



**HAL**  
open science

## MPEG-4 AVC stream watermarking by m -QIM techniques

Marwen Hasnaoui, Maher Belhaj Abdallah, Mihai Mitrea, Françoise Preteux

► **To cite this version:**

Marwen Hasnaoui, Maher Belhaj Abdallah, Mihai Mitrea, Françoise Preteux. MPEG-4 AVC stream watermarking by m -QIM techniques. *Electronic Imaging 2011: Multimedia on Mobile Devices 2011; and Multimedia Content Access: Algorithms and Systems*, Jan 2011, San Francisco, United States. pp.78810L, 10.1117/12.876282 . hal-00848646

**HAL Id: hal-00848646**

**<https://hal.science/hal-00848646>**

Submitted on 26 Jul 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# MPEG-4 AVC stream watermarking by $m$ -QIM techniques

M. Hasnaoui\*, M. Belhaj<sup>1</sup>\*, M. Mitrea\*, F. Preteux\*\*

\* Institut Télécom ; Télécom SudParis, ARTEMIS Department

\*\* MINES ParisTech

## ABSTRACT

The present paper is devoted to the MPEG-4 AVC (a.k.a. H.264) video stream protection by means of watermarking techniques. The embedding process is carried out in quantized index domain and relies on the  $m$ -QIM ( $m$ -array Quantisation Index Modulation) principles. In order to cope with the MPEG-4 AVC peculiarities, the Watson's perceptual model is reconsidered and discussed. The experimental results correspond to the MEDIEVALS (a French National Project) corpus of 4 video sequences of about 15 minutes each, encoded at 512kbps. The transparency is assessed by both subjective and objective measures. The transcoding (down to 64kbps) and geometric (StirMark) attacks result in BER of 6.75% and 11.25%, respectively. In order to improve robustness, an MPEG-4 AVC syntax-driven counterattack is considered: this way, the two above mentioned attacks lead to BER of 2% and 10%, respectively. Finally, the overall theoretical relevance of these results is discussed by estimating the related channel capacities.

**Keywords:** MPEG-4 AVC watermarking,  $m$ -array QIM; transcoding, geometric attacks.

## I. INTRODUCTION

In order to protect ownership and to ensure piracy tracking-down for digital video, watermarking solutions<sup>1,2</sup> can be considered. A watermark represents some embedding digital information into a cover content (an MPEG-4 AVC stream, in the particular case of this paper). In order to be effective, the watermark should be perceptually invisible (*transparent*) and *robust* against the attacks; moreover, it should allow the insertion of a sufficient amount of information (*e.g.* a serial number identifying a user).

In order for the watermarking techniques to be easily integrated into practical applications, an additional constraint is imposed: they should not be computationally complex, *i.e.* they should not require sophisticated operations, like decoding/re-encoding, spectral representations, *etc.* One way of achieving such a desideratum is to consider compressed domain watermarking techniques, *i.e.* techniques inserting the mark directly into the compressed stream.

Nowadays, the state-of-the art cannot bring into light any MPEG-4 AVC watermarking method reaching the trade-off among transparency, robustness and data payload<sup>3-5</sup>; however, these three issues are individually addressed with good results.

D. Zou and J. A. Bloom<sup>3</sup> introduce a substitution watermarking method for MPEG-4 AVC stream. While ensuring a very fast technique, the paper doesn't evaluate robustness against attacks.

A. Golikeri, P. Nasiopoulos and Z. J. Wang<sup>4</sup> propose an ST-QIM (Spread Transform -QIM) watermarking method. Although improving the performances of the traditional ST-DM, this method features a quite small data payload (one bit per macro-block) and has no robustness against the geometric attacks.

M. Noorkami advances a correlation-watermarking method<sup>5</sup> based on perceptual masking principles: the mark is a simple bipolar message inserted according to a psycho-visual cost. This method provides a very good transparency with a high data payload (3 times better than existed methods). Its main weakness point remains the robustness, even the mundane transcoding attack being able to destroy the mark.

M. Belhaj, M. Mitrea, S. Duta and F. Preteux<sup>6</sup> introduce a binary ST-QIM for MPEG-4 AVC stream watermarking. By combining ST-DM principles, a perceptual shaping mechanism, and an information-theory driven selection criterion,

---

<sup>1</sup> Corresponding author: Maher Belhaj, PhD Student, 9, rue Charles Fourier, 91011 Evry, France. Phone: + 33 1 60 76 45 66, Fax: + 33 1 60 76 43 81. maher.belhaj\_abdallah@it-sudprais.eu, www.it-sudparis.eu/artemis

they achieved a good transparency and robustness against transcoding and geometric attacks. These principles are extended in M. Hasnaoui, M. Belhaj, M. Mitrea and F. Preteux<sup>7</sup>; in this respect,  $m$ -arry ST-QIM techniques are considered, thus resulting in an data payload increased by a factor of  $\log_2 m$ .

