



**HAL**  
open science

## MPEG-4 AVC stream watermarking by ST-mDM 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 ST-mDM techniques. Electronics, Circuits, and Systems (ICECS), Dec 2010, Athens, Greece. pp.487 - 490, 10.1109/ICECS.2010.5724555 . hal-00848647

**HAL Id: hal-00848647**

**<https://hal.science/hal-00848647v1>**

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 ST-*m*DM techniques

Marwen Hasnaoui, Maher Belhaj, Mihai Mitrea  
Institut Télécom;Télécom SudParis, ARTEMIS Department  
France  
{marwen.hasnaoui, maher.belhaj\_abdallah,  
mihai.mitrea}@it-sudparis.eu

Françoise Prêteux  
MINES ParisTech  
France  
francoise.preteux@mines-paristech.fr

**Abstract** — The present paper introduces a new insertion/detection technique for MPEG-4 AVC stream protection by means of watermarking techniques. The embedding process relies on the ST-DM (Spread Transform – Dither Modulation) principles and adapts/extends the Watson’s perceptual model so as to cope with the compressed domain peculiarities. In contrast to the state-of-the art techniques considering only binary insertion techniques, our paper demonstrates the first *m*-ary DM technique; in such a way, for a prescribed transparency and robustness, the data payload is increased by a factor of  $\log_2 m$ . The experimental results correspond to the MEDIEVALS (French National Project) corpus of 4 video sequences of about 15 minutes each. The robustness against noise addition, transcoding and geometric (StirMark) attacks is proved. The transparency is assessed by both subjective and objective measures.

**Keywords** MPEG-4 AVC watermarking; *m*-ary DM; transcoding, geometric attacks

## I. INTRODUCTION

In order to protect ownership and to ensure piracy tracking-down for digital video, watermarking solutions [1-2] can be considered. A watermark represents some digital information embedded 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 [3-5] but rather individually addresses these issues.

D. Zou and J. A. Bloom [3] introduce a substitution watermarking method for MPEG-4 AVC stream. While ensuring a very fast insertion, the paper doesn’t evaluate either transparency or robustness against attacks.

A. Golikeri, P. Nasiopoulos and Z. J. Wang [4] propose an ST-DM watermarking method. Although improving the performances of the traditional ST-DM, this method features a quite small data payload and has no robustness against the geometric attacks.

A. Abrardo and M. Barni [5] present a new dirty paper coding watermarking scheme by replacing orthogonal with quasi-orthogonal codes. This scheme is robust against gain attacks. However the paper doesn’t evaluate transcoding and geometric attacks.

M. Belhaj, M. Mitrea, S. Duta and F. Preteux [6] introduce a binary ST-DM for MPEG-4 AVC stream watermarking. By combining ST-DM principles, a perceptual shaping mechanism, and an information-theory driven selection criterion, good transparency and robustness against transcoding and geometric attacks are obtained. In order to also increase the data payload, the present paper reconsiders the study in [6] and introduces an *ST-mDM* (multi-symbol ST-DM) technique. The main contribution consists in deriving the optimal decision rule under additive noise constraints and in applying it under the framework of the MEDIEVALS French national project. The following structure was considered. Section II presents the place the ST-DM holds under the general framework of the Quantisation Index Modulation (QIM) techniques. Section III is devoted to the *ST-mDM* algorithm. The experimental results are presented and discussed in Section IV. Section V concludes the paper.

## II. QIM PRINCIPLES

### A. Binary QIM

In quantization based methods, the host data  $x$  is quantized using multiple quantizers and the index of the quantizer is chosen based on the message to be embedded [7]. 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:

$$Q_{\Delta}(x) = \Delta \text{Round}(x/\Delta) \quad (2)$$

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:

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

The detection is computed as follows:

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

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 normed vector  $u$ . This technique is called Spread Transform as proposed by Chen and Wornell in [8]. The resulting scalar value is then quantized before being added in the component of the signal:

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

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

$$Y(d) = Q_{\Delta}(r^T u - k\Delta) - r^T u + k\Delta \quad (6)$$

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.

