Joint informed embedding and spread spectrum video watermarking

Sorin Duță[1], Mihai Mitrea[1,2], Françoise Prêteux[1]

[1] ARTEMIS Department, GET/INT
9, Rue Charles Fourier, 91011 Evry Cedex
[2] Faculty of Electronics and Telecommunications, POLITEHNICA University of Bucharest, Romania

{sorin.duta,mihai.mitrea,francoise.preteux}@int-evry.fr

Abstract

In the Internet era, any piece of digital/digitalised art (be it image, video, audio, 3D ...) can be anytime and anywhere replicated with a simple click, thus frustrating the artists/producers from a large part of their economic benefits. Imposing itself as a viable solution to the copyright enhancement problem, watermarking represents a research field which exploded in the last decade.
The present paper reports on an original video watermarking method for very low rate video based on: (1) the synergy between spread spectrum and informed embedding approaches (patent pending) and (2) an accurate statistical modelling of some real-life attacks. The detection is oblivious (it does not require the unmarked object). Firm results concerning transparency (no visible differences between the marked and the unmarked video) and robustness (with respect to both mundane transforms, as the compressions, and malicious attacks, as the StirMark attack) are obtained. The data payload is increased more than 10 times with respect to the state-of-the-art. Beyond traditional watermarking, our method can be applied for emerging enriched multimedia applications: interactive television, video on demand, scalable enriched content streaming, and adaptive indexing.

Key words

Low rate video watermarking, robustness, transparency, informed embedding, spread spectrum.

1 Introduction

When considering the Information Society in general and the Internet in particular, the art producers find themselves in a quite awkward position. On the one hand, a digital dimension is added as a completing element giving art a whole new perspective. On the other hand, this very dimension opens the door to author spoliation: any piece of digital/digitalised art (be it image, video, audio, 3D …) can be anytime and anywhere replicated with a simple click. For instance [1], in 2004, the DVD piracy summed up to a total amount of $ 512 billion. When expressing this phenomenon in terms of markup [1], it turns out to be more "profitable" than cocaine traffic, Fig. 1.
Watermarking is the potential solution to such a problem. It provides a mean to persistently associate copyright information with virtually any digital representation. Despite its very short history, we already dispose of a sound theoretical background [2-5] and of various derived technical solutions.
The present paper reports on an original watermarking method aimed to protect the property rights connected to video distribution in mobile networks. Section 2 presents the watermarking main definitions. Section 3 describes a new watermarking scheme (patent pending) also relying on an accurate statistical analysis of some real-life attacks (*i.e.* rotation, linear filtering, StirMark) described in Section 4. Section 5 experimentally supports our method while Section 6 concludes the paper and opens the perspectives for our future work.

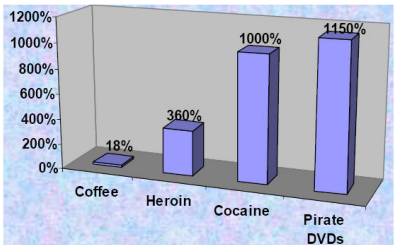


Figure 1 - *A comparison between four of the most profitable black market products*

2 Watermarking definitions

In its largest acceptation, *watermarking* stands for the practice of imperceptibly modifying an original piece of media in order to embed a message.
This embedded message is referred to as *mark* or *watermark*. Generally, it conveys copyright information and should be generated starting from some secret information referred to as *key*. According to the targeted application, the size (in bits) of the copyright information (*the data payload*) may vary.
When the embedded message does not alter the visual quality of the considered object, the watermarking procedure features *transparency*.
The *robustness* refers to the ability of the watermark to survive signal processing operations. Two classes of such operations should be considered. The first class contains the common transformations applied to the video sequence, *e.g.* compression, change of file format, *etc*. The second class is represented by *the attacks*. These are