The present paper reconsiders and extends the authors previous studies reported in<sup>7</sup>. In this respect, in addition to the  $m$ -arry ST-QIM method (summarised in Section II), a perceptual shaping model, devoted to the MPEG-4 AVC peculiarities is discussed (Section III). The watermarking method combining  $m$ -QIM and perceptual shaping mechanisms is presented in Section IV and the related experimental results in Section V. Section V also advances an MPEG-4 AVC syntax-driven counterattack and establishes the overall theoretical relevance of the paper by estimating the related channel capacities. Conclusions are drawn and perspectives are opened in Section VI.

## II. $m$ -arry ST-QIM principles

### A. Binary QIM

In quantization based insertion techniques, the host data  $x$  is quantized using multiple quantizers and the index of the quantizer is chosen based on the message to be embedded<sup>8</sup>. To build quantizers, dither modulation (DM) can be used. For a host signal  $x$  and a binary message  $d$ , the watermarked signal sample  $y$  is computed by:

$$\begin{cases} q = Q_{\Delta}(x - \Delta(d/2 + k)) - x + \Delta(d/2 + k) \\ y = x + \alpha q \end{cases} \quad (1)$$

where  $\Delta$  is a fixed quantization step size,  $k$  a random key and  $\alpha$  a fixed parameter,  $0 < \alpha \leq 1$ . The quantizer  $Q_{\Delta}$  is defined as follows:

$$Q_{\Delta}(x) = \Delta \text{Round}\left(\frac{x}{\Delta}\right). \quad (2)$$

At the decoder, the embedded message bit is recovered by a scalar quantization of the received signal sample,  $r$  (a corrupted version of  $y$ ).

The  $Y(d)$  detection variable is computed as follows:

$$Y(d) = Q_{\Delta}(r - kd) - r + k\Delta. \quad (3)$$

The decision is made according to the following rule:

$$\begin{cases} \text{if } |Y(d)| < (1 - \alpha)\Delta/2, & \hat{d} = 0 \\ \text{if } |Y(d)| \geq (1 - \alpha)\Delta/2, & \hat{d} = 1 \end{cases} \quad (4)$$

Note that one message bit is embedded in one sample.

### B. Spread transform dither modulation

ST-DM is a particular form of QIM. The watermark is not directly embedded into the original signal  $x$  but into the projection  $x'$  of  $x$  onto a randomly generated normalized vector  $u$ . This technique is called Spread Transform as proposed by Chen and Wornell<sup>8</sup>. The resulting scalar value is then quantized before being added in the component of the signal:

$$\begin{cases} q = Q_{\Delta}(x'u - \Delta(d/2 + k)) - x'u + \Delta(d/2 + k) \\ y = x + (\alpha q)u \end{cases} \quad (5)$$

and the corresponding detection is given by equation (4) where the detection function  $Y(d)$  is:

$$q = Q_{\Delta}(r'u - \Delta k) - r'u + \Delta k. \quad (6)$$

ST-DM is more robust to re-quantization (*e.g.* JPEG compression) than regular QIM. However, ST-DM introduces relatively higher perceptual distortions. To reduce perceptual distortion, perceptual models can be considered.

### C. ST-*m*QIM extension

The ST-*m*QIM is an extension of regular ST-DM. Be there a binary message to be inserted; instead of directly inserting it, a message  $d$  encoded into an  $m$ -ary alphabet  $D = \{-(m-1)/2, -(m-2)/2, \dots, 0, (m-2)/2, (m-1)/2\}$  is considered, with the aim of increasing the data payload by a factor  $\log_2(m)$ .

The insertion rule is given by (7):

$$q = Q_\Delta(x^t u - \Delta(d/m + k)) - x^t u + \Delta(d/m + k). \quad (7)$$

While keeping practically the same insertion rule, see (7) vs. (5), this amendment will lead us to modify the decision rule given in (6). The decision rule<sup>7</sup> optimises the probability error under the additive noise hypothesis. Note that the decision is based on the value of  $Y(d)$  which is a quantization error belonging to the  $[-\Delta/2, \Delta/2]$  interval; hence, specifying a decision rule means to divide the decision region  $[-\Delta/2, \Delta/2]$  into  $m$  nonoverlapping intervals.

$Y(d)$  can be written as follows:

$$\begin{aligned} q &= Q_\Delta(B(d)) - B(d) \\ B(d) &= (\alpha - 1)q + \Delta \frac{d}{m}. \end{aligned} \quad (8)$$

Given  $q$  a quantization error lying in the  $[-\Delta/2, \Delta/2]$  interval, then

$$\frac{\Delta((\alpha - 1)m + 2d)}{2m} \leq B(d) \leq \frac{\Delta((1 - \alpha)m + 2d)}{2m}. \quad (9)$$

Be there  $I_{\text{sup}(d)} = \frac{\Delta((1 - \alpha)m + 2d)}{2m}$  and  $I_{\text{inf}(d)} = \frac{\Delta((\alpha - 1)m + 2d)}{2m}$ .