### III. ST- $m$ DM EXTENSION

The ST- $m$ DM 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  $D = \{-(m-1)/2, -(m-2)/2, \dots, 0, \dots, (m-2)/2, (m-1)/2\}$  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). In the present paper, we introduce a decision rule optimising 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  $[-\frac{\Delta}{2}; \frac{\Delta}{2}]$  interval; hence, specifying a decision rule means to divide the decision region  $[-\frac{\Delta}{2}; \frac{\Delta}{2}]$  into  $m$  nonoverlapping intervals.

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

$$Y(d) = Q_{\Delta}(B(d)) - B(d) \quad (8)$$

where  $B(d) = (\alpha - 1)q + \Delta \frac{d}{m}$ .

Given  $q$  a quantization error lying in the  $[-\frac{\Delta}{2}; \frac{\Delta}{2}]$  interval, then

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

Be there  $I_{sup}(d) = \frac{\Delta((1-\alpha)m+2d)}{2m}$  and  $I_{inf}(d) = \frac{\Delta((\alpha-1)m+2d)}{2m}$ . To avoid overlapping, there will be one decision interval for each element of the alphabet; hence,  $Q_{\Delta}(B, d, k)$  should take eventually one value. Be there  $Q_{\Delta}(B, d, k) = 0$ . Then, we have:

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

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

$$\begin{cases} -I_{sup}(d) \leq Y(d) \leq -I_{inf}(d) \\ \alpha > \frac{m-1}{m} \end{cases} \quad (11)$$

For a fixed parameter  $\alpha$ ,  $I_{sup}$  and  $I_{inf}$  are increasing functions of  $m$ . Hence, if each two successive symbols ( $d, d+1$ ) have nonoverlapping decision intervals, then we will have  $m$  nonoverlapping decision intervals. In this respect,  $I(d)$  and  $I(d+1)$  must verify the equation (12):

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

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  $Y(d)$  is between two decision regions, we decide for the region which is closest to  $Y(d)$ .

*Example:*

For  $m = 5$  we have  $\alpha^* = 0.8$ . The decision regions are obtained as illustrated in Figure. 1.

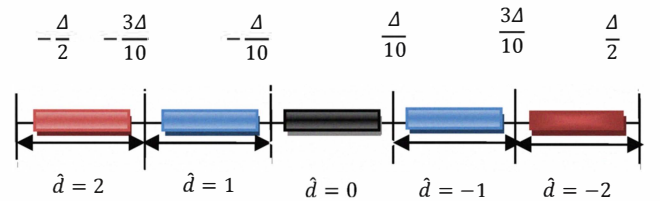


Figure 1. Decision regions for  $m=5$  and  $\alpha \geq \alpha^*$ .

### IV. ST- $m$ DM MPEG-4 AVC WATERMARKING

The method presented in this paper inserts the mark in the AC luma coefficients of the 4\*4 sub-macroblocks in the  $I$  frames [9] 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 (13) are considered:

$$\mu_x - \frac{\sigma_x}{16} < 1 - \|x\|/\|r\| < \mu_x + \frac{\sigma_x}{16} \quad (13)$$

where  $x$  and  $r$  are the original and the watermarked blocks,  $\| \cdot \|$  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.

### Pseudo-code for the selection

#### Begin video

If the current frame is  $I$  then

#### Begin frame

1- Select randomly  $n$  nonzero blocks that are disjoint

If the current block is selected then

- Compute the energy  $\|x\|$  of the bloc.
- Mark the block according to the equations (14) and (15)
- Compute the energy  $\|r\|$  of the block after marking
- Record the position,  $\|x\|$  and  $\|r\|$  for each selected block

Else go to the next block

#### End frame

Else go to the next frame

#### End video

1- Compute  $\sigma_x$  and  $\mu_x$

2- Compute the distortion  $\|x\| / \|r\|$  for each saved block

If the distortion verifies equation (13) then the bloc is selected and its position is recorded.