malicious transforms designed to make the watermark detection unsuccessful while preserving a good visual quality for the considered video.
When the unmarked video is not required during the detection procedure, the method is *oblivious.*
The probability of false alarm expresses the probability of taking an unmarked object for a marked one. Its upper limit is application dependent.
By summarising these four requirements, it can be noticed that they are contradictory, *e.g.* the better the transparency the weaker the robustness. Hence, for each and every method, a trade-off among them should be reached.
The communication theory represents a generic framework for the watermarking applications, Fig. 2. The mark is generated starting from the message to be embedded and from the key. At the detection side, the message should be recovered. Hence, it represents a sample from the information source. Every factor that makes it difficult for the message to be recovered (the original video itself and the transformations applied to the marked video) is considered as noise.

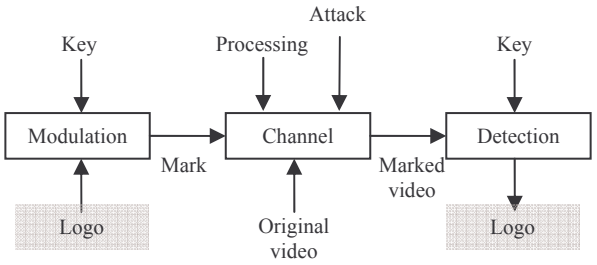


Figure 2 - *Watermarking application as a noisy channel*

From the owner point of view, the mark is corrupted by three noise sources: the original video, the distortions and the attacks. However, despite this theoretical model, a human observer (the person who buys the video) is interested in that video. From his/her point of view, the mark is responsible for the artefacts induced in the marked video. In order to keep these artefacts as un-disturbing as possible, the mark should have a very low power. Hence, a watermarking technique should allow the reliable detection of a very low power signal. Under the communication framework, the Spread Spectrum (SS) techniques have already offered good solutions to such problems. In its largest acceptation, an SS technique represents a communication technique in which the information is transmitted into a very large band. Consequently, an SS watermarking technique is a method in which the mark is spread over the original video, while occupying a much larger band than strictly necessary. In practice, the SS watermarking techniques feature good robustness but a quite small data payload.
The informed embedding (IE) is a different approach which exploits the knowledge about the original video in order to optimise the embedding procedure, *i.e.* it takes advantage on the fact that the main noise component is known at the embedder. From the theoretical point of view [6, 7], such a noise should not alter the channel capacity (*i.e.* the maximal amount of information which can be *theoretically* transmitted through the channel). Hence, the informed embedding approach is *a priori* very promising, at least from the theoretical point of view. In practice, they allow an impressive data payload but, generally, a quite poor robustness.
The watermarking procedure we designed (Section 3) synergistically combines the SS and IE principles, in order to reach the trade-off between data payload and robustness.

3 Method presentation

In order to pass from some theoretical concepts to a real life application, this paper adapts and extends the principles in [8] and also takes into account the results of the statistical investigation in Section 4.

☛ ***Mark modulation***

The M bits corresponding to the logo to be inserted are encoded according to the SS principles, by means of a modified trellis code, [8], [9].
The trellis has K states and 2 arcs exiting each state (each transition codes one bit). Each arc is labelled with an N length vector which components are real numbers.
These labels are computed starting from the key, *i.e.* they are known only by the true object owner.
The mark thus obtained is a vector denoted by g, with $M \times N$ real number components.

☛ ***Video representation***

The mark is inserted into a vector of salient characteristics of the video sequence, obtained as follows.
Be there a colour video sequence consisting in L frames. Each frame is represented in the HSV (hue-saturation-value) space [10]; the V component is normalised to [0,1] interval.
For each frame in the video sequence, the 2D-DWT is applied to the V component, at an N_r resolution level. The DWT coefficients corresponding to the LH and HL lowest frequency sub-bands, Fig. 3, are recorded into a vector. The coefficient hierarchy is built up by sorting in a decreasing order the vector previously obtained and the largest $M \times N$ rank values alongside with their original locations are recorded into two vectors, denoted by c_0 and λ, respectively.