To avoid overlapping, there will be one decision interval for each element of the alphabet; hence,  $Q_\Delta(B(d)) = 0$  should take eventually one value. Be there  $Q_\Delta(B(d)) = 0$ . Then, we have:

$$\begin{cases} I_{\text{sup}(d)} < \frac{\Delta}{2} \\ \frac{\Delta}{2} < I_{\text{inf}(d)} \end{cases}. \quad (10)$$

From (8), (9) and (10) we have:

$$\begin{aligned} -I_{\text{sup}(d)} &\leq B(d) \leq I_{\text{inf}(d)} \\ \alpha &> \frac{m-1}{m}. \end{aligned} \quad (11)$$

For a fixed parameter  $\alpha$ ,  $I_{\text{sup}(d)}$  and  $I_{\text{inf}(d)}$  are increasing functions of  $m$ . Hence, if each two successive symbols ( $d$ ,  $d+1$ ) have no overlapping decision intervals, then we will have  $m$  no overlapping decision intervals. In this respect,  $I(d)$  and  $I(d+1)$  must verify the equation (12):

$$I_{\text{sup}(d)} - I_{\text{inf}(d+1)} < 0. \quad (12)$$

Equation (12) implies that:  $\alpha > \frac{m-1}{m}$  (our assumption in (11)).

Therefore, the optimal value of  $\alpha$  is  $\alpha^* = \frac{m-1}{m}$ .

We note that if  $Y(m)$  is between two decision regions, we decide for the region which is closest to  $Y(m)$ .

Example: For  $m = 5$  we have  $\alpha^* = 0.8$  and the corresponding decision regions are illustrated in Figure 1.

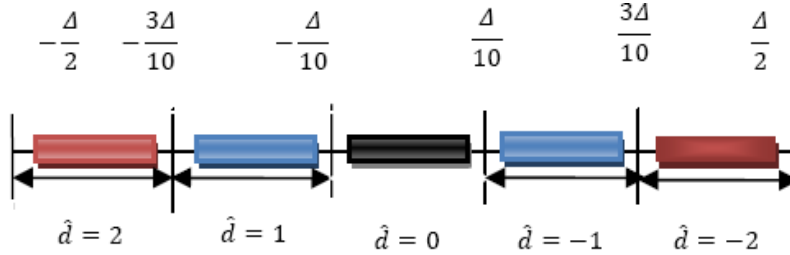


Figure 1. Decision regions for  $m = 5$  and  $\alpha \geq \alpha^*$ .

### III. PERCEPTUAL MASK

In its largest acceptance, perceptual masking consists in matching the mark to the human visual system peculiarities.

The first model of the psycho-visual metric was proposed in 1983, when Waton<sup>12</sup> wrote “*What does the eye see the best?*”, thus establishing a model mapping the visual signal perceived by a human observer into an internal representation on which the observer’s decision is based.

In 1990, Peterson<sup>13</sup> proposed the IIP (Image-Independent Perceptual) approach as an experimental estimation of the visibility threshold in DCT transform coefficients.

Three years later, an improvement of the IIP matrix was presented by Watson<sup>11</sup>. This way, the contrast and the global luminance influence to the visibility threshold was considered in order to establish a new three components perceptual mask (the so-called *8x8 Watson contrast sensitivity table*). This mask has already been used for watermarking in uncompressed or compressed domain, aiming at the increase of the data payload and at the improvement of the robustness, while limiting the visual distortions. However, as this mask was designed for classical 8x8 DCT, it can be directly exploited in MPEG-2 applications but has no straightforward adaptation for the MPEG-4 AVC framework. Actually, from this point of view, MPEG-4 AVC comes across with three new issues. First, the DCT is no longer applied to 8x8 blocks but to 4x4 blocks. Secondly, the DCT itself is different, MPEG-4 AVC considering a transform designed so as to integers. Finally, this integer DCT is no longer applied to pixels (as it is the case in JPEG and MPEG-2) but to the intra/inter prediction errors.

In order to avoid the first type of mismatch, M. Noorkami [5] considered the basic IIP matrix and derived its scaled and sub-sampling version, suitable for 4x4 DCT. The studies reported in<sup>6,7</sup> also consider the second and the third issues, as follows.