### A. Embedding process

The embedding process combines *ST-mDM* and perceptual shaping. It is structured into three main modules as shown in Figure 2: *perceptual shaping*, *mark generation* and *mark embedding*.

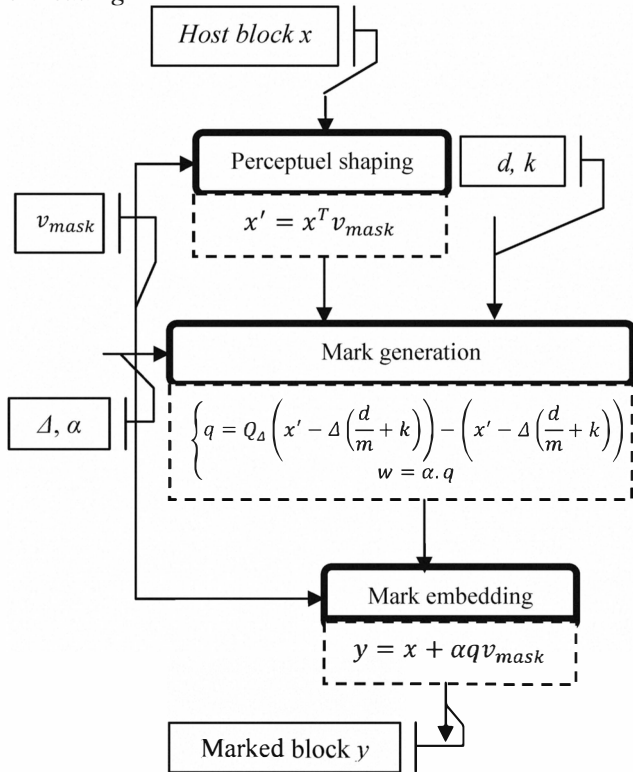


Figure 2. The embedding synopsis: three inputs (the message  $d$ , the host  $x$  and the key  $k$ ) and three parameters (the perceptual mask  $v_{mask}$ , the quantization step  $\Delta$  and the scaling factor  $\alpha$ ) are considered to compute the marked data  $y$ .  $m$  is the number of symbols in the message alphabet.

#### Perceptual shaping:

The watermark is not directly embedded into the original  $4 \times 4$  block  $x$  but into the projection  $x'$  of  $x$  onto a perceptual mask  $v_{mask}$ . The perceptual mask is obtained by adapting and

extending for MPEG-4 AVC the popular Watson model developed for still images [10, 6].

#### Mark generation:

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

$$\begin{cases} q = Q_{\Delta} \left( x^T v_{mask} - \Delta \left( \frac{d}{m} + k \right) \right) - \left( x^T v_{mask} - \Delta \left( \frac{d}{m} + k \right) \right) \\ w = \alpha q \end{cases} \quad (14)$$

#### Embedding mark:

This module generates the watermarked  $4 \times 4$  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:

$$y = x + w v_{mask} \quad (15)$$

### Pseudo-code for the embedding

#### Begin video

If the current frame is  $I$  then

#### Begin frame

If the current block is selected then

for  $i=1$  to 15 do

$x^i = x^i + \text{CurrBck\_coeff}[i]$  // compute the projection

Compute the mark  $w$  according to (14)

for  $i=1$  to 15 do

$\text{CurrBck\_coeff}[i] = \text{CurrBck\_coeff}[i] + w v_{mask}[i]$

Else go to the next block

#### End frame

Else go to the next frame

#### End video

### B. Detection process

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

### Pseudo-code for the detection

#### Begin video

If the current frame is  $I$  then

#### Begin frame

If the current block is selected then

for  $i=1$  to 15 do

$y^i = y^i + \text{CurrBck\_coeff}[i]$  // compute the projection

- Compute the decision value  $Y(d)$  of the block according to (3).