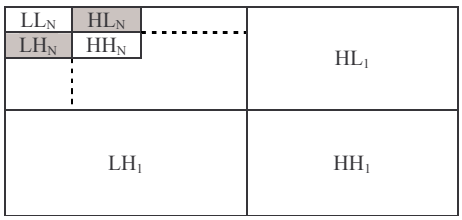


Figure 3 - *The chosen sub-bands*

☛ ***The detection***

We aim at establishing whether a suspicious video sequence was marked or not, *i.e.* finding out whether it contains the M bit logo or not.

We first extract from the video the vector susceptible to convey the mark. In this respect, the DWT is applied as above. Then, the coefficients corresponding to the λ locations are recorded, thus obtaining a $\hat{c}_W$ vector with $M \times N$ real components.

This vector is the input of a Viterbi decoder [9]. The decoder is pair designed with the trellis encoder. The cost involved in the Viterbi algorithm is the (un-normalised) correlation coefficient between the input sequence and the transition labels. This cost is to be maximised. High performances are obtained for uncorrelated labels.

☛ ***Informed embedding***

Designed by adapting the principles in [8], the embedding procedure aims at finding a c_W vector which is as close as possible to the c_0 vector and for which the Viterbi decoder produces the same output as for the g vector.

This c_W vector is iteratively computed, Fig. 4.

In the first iteration, c_W is initialised with c_0. Further on, a vector denoted by b is computed by applying the Viterbi decoder to $c_W + n$, and by trellis encoding the resulting bits. Here, n is a vector of $M \times N$ length, whose components are sampled from a noise source modelling the channel perturbations. Section 4 presents an accurate statistical investigation on the noise behaviour.

The c_W vector is now modified according to the following formula:

$$c_W \leftarrow c_W + \alpha \cdot (g - b) \big/ |g - b| .$$

The α scalar value is computed as follows:

$$\alpha = R_t - R(g, b, c_W) ,$$

where $R(g, b, c_W) = c_W \cdot (g - b) \big/ |g - b|$ and R_t is a scalar. The dot product between the c_W and the $(g - b)$ vectors is the un-normalised correlation coefficient.

The loop of b computation and c_W modification is repeated until the condition $R(g, b, c_W) \geq R_t$ is reached several times successively (*e.g.* 100 times – $N_j = 100$). If the equality between the g and the b vectors is reached before the $R(g, b, c_W) \geq R_t$ condition is achieved, then the b vector is computed without modifying c_W. If such a situation is encountered many times successively (*e.g.* 100 times – $N_i = 100$) then we consider that the g mark was successfully embedded into the c_W vector: regardless the added noise, the decoder can recover the message.

The c_W vector thus computed replaces the c_0 salient vector in the original video.