#### A. Perceptual shaping for integer DCT applied to pixels

The principle is to start from the Noorkami (sub-sampled) IIP matrix and to adapt it according to the dependencies between the DCTs involved in the MPEG-2 and MPEG-4 AVC. In other words, we shall consider that modifications induced in the pixels have the same effects and we shall write the equations précising how these modifications are translated into different MPEG-2 and MPEG-4 AVC syntax elements.

The Noorkami matrix (further denoted by  $T$ ) is presented in (13):

$$T = \begin{bmatrix} 0.7 & 0.58 & 1.2 & 2.39 \\ 0.58 & 1.12 & 1.49 & 2.3 \\ 1.2 & 1.49 & 3.07 & 4.35 \\ 2.39 & 2.3 & 4.35 & 7.25 \end{bmatrix} \quad (13)$$

A value in the  $T$  matrix represents the visibility threshold, *i.e.* the maximal value of a distortion added on a pixel (classical) DCT coefficient which is still transparent (imperceptible) for a human observer.

The integer DCT (considered in MPEG-4 AVC) was described so as to reduce the codec complexity with respect to the MPEG-2 classical (floating point) DCT. The standard way of applying the integer and the classical transforms are given by (14) and (15), respectively:

$$Y = C \cdot X \cdot C^t \otimes E, \quad (14)$$

$$C = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{bmatrix}, \quad E = \begin{bmatrix} 0.25 & 0.158 & 0.25 & 0.158 \\ 0.158 & 0.1 & 0.158 & 0.1 \\ 0.25 & 0.158 & 0.25 & 0.158 \\ 0.158 & 0.1 & 0.158 & 0.1 \end{bmatrix},$$

where:

- $X$  is an original 4x4 matrix, while  $Y$  is the transform matrix,
- “ $\cdot$ ” is the matrix product,
- “ $\otimes$ ” is the scalar product.

$$Y = A \cdot X \cdot A^t, \quad (15)$$

$$C = \begin{bmatrix} 0.5 & 0.5 & 0.5 & 0.5 \\ 0.653 & 0.271 & -0.271 & -0.653 \\ 0.5 & -0.5 & -0.5 & 0.5 \\ 0.271 & -0.653 & +0.653 & -0.271 \end{bmatrix},$$

where  $X$  is an original 4x4 matrix, while  $Y$  is the transform matrix.

By combining (13), (14) and (15), the perceptual mask for the MPEG-4 AVC DCT can be obtained, (16):

$$T_{AVC} = C \cdot (A^t \cdot T \cdot A) \cdot C^t \otimes E. \quad (16)$$

From the intuitive point of view, (16) is illustrated in Figure 2: a value in  $T_{AVC}$  gives the new visibility threshold, *i.e.* the maximal value of a distortion added on an integer DCT coefficient which is still transparent (imperceptible) for a human observer.

Note that the matrix in (16) corresponds to the distortions induced in the pixel values and not in the prediction errors, as it is the real case in the MPEG-4 AVC.

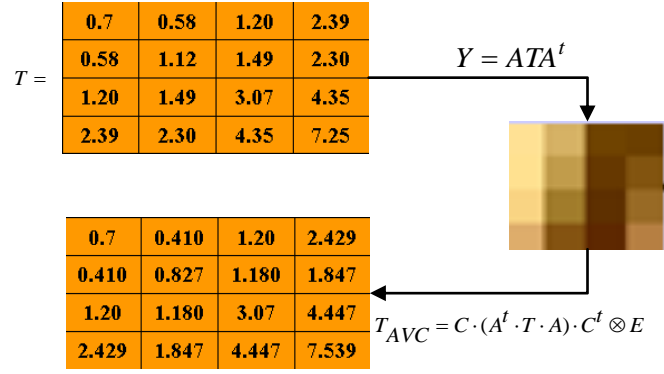


Figure 2. MPEG-4 AVC adaptation of Watson perceptual mask.

### B. Perceptual shaping for integer DCT applied to prediction errors

The MPEG-4 AVC standard considers the integer DCT transform applied to the prediction errors between pixels and their neighbors. Hence, in the watermarking practice, we should take into consideration a perceptual mask describing the maximal value of a distortion added to a prediction error which is still transparent for a human observer. In order to compute such a perceptual mask, the standard MPEG-4 AVC prediction modes should be considered. Actually, the standard makes provision for 9 different prediction modes: Vertical, Horizontal, DC, Horizontal/Up, Vertical/Left, Horizontal/Down, Vertical/Right, Diagonal down/Right, Diagonal down/Left. However, in practice, the most frequent ones are the vertical, horizontal and DC modes (as illustrated in Figure 3).