- Detect the mark using the detection rule

Else go to the next block

#### End frame

Else go to the next frame

#### End video

## V. EXPERIMENTAL RESULTS

The *ST-mDM* was implemented in the MPEG-4 AVC reference software version JM86.

### A. Videos corpus

The video corpus consists of 4 video sequences (further denoted by  $a$ ,  $b$ ,  $c$  and  $d$  respectively) 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 (13) 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.

### 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 domain: 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-mDM* features robustness against this attack: an average value of 6.75% was obtained for the bit error rate (BER), *cf.* Table I.

Finally, the robustness against geometric attacks was investigated. This time (*cf.* Table I), 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 3.

TABLE I. BIT ERROR RATE AFTER ATTACKS ( $m=5, \alpha=0.84, \lambda=150$ ).

	video <i>a</i>	video <i>b</i>	video <i>c</i>	video <i>d</i>
Transcoding	7%	6%	9%	5%
Geometric	9%	16%	7%	13%

From the information theory point of view, this robustness is connected to channels with high mutual information, as illustrated in Table II. The channel corresponding to the watermarking in this study has as input the  $m= 5$  alphabet  $\{-2, -1, 0, 1, 2\}$  and the output is represented by the decision according to the rule expressed in Section III. Table II 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 II. 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



Figure 3: The original ARTEMIS (left) binary logo recovered after transcoding (middle) and geometric attacks (right).

### D. Transparency

In order to prove the transparency of the watermarked videos, three types of metrics (Table III) 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 contour – SC and normalized cross correlation - NCC), and *psycho-visual measures* (digital video quality - DVQ). The values filled-in in Table III point 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 III. OBJECTIVE EVALUATION OF THE TRANSPARENCY.

video	PSNR	AAD	PMSE	CQ	SC	NCC	DVQ
<i>a</i>	60	0.161	0.0002	108	0.99	1.00	0.05
<i>b</i>	66	0.119	0.0002	100	0.99	0.99	0.05
<i>c</i>	52	1.531	0.0005	106	0.98	1.00	0.06
<i>d</i>	64	0.097	0.0006	93	0.99	1.00	0.07

## VI. CONCLUSION

The present study introduces the first *ST-mDM* watermarking method for MPEG-4 AVC stream. While applied to the MEDIEVALS corpus, the method proved good transparency (both objective and subjective evaluation criteria), strong robustness (against noise addition, transcoding and geometric attacks). The robustness is also, *a posteriorie*, evaluated by basic information theory tools. The data payload averaged 130 bits per minute, *i.e.* about 20 times larger than the minimal limit imposed by the DCI standard.

## 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, pages 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] A. Abrardo, and M. Barni, "Informed Watermarking by Means of Orthogonal and Quasi-Orthogonal Dirty Paper Coding,"IEEE TRANSACTIONS ON SIGNAL PROCESSING, VOL. 53, NO. 2, FEBRUARY 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, June 2010.
- [7] J.J.Eggers R. Bäuml R. Tzschoppe, and B. Girod, "Scalar costa scheme for information embedding" IEEE TRANSACTIONS ON SIGNAL PROCESSING, VOL. 51, NO. 4, APRIL 2003.
- [8] B. Chen, and G.W. Wornell, "Digital watermarking and information embedding using dither modulation," in Proc. IEEE Workshop Multimedia Signal Process, Redondo Beach, CA, pp. 273-278, Dec. 1998.
- [9] I. Richardson, H.264 and MPEG-4 Video Compression, The Robert Gordon University, Aberdeen, UK, 2003.
- [10] A.B. Watson, "DCT Quantization Matrices Optimized for individual Images", Proc. SPIE, Vol.1913, pp. 202-216.