this table we can see that the performance of the proposed method is slightly better than that of Lee's and Ling's methods [3,6], in which we observe that in the proposed scheme, the BER is not affected by frame rate reduction and only the spatial resolution changes degrades slightly the robustness performance. The combination of frame rate reduction and spatial resolution causes severe damage over Chen's and Preda's methods [4,5], because loss of temporal synchronization causes problem of the watermark extraction in their methods.

Another common operation in practice is to change the bit rate, spatial resolution and frame rate parameters as a result of conversion between common profiles of compression standards, which is typical case of heterogeneous transcoding. An example of this task is a video encoded over MPEG-1 to be converted to MPEG-2 (Main profile) and MPEG-4 SP (Simple profile). Unlike MPEG-2, MPEG-4 SP

Table 5
Comparison of the watermark robustness of proposed scheme and four video watermarking schemes against a typical heterogeneous transcoding applied to watermarked video with MPEG-1, 15 Mbps, 30 FPS and CIF format.

Target video	Properties	BER				
		Proposed	Lee [3]	Chen [4]	Preda [5]	Ling [6]
MPEG-2 Main Profile	15 Mbps, 30 FPS, 480 × 270 pixels	0.001	0.000	0.002	0.005	0.052
MPEG-4 Simple Profile	8 Mbps, 15 FPS, 176 × 144 pixels	0.083	0.157	0.417	0.410	0.292

does not support frames type B or interlaced video frames, additionally MPEG-4 SP works with lower spatial resolution and lower frame rate than MPEG-2. To simulate this heterogeneous transcoding case, first a watermarked video sequence is encoded by MPEG-1 compression standard with 15 Mbps, 30 FPS and CIF (352×288) format. Next this watermarked video is transcoded generating two video sequences; the first one with MPEG-2 main profile, i.e. a spatial resolution of 480×270 pixels, a bit rate of 15 Mbps and 30 FPS; and the second one with MPEG-4 SP, i.e. a spatial resolution of 176×144 pixels, a bit rate of 8 Mbps and 15 FPS. The watermark sequences are extracted from these two watermarked videos and their respective results are shown in Table 5. From this table we can conclude that the proposed video watermarking scheme could detect successfully the embedded copyright information after an aggressive heterogeneous transcoding, showing better performance compared with four video watermarking schemes [3–6]. It is worth noting that the target video properties after heterogeneous transcoding, such as spatial resolutions, the frame rates and bit-rates are common in the practical situation, although the watermark sequence in the proposed scheme will be survived against more aggressive cases as shown in Sections 5.3.2–5.3.4.

5.4. Computational complexity

It is important to consider the possibility of applying the proposed watermarking scheme under the real-time requirement. In the proposed scheme, the watermarking is performed in the raw video data without considering any special compression standard to obtain the watermark robustness against transcoding. However, generally base-band watermarking schemes have disadvantage compared with compressed domain watermarking schemes when real-time operation is desired, because time consuming 2D-DCT and inverse 2D-DCT computation are required in base-band watermarking scheme.

In the proposed video watermarking scheme, the watermark embedding and extraction are performed in blocks of 8×8 2D-DCT coefficients, and the computation of the dynamic step size of the QIM algorithm, except visual attention region estimation, is also performed directly in the blocks of 8×8 2D-DCT coefficients. Therefore, if visual attention regions are estimated directly in the 2D-DCT domain using fast inter-transformation between 2D-DCT block and its sub-block [26], the proposed video watermarking scheme can be operated in the compressed domain, avoiding 2D-DCT and inverse 2D-DCT operations. The idea of use of the fast inter-transformation of 2D-DCT blocks is the same with the compressed domain video watermarking schemes proposed in [3,6]. The computing times required in watermark embedding and extraction process, realizing the above mentioned strategy, are approximately 4.47 and 4.39 s, respectively. These computing times are obtained by Matlab ver. R2010a in Intel Core-i5 2.5 GHz.

6. Conclusions

In this paper, we proposed a video watermarking technique robust against several signal processing distortions, frame-based attacks and especially video transcoding. To improve robustness and accuracy in the detection, and at the same time obtain a good quality of video sequences, the watermark sequence was embedded and detected using a quantization index modulation (QIM) algorithm with adaptive step size, which is calculated using four HVS-based criteria, such as texture and luminance masking, a motion threshold and visual attention region. The experimental results show fairly good watermark imperceptibility since the obtained PSNR values are near to 47 dB and the SSIM is close up to 1, taking in account that the SSIM index is a good indicator of the perceptual quality degradation. The experimental results show the watermark robustness to the five most common video compression standards encoding, such as MPEG-2, MPEG-4, H.264 AVC, VP6 and VC-1 with different bit rates, also the embedded watermark is robust to signal processing, intentional video frame-based attacks and changing spatial resolution. The watermark robustness is evaluated in two practical situations, which are homogeneous and heterogeneous transcoding cases. In both cases, the BERs of the proposed scheme are smaller than 0.1, which provides a reliable copyright protection over the digital video contents in the practical situations. The robustness performance of the proposed scheme is compared with four recently reported robust video watermarking schemes [3–6] under the same conditions. The comparison results show the better performance of the proposed scheme especially in heterogeneous transcoding, which is applied commonly in our daily life.