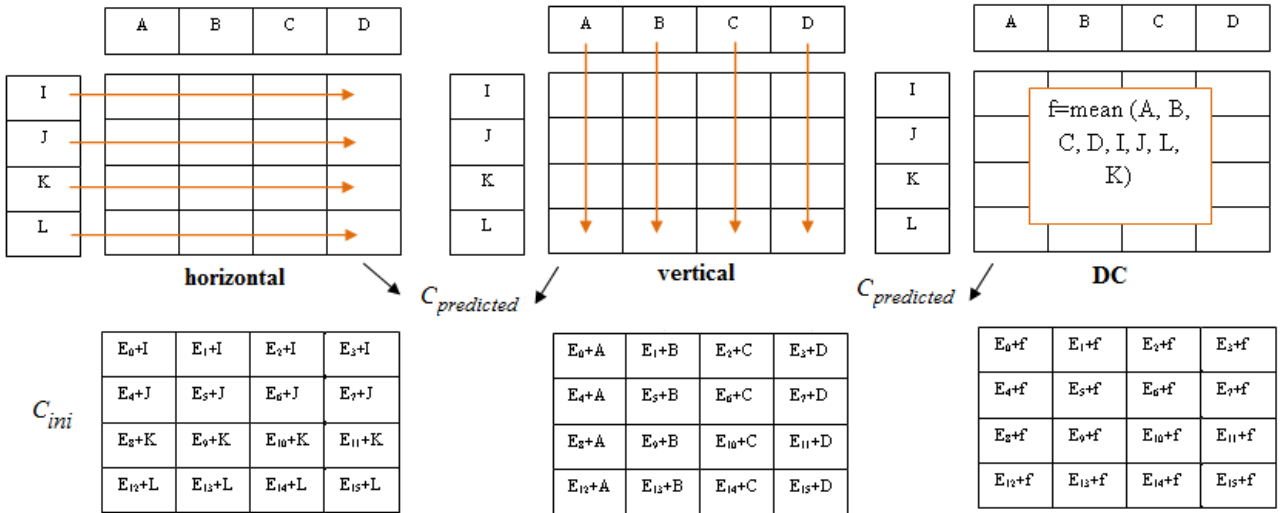


Figure 3. Most frequent used prediction modes: horizontal (left), vertical (middle) and DC (right). (A..D, I..L) adjacent pixels,  $C_{ini}$  is the matrix the current block coefficients,  $E$  is the matrix of the integer DCT prediction errors and  $C_{predicted}$  is the matrix of the predicted coefficients.

$$E = C_{ini} - C_{predicted} = \begin{bmatrix} E_0 & E_1 & E_2 & E_3 \\ E_4 & E_5 & E_6 & E_7 \\ E_8 & E_9 & E_{10} & E_{11} \\ E_{12} & E_{13} & E_{14} & E_{15} \end{bmatrix}. \quad (17)$$

The perceptual mask matrix in the integer DCT domain, applied to the prediction error (further denoted as  $T_{AVC\_pred}$ ) is defined as:

$$T_{AVC\_pred} = \max_{transparency} (mark) = \max_{transparency} (E_{marked} - E_{original}). \quad (18)$$

where an additive mark insertion in the integer DCT of the prediction error has been considered. When the vertical prediction mode is employed, (19) can be rewritten as:

$$T_{AVC\_pred} = \max_{transparency} \left( C_{ini\_marked} - \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} C_{ref\_UP} - C_{ini} + \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} C_{ref\_UP} \right). \quad (19)$$

$$C_{ref\_UP} = \begin{bmatrix} A & B & C & D \\ D & E & F & G \\ H & I & J & K \\ L & M & N & O \end{bmatrix} \quad C_{ref\_LEFT} = \begin{bmatrix} I & M & Q & U \\ J & N & R & V \\ K & O & S & W \\ L & P & T & X \end{bmatrix}$$

$$T_{AVC\_pred} = \max_{transparency} (C_{ini\_marked} - C_{ini}) = T_{AVC}. \quad (20)$$

Equation (20) demonstrates that the perceptual masking matrix corresponding to the prediction error in the vertical mode is identical to the masking matrix in the coefficient domain. Actually, through analogy equations similar to (19) can be written for all the MPEG-4 AVC prediction modes, thus demonstrating the  $T_{AVC}$  matrix in (16) can be used for perceptual masking in the prediction error DCT domain.

#### IV. MULTI SYMBOL ST-MD TECHNIQUE FOR MPEG-4 AVC

The method presented in this paper inserts the mark in the AC luma coefficients of the 4\*4 sub-macroblocks in the I frames of MPEG-4 AVC stream; hence, the  $x$  vector in Section II stands for a 15 component vector obtained by zig-zag scanning such a sub-macroblock.

The mark is inserted after a block selection based on an energy criterion. In this respect only the blocks whose energies prior to and after the insertion verify the condition (21) are considered:

$$\mu_x - \frac{\sigma_x}{16} < 1 - \frac{\|x\|}{\|y\|} < \mu_x + \frac{\sigma_x}{16}. \quad (21)$$

where  $x$  and  $y$  are the original and the watermarked blocs,  $\| \cdot \|$  represents the energy of the block, while  $\mu_x$  and  $\sigma_x$  stand for the mean and standard deviation of the energy in the block.