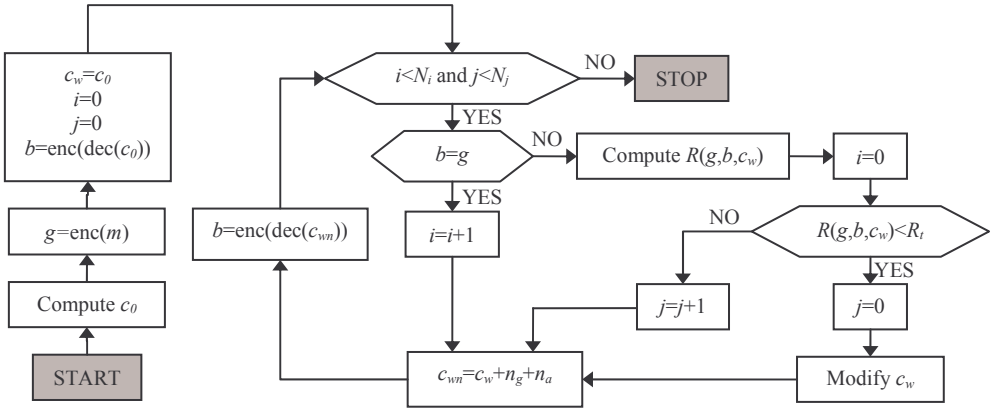


Figure 4 - *The informed embedding algorithm*

4 Noise statistical investigation

Concerning the effects of the video processing techniques and attacks, no reliable statistical investigation procedure has yet been advanced. However, in the literature, the popular AWGN (additive white Gaussian noise) model is assumed [2-5]. In order to find out whether this general assumption does hold in the particular case of our method (*i.e.* when embedding the mark into the DWT coefficient hierarchy) we reconsidered a statistical approach which already proved its efficacy in watermarking [11, 12].

The algorithm is structured in two parts. Steps 1 - 6 start from the attacked and original videos (2D random process - r.p.) and defines the noise as a 1D r.p. Steps 7 - 12 represent the statistical investigation on this r.p. Be there the same colour video sequence consisting in L frames represented in the HSV space, with the V component normalised to [0,1] interval.

Noise_modelling algorithm

For each frame in the video sequence:

Step 1: Apply the 2D-DWT to the V component, at an N_r resolution level.

Step 2: Record in a vector the DWT coefficients corresponding to the LH and HL lowest frequency sub-bands, see Fig. 3.

Step 3: Build up the coefficient hierarchy by sorting in a decreasing order the vector obtained in the previous step; record the largest N rank values alongside with their original locations; the resulting vectors are denoted by c_0 and λ, respectively.

Step 4: Apply the chosen attack to the considered frame.

Step 5: Resume the Steps 1 and 2 on the attacked video sequence; record the coefficients corresponding to the λ positions, thus obtaining the c_a vector.

Step 6: Compute the difference between the two vectors: $n = c_a - c_0$.

For each rank in the hierarchy and for all frames (i.e. for each component in all n vectors):

Step 7: Record in an x vector the values taken by a chosen r rank, $r \in [1, N]$, along all the L frames. In other words, the x vector is obtained by concatenating the r^{th} component from each of the L noise vectors.

Step 8: Periodically sample the x vector. When the D sampling period is large enough, by shifting the sampling origin, a partition into D classes with L/D independent elements is obtained.

Step 9: Verify whether all classes in the partition obey to the Gaussian law. In this respect, a Chi-square test [13] is run on each class. The ratio of the number of tests which are not passed to the D value can be considered as a measure of the overall Gaussian behaviour.

Step 10: Verify whether the D sampling period is large enough so as to afford the independence among the data in a partition class. Therefore, we run a Ro test on correlation [13] (for Gaussian data, the correlation and the independence are equivalent). As in the previous step, the relative number of tests which were not passed (*i.e.* the ratio of the tests which were not passed to the D value) is computed.

Step 11: A homogeneity investigation on the data in the same partition class is carried out by combining the Fisher F test and the Student T test [14].

Step 12: A homogeneity investigation among the D partition classes is performed by combing the same F and T tests.

5 Experimental results

5.1 The statistical investigation on attacks

The corresponding experiments have been carried out on 10 video sequences of about 25 minutes each (a number of $L = 35000$ frames in each sequence). These sequences were coded at a very low rates (64kbit/s, 192x160 pixel frames), as imposed by the mobile networks constraints.