Acknowledgments

The authors thank the National Council of Science and Technology (CONACYT) of Mexico, the National Polytechnic Institute and the National Autonomous University of Mexico (UNAM) for financial support of this work.

References

[1] A. Vetro, C. Christopoulosa, H. Sun, Video transcoding architectures and techniques: an overview, IEEE Signal Processing Magazine 20 (2003) 18–29.

[2] G. Doërr, J.L. Dugelay, A guide tour of video watermarking, Signal Processing: Image Communication 18 (2003) 263–282.

[3] M.J. Lee, D.H. Im, H.Y. Lee, K.S. Kim, H.K. Lee, Real-time video watermarking system on the compressed domain for high-definition video contents: practical issues, Digital Signal Processing 22 (2012) 190–198.

[4] H. Chen, Y. Zhu, A Robust Video Watermarking Algorithm Based on Singular Value Decomposition and Slope-Based Embedding Technique, Multimedia Tools Applications, Springer http://dx.doi.org/10.1007/s11042-012-1238-2.

[5] R. Preda, D.N. Vizireanu, Robust wavelet-based video watermarking scheme for copyright protection using the human visual system, Journal of Electronic Imaging 20 (2012) 013022-1–013022-8.

[6] H. Ling, L Wang, F. Zou, Z. Lu, P. Li, Robust video watermarking based on affine invariant regions in the compressed domain, Signal Processing 91 (2011) 1863–1875.

[7] R.B. Wolfgang, C.I. Podilchuk, E.J. Delp, Perceptual watermarks for digital images and video, Proceedings of the IEEE 87 (1999) 1108–1126.

[8] H.S. Jung, Y.Y. Lee, S.U. Lee, RST-resilient video watermarking using scene-based feature extraction, EURASIP Journal of Applied Signal Processing 14 (2004) 2113–2131.

[9] Z. Liu, H. Liang, X. Niu, Y. Yang, A robust video watermarking in motion vectors, in: Proceedings of the 7th International Conference on Signal Processing, 31 August–4 September 2004, Beijing, China, 2004, pp. 2358–2361.

[10] M. Noorkami, R.M. Mersereau, Improving perceptual quality in video watermarking using motion estimation, in: IEEE International Conference on Image Processing, 8–11 October 2006, Atlanta, GA, USA, 2006, pp. 520–523.

[11] S. Biswas, S.R. Das, E.M. Petriu, An adaptive compressed MPEG-2 video watermarking scheme, IEEE Transactions on Instrumentation and Measurement 54 (2005) 1853–1861.

[12] A.M. Alattar, E.T. Lin, M.U. Celik, Digital watermarking of low bit-rate advanced simple profile MPEG-4 compressed video, IEEE Transactions on Circuits and Systems for Video Technology 13 (2003) 787–800.

[13] B. Chen, G. Wornell, Quantization index modulation: a class of provably good method for digital watermarking and information embedding, IEEE Transaction on Information Theory 47 (2001) 1423–1443.

[14] Y. Y. Zhao, Y. Zhao, Improved quantization watermarking with an adaptive quantization step size and HVS, in: Knowledge-Based Intelligent Information and Engineering Systems, Lecture Notes in Computer Science, vol. 3681, 2005, pp. 1212–1218.

[15] Q. Li, I.J. Cox, Using perceptual models to improve fidelity and provide resistance to volumetric scaling for quantization index modulation watermarking, IEEE Transactions on Information Forensics and Security 2 (2007) 127–139.

[16] I. Ahmad, X. Wei, I. Ahmad, X. Wei, Y. Sun, Y.Q. Zhang, Video transcoding: an overview of various techniques and research issues, IEEE Transactions on Multimedia 7 (2003) 353–364.

[17] J. Xin, C.W. Lin, M.T. Sun, Digital video transcoding, Proceedings of the IEEE 93 (2005) 84–97.

[18] T. Shanableh, M Ghanbari, Heterogeneous video transcoding to lower spatio-temporal resolutions and different encoding formats, IEEE Transactions on Multimedia 2 (2000) 101–110.

[19] H.Y. Tong, A. N. Venetsanopoulos, A perceptual model for JPEG applications based on block classification, texture masking and luminance masking, in: Proceedings of International Conference on Image Processing, 4–7 October 1998, Chicago, USA, 1998, pp. 428–432.

[20] Y. Jia, W. Lin, A.A. Kassim, Estimating just-noticeable distortion for video, IEEE Transactions on Circuits and Systems For Video Technology 16 (2006) 820–829.

[21] N. Bruce, J. Tsotsos, Saliency, attention and visual search: an information theoretic approach, Journal of Vision 9 (2009) 1–24.

[22] O. Le Meur, P. Le Callet, D. Barba, D. Thoreau, A coherent computational approach to model bottom-up visual attention, IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (2006) 802–817.

[23] ⟨http://trace.eas.asu.edu/yuv/index.html⟩.

[24] ⟨http://www.mediacoderhq.com/⟩.

[25] P. Chan, M. Lyu, A DWT-based digital video watermarking scheme with error correcting code, in: Proceedings of the 5th International Conference of Information and Communications Security, 10–13 October, Huhehaote City, China, 2003, pp. 202–213.

[26] B. Davis, S. Nawab, The relationship of transform coefficients for differing transforms and/or differing subblock sizes, IEEE Transactions on Signal Processing 52 (2004) 1458–1461.