##### A. Embedding process

The embedding process combines *ST-mQIM* and perceptual shaping. It is structured into three main modules as shown in Figure 4: *perceptual shaping*, *mark generation* and *mark embedding*.

##### *Perceptual shaping:*

The watermark is not directly embedded into the original 4\*4 block  $x$  but into the projection  $x'$  of  $x$  onto a perceptual mask, denoted by the  $v_{mask}$  vector; actually, the  $v_{mask}$  is obtained by scanning the  $T_{AVC}$  mask in a zig-zag order.

##### *Mark generation:*

The mark to be inserted into the host  $x$  vector depends of the m-array message  $d$  and of the original vector  $x$ :



$$q = Q_{\Delta}(x^t v_{mask} - \Delta(\frac{d}{m} = k)) - x^t v_{mask} + \Delta(\frac{d}{m} + k). \quad (22)$$

*Embedding mark:*

This module generates the watermarked 4\*4 MPEG-4 AVC block. It has as input the mark  $w$ , the original content and the perceptual mask  $v_{mask}$ . In the present paper the insertion follows a simple additive rule as follows:

$$y = x + w.v_{mask}. \quad (23)$$

### B. Detection process

For each supposed marked block, the detector starts by projecting the vector of 15 AC coefficients of the 4\*4 sub-macroblocks in the  $I$  frames onto the perceptual mask  $v_{mask}$ . Then the detection is achieved by following the rule described in Section II.

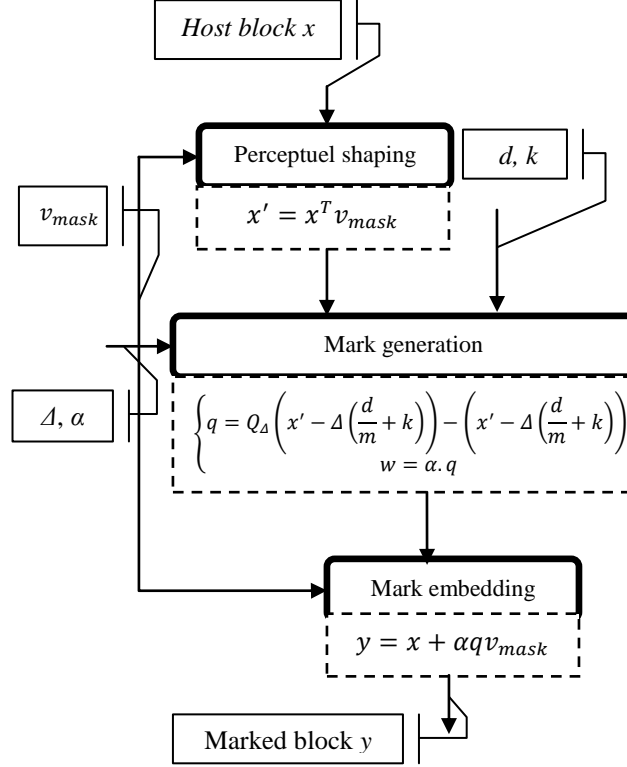


Figure 4. The embedding synopsis: three inputs (the message  $d$ , the host  $x$  and the key  $k$ ) and three parameters (the perceptual mask  $v_{mask}$ , the quantization step  $\Delta$  and the scaling factor  $\alpha$ ) are considered to compute the marked data  $y$ .  $m$  is the number of symbols in the message alphabet.

## V. EXPERIMENTAL RESULTS

The ST- $m$ QIM was implemented in the MPEG-4 AVC reference software version JM86 [14].

### A. Videos corpus

The video corpus consists of 4 video sequences of 15 minutes each. They were encoded in MPEG-4 AVC Baseline Profile (no B frames, CAVLC entropy encoder) at 512 kb/s. The GOP size is set to 10. The frame sizes were chosen according CIF specification (352\*288).

### B. Data payload

As (21) is an equation depending on the particular video sequence to be watermarked, the data payload also depend on the particular video sequence and cannot be a priori predicted. The data payload corresponding to the processed corpus averaged 130 bits/minute, *i.e.* 20 times larger then the minimal limit imposed by the DCI standard.

### C. Robustness

First, the robustness against additive noise was verified by considering a bipolar white noise (-1/1), added in the 4\*4 MPEG-4 AVC coefficient domains: each and every time, all the message symbols were correctly recovered.

Secondly, the robustness against transcoding is evaluated by compressing the watermarked files down to 64kb/s. It can be considered that ST-*m*QIM features robustness against this attack: an average value of 6.75% was obtained for the bit error rate (BER).