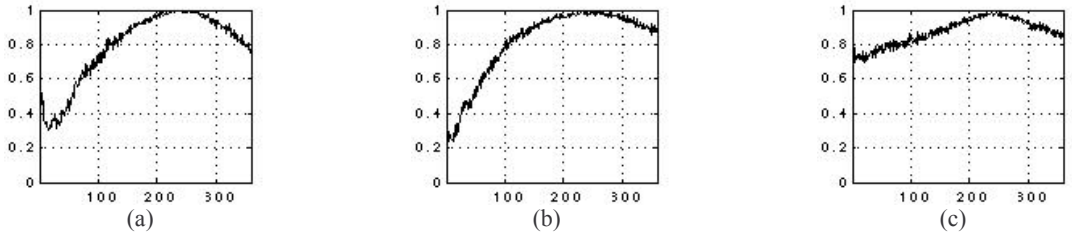
The (9,7) bi-orthogonal 2D-DWT [15] was applied at an $N_r = 3$ resolution level. The first $N = 360$ ranks of the hierarchy in each frame were considered. All the statistical tests were applied at an $\alpha = 0.05$ significance level.

The results obtained when applying the Step 9 are synoptically represented in Figs. 5, for three types of attacks: a rotation of 2 degrees (Fig. 5.a), a Gaussian linear filtering (Fig. 5.b), and the StirMark [16] attack (Fig. 5.c). The abscissa corresponds to the considered rank (from 1 to $N = 360$) while the ordinate depicts the ratio of the Chi-square tests which were not passed to the D sampling period (in these plots, $D = 550$ frames). Figs. 5 prove that for each type of attack and for each investigated rank, the Gaussian assumption is refuted: very large values are encountered for the rejecting ratio.

Figs. 6 are drawn for the same attacks and correspond to the Step 10. They *a posteriori* validate the $D = 550$ sampling period: each and every time, the ratio of refuted tests is about α. Note that the Ro tests are not properly run: they are meaningful only when the observation data are Gaussian distributed. However, as they are so nicely passed, we considered them as an additional (yet not theoretically sound) support for the sampling period.

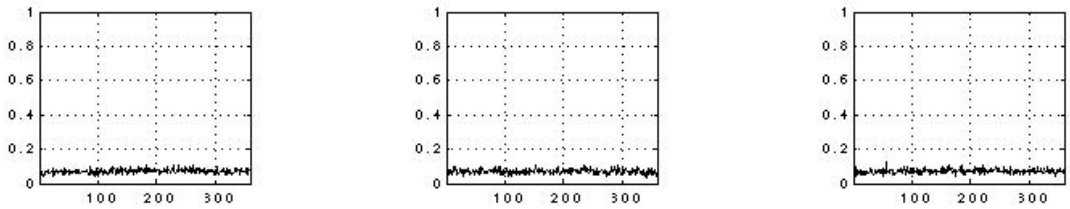
As the Chi-square tests were not passed, the Steps 11 and 12 become meaningless.

When inspecting Figs. 5 & 6, it becomes obvious that the Gaussian assumption does not hold for this application. Consequently, in order to obtain good watermarking performances, we considered in our method (Fig. 4) the very particular way in which some attacks act. When we tested the robustness against the attacks by computing the vector as a sum of the vector and of two noise sources: , corresponding to a Gaussian noise, and , corresponding to the StirMark attack. Note that the plots represented in Figs. 5&6 were obtained for a particular video sequence; however, the experiments resumed on the other 9 sequences led to the same general behaviour.



(a) (b) (c)

Figure 5 - *The Gaussian investigation on the noise r.p.: the ratio of Chi-square tests which are not passed to the D sampling period (ordinate) vs. the investigated rank (abscissa)*



(a) (b) (c)

Figure 6 - *The sampling period validation: the ratio of Ro tests which are not passed to the D sampling period vs. the investigated rank*

5.2 Robustness & transparency in practice

The watermarking experiments were run on 20 video sequences, each of them having 1000 frames (40 sec). These sequences are coded at 64 kbit/s and have 192×160 pixel frames.