Finally, the robustness against geometric attacks was investigated. An average BER of 11.25% was obtained. In practical applications, in order not to be impeded by such a BER value, it would be preferable to connect the inserted message to some visual information, as illustrated in Figure 5.

From the information theory point of view, this robustness is connected to channels with high mutual information, as illustrated in Table I. The channel corresponding to the watermarking in this study has as input an  $m = 5$  letter alphabet  $\{-2, -1, 0, 1, 2\}$  and the output is represented by the decision according to the rule expressed in Section II. Table I illustrates the noise matrices and the related mutual information corresponding to this situation, estimated on the corresponding video sequences. The values thus obtained ( $I(X,Y)=2.02$ ,  $I(X, Y)= 1.2$ ) clearly point to robustness.

TABLE I. NOISE MATRICES FOR TRANSCODING (LEFT) AND GEOMETRIC (RIGHT) ATTACKS .

	TRANSCODING – $I(x,y)=2.02$					GEOMETRIC – $I(x,y)=1.02$				
	-2	-1	0	1	2	-2	-1	0	1	2
-2	0.90	0.02	0.02	0.01	0.05	0.77	0.11	0.02	0.01	0.09
-1	0.06	0.87	0.05	0.01	0.01	0.12	0.69	0.15	0.03	0.01
0	0	0.06	0.89	0.04	0.01	0.03	0.14	0.73	0.08	0.02
1	0.01	0.02	0.04	0.90	0.03	0.01	0.04	0.06	0.81	0.08
2	0.02	0.01	0.11	0.04	0.82	0.07	0.02	0.04	0.11	0.76
CAPACITY	2,13					2,04				



Figure 5. The original ARTEMIS (left) binary logo recovered after transcoding (middle) and geometric attacks (right).

### D. Transparency

In order to prove the transparency of the watermarking method, three types of metrics (Table II) have been considered: pixel difference-based measures (peak signal to noise ratio – PSNR, absolute average difference – AAD and peak mean square error – PMSE), correlation based measures (correlation quality – CQ, structural continent – SC and normalized cross correlation - NCC), and psycho-visual measures (digital video quality - DVQ). The average value of the measures corresponding to the five video sequences in the MEDIEVALS corpus was computed and filled-in in Table II which points to a good transparency. This conclusion was also strengthened by a panel of 5 human observers: while involved in a Two-alternative forced choice test, they agreed on the method transparency.

TABLE II. OBJECTIVE EVALUATION OF THE TRANSPARENCY.

	PSNR	AAD	PMSE	CQ	SC	NCC	DVQ
Average value (5 video sequences)	60.5	0.477	0.0003	101	0.98	0.99	0.05

### E. MPEG-4 AVC syntax-driven counter-attack

The aim of this section is to prove that by simply resetting the watermarked stream with the original encoding parameters the bit error rate will be significantly decreased. In other words, when reencoding the watermark sequence at a lower rate, the errors in mark detection are mainly caused to the syntax desynchronisation and not to the lost of information.

The principle is illustrated in Figure 6 and consists in re-encoding the attacked watermark video with the original encoding parameters; actually, in our experiments, three parameters have been considered: the rate, the QP and the prediction mode.

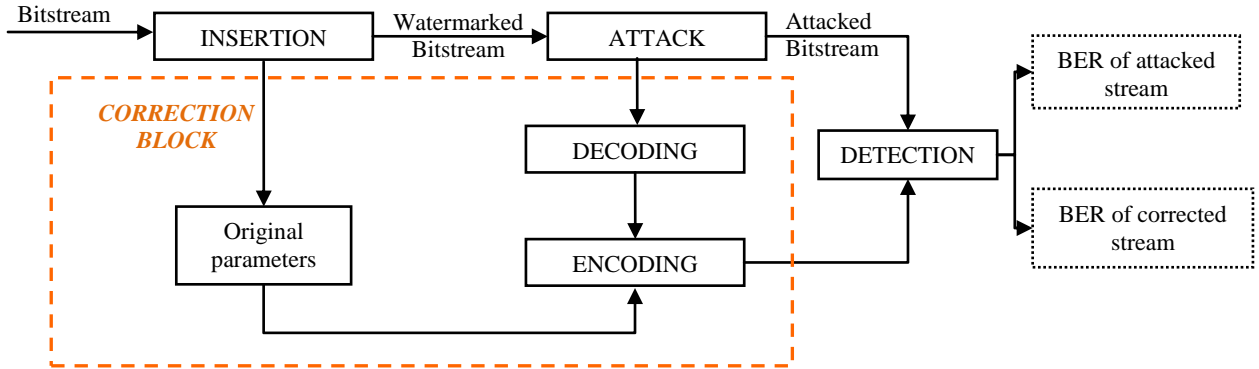









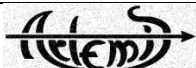




Figure 6. MPEG-4 AVC syntax-driven counter-attack.

This process was tested for transcoding and StirMark attacks. In the former case, the watermarked sequence was compressed from 512kbps down to 256, 128 and 64 kbps. In the latter case, the StirMark attack was individually applied to each frame in the decompressed sequence, then the attack sequence re-encoded with the original (unwatermarked) parameters. It can be considered that such counterattack is successful: an average gain of 3% was obtained for the bit error rate (BER), cf. Table III. The visual impact of this counterattack is also illustrated in Table III.

TABLE III. BER RESULT AGAINST TRANSCODING AND GEOMETRIC ATTACKS WITHOUT CORRECTIN, AFTER QP CORRECTION AND AFTER QP AND PREDICTION CORRECTION .

Video	Transcoding 64 kbps	Transcoding 128 kbps	Transcoding 256 kbps	StirMark
Attacked video	 6.75%	 5.2%	 5%	 11.25%
Corrected video QP	 4.6%	 3%	 3%	 10.25%
Corrected video QP and prediction mode	 2%	 2%	 1%	 8.65%

## VI. CONCLUSION

This paper resumes and extends the author's study devoted to MPEG-4 AVC stream watermarking by means of  $m$ -QIM methods. In this respect, the perceptual masking and the syntax-driven counter attacks are discussed. The overall experimental results point to the viability of such an approach: the transparency is proved by objective and subjective measures and the robustness is ensured against noise addition, transcoding and geometric attacks. The robustness is also, *a posteriori* validated by basic information theory tools.

Further work will be devoted to accurate noise matrix estimation (hence accurate capacity evaluation) and to the design of the corresponding optimal insertion rule.

## REFERENCES

- [1] I. Cox, M. Miller, and J. Bloom, Digital watermarking, Morgan Kaufmann Publishers, 2002.
- [2] M.L. Miller, I.J. Cox, J-P.M.G. Linnartz, and T. Kalker, "A Review of Watermarking Principles and Practices," Digital Signal Processing for Multimedia Systems, K. K. Parhi, T. Nishitani (eds.), Marcell Dekker, Inc. NY, 461-485(1999).
- [3] D. Zou, and J. A. Bloom, " H.264/AVC Stream Replacement Technique for video watermarking", IEEE International conference on Acoustics, Speech, and signal processing, ICASSP (2008).
- [4] A. Golikeri. P. Nasiopoulos, and Z. J. Wang, " Robust digital video watermarking scheme for H.264 advanced video coding Standard", Journal of Electronic Imaging 16 (4), 043008 (Oct-Dec 2007).
- [5] M. Noorkami, R.M. Mersereau "Compressed-Domain Video Watermarking for H.264", Proc. of the 2005 IEEE International Conference on Image Processing, ICIP, Genoa, Italy, 2, 890-893, (Sep 2005).
- [6] M. Belhaj, M.Mitrea, S. Duta, and F. Preteux, "MPEG-4 AVC robust video watermarking Based based on QIM and perceptual masking", International Conference on communication, Bucharest, (Jun 2010).
- [7] M. Hasnaoui, M. Belhaj, M.Mitrea, S. Duta, and F. Preteux, " MPEG-4 AVC stream watermarking by ST-mDM techniques", The 17th IEEE International Conference on Electronics, Circuits, and Systems, Athens, (Dec 2010).
- [8] J.J.Eggers R. Bäuml R. Tzschoppe, and B. Girod, "Scalar costa scheme for information embedding" IEEE TRANSACTIONS ON SIGNAL PROCESSING, 51(4), 1003-1019, (Apr 2003).
- [9] B. Chen, and G.W. Wornell, "Digital watermarking and information embedding using dither modulation," in Proc. IEEE Workshop Multimedia Signal Process, Redondo Beach, CA, 273-278, (Dec 1998).
- [10] I. Richardson, H.264 and MPEG-4 Video Compression, The Robert Gordon University, Aberdeen, UK, (2003).
- [11] A.B. Watson, "DCT Quantization Matrices Optimized for individual Images", Proc. SPIE, 1913, 202-216.
- [12] Watson AB, Barlow HB, Robson JG. "What does the eye see the best", Macmillan journalLt., 419 - 422 (Mar 1983).
- [13] H. A. Peterson, H. Peng, J. H. Morgan, W. B. Pennebaker, "Quantization of color image components in the DCT domain", in B. E. Rogowitz, M. H. Brill, J. P. Allebach, eds., Human Vision, Visual Processing, and digital Display II, Proc. SPIE, 1453, 210-222, (1991).
- [14] <https://ipbt.hhi.fraunhofer.de>