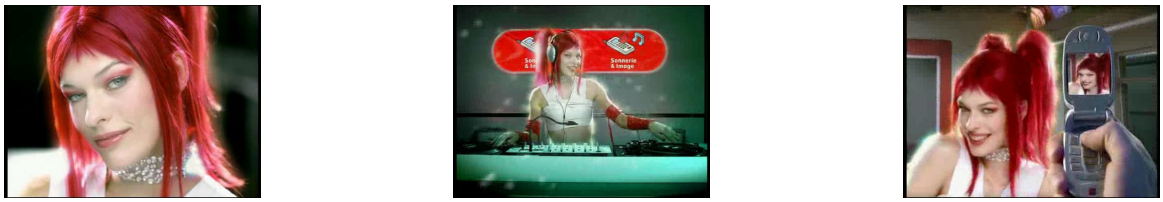
The 2D-DWT is applied at a $N_r = 3$ resolution level.

The original message to be inserted is represented on $M = 1000$ bits and corresponds to the binary SFR logo (just for illustration see Fig. 9.a). Each bit from this message is trellis encoded by a $N = 360$ real number label. These numbers are extracted from a random generator obeying a Gaussian distribution of $\mu = 0$ mean and $\sigma = 0.005$ standard deviation.

The R_t parameter involved in the embedding scheme (Fig. 4) was set to $R_t = 2$.

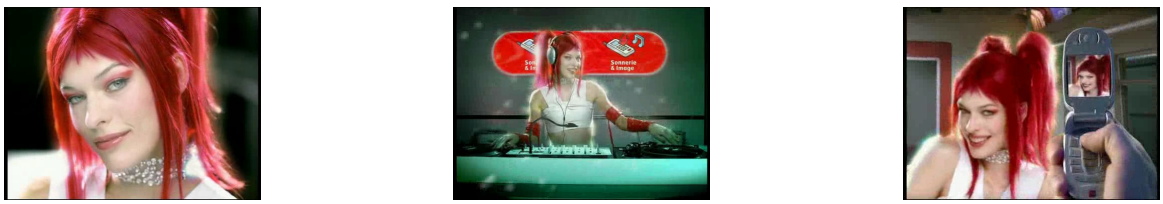
In order to subjectively evaluate the transparency, 25 human observers of different ages were involved in our experiments: 5 researchers deeply involved in the image/video processing, 5 researchers working in fields not connected with video processing, 5 persons with various educational backgrounds (foreign languages, history, law), 6 students, 1 film director, 1 film producer and 2 painters. They agreed that the method features fidelity. In order to also offer an objective measure of the transparency, the UIQI (Universal Image Quality Index [17]) was computed for each frame in the video sequence: their minimal, maximal and mean values are 0.9798, 0.9994, and 0.9981 respectively (a UIQI of 1 corresponds to identical images). Frames from original and marked *Advertising* sequences are represented in Figs. 7 and 8.

The method also features very good robustness. First, we check up the resistance against the mundane video processing: change of file format (from mpg to avi), linear and non-linear filtering (Gaussian, Laplace, median), small rotations (each frame was randomly rotated up to 2 degrees), noise addition, spatial and temporal cropping (up to 25% of frames have been randomly dropped). Each and every time, the visual logo has been successfully recovered. Secondly, the StirMark attack was individually applied to each frame in the sequence: although the commercial value of the video sequence was completely destroyed during this attack, the logo was still recovered. Fig. 9 illustrates the robustness. The logos recovered after the file format changing, Laplace filtering and the StirMark attack are represented in Fig. 9 a, b, and c, respectively.



(a) (b) (c)

Figure 7 - *Original frames sampled from the Advertising sequence*



(a) (b) (c)

Figure 8 - *Transparency for video watermarking: marked frames sampled from the marked Advertising sequence, and corresponding to the originals in Fig. 7*

(a) (b) (c)

Figure 9 - *Robustness for video watermarking: the SFR logo recovered after the file format changing (a), Laplace filtering (b) and the StirMark attack (c). Note that the logo (a) is practically identical to the original logo*

6 Conclusion

The paper presents a video watermarking method (patent pending [18]) which synergistically combines the spread spectrum and informed embedding principles in order to reach the trade off between data payload and robustness: we inserted a binary logo into a very low rate video sequence of 40s and we recovered it after the StirMark attack. This means increasing data payload by more than 10 times as compared to the state-of-the-art.

The method is based on the original informed embedding scheme represented in Fig. 4 and on the attack statistical analysis described in Section 4.

Fig. 4 is generic enough so as to be applied to other media types; very good results were already obtained for 3D objects and for audio (speech, music) signals.

Section 4 proves that the popular Gaussian law cannot model all the attacks, at least not when inserting the mark in the DWT hierarchy. Consequently, we considered in Fig. 4 both a Gaussian and a StirMark noise generator. Moreover, Fig. 4 allows the user to add some extra noise generator, modelling emerging/future designed attacks.

To conclude with, beyond traditional watermarking (copyright protection), this method has the potential to be extended for emerging applications, such as: interactive television, video on demand, scalable enriched content streaming, and adaptive indexing.

Acknowledgement

This study was partly supported by the French SFR mobile services provider (Vodafone Group).

References

[1] International Intellectual Property Alliance. 2006 Special 301 Report on Global Copyright Protection and Enforcement. http://www.iipa.com/special301_TOCs/2006_SPEC301_TOC.html

[2] I. Cox, M. Miller, J. Bloom. *Digital Watermarking*. Morgan Kaufmann Publishers, 2002.

[3] M. Arnold, M. Schmucker, S. Wolthusen. *Techniques and Applications of Digital Watermarking and Content Protection*. Artech House, 2003.

[4] F. Davoine, S. Pateux. *Tatouage de documents audiovisuels numériques*. Lavoisier, 2004.

[5] S. Katenbeisser, F. Petitcolas. *Information Hiding – Techniques for Steganography and Digital Watermarking*. Artech House, 2000.

[6] C.E. Shannon. Channels with Side Information at the Transmitter. *IBM Journal*: pp. 289-293, Oct. 1958.

[7] M. Costa. Writing on dirty paper. *IEEE Transactions on Information Theory*, Vol. IT-29, pp. 439-441, 1983.

[8] M. Miller, G. Doerr, I. Cox. Applying informed coding and embedding to design a robust high-capacity watermark. In *IEEE Trans. on Image Processing*, Vol. 13, No. 6, pp. 792-807, 2004.

[9] S. Lin, D. J. Costello Jr. *Error Control Coding: Fundamentals and Applications*. Prentice-Hall, 1983.

[10] The MPEG-7 International Standard, Text of ISO/IEC International Standard 15938-3 – Information Technology – Multimedia Description Interface, Part 3 Visual, Geneva, Switzerland, September 2001.

[11] M. Mitrea, F. Preteux, A. Vlad, C. Fetita. The 2D-DCT Coefficient Statistical Behaviour: A Comparative Analysis on Different Types of Image Sequences. *JOAM*, Vol.6, No.1, pp. 95-102, 2004.

[12] M. Mitrea, F. Prêteux, M. Petrescu. Very Low Bitrate Video: A Statistical Analysis in the DCT Domain. *LNCS*, Vol. 3893, pp. 99-106, 2006.

[13] R.E. Walpole and R.H. Myres. *Probability and Statistics for Engineers and Scientists*. MacMillan Publishing, 1989.

[14] B.R. Frieden. *Probability, Statistical Optics and Data Testing*. Springer-Verlag, N.Y., 1983.

[15] A. Chalderbank, I. Daubechies, W. Sweldens, B. Yeo. Wavelet transforms that map integers to integers. *Appl. Comput. Harmon. Anal.*, Vol.5, No.3, pp. 332-369, 1998.

[16] F. Petitcolas, R. Anderson, and M. Kuhn. Attacks on copyright marking systems. *LNCS*, Vol. 1525, 1998.

[17] Z. Wang, A. Bovik. A Universal Image Quality Index. *IEEE Signal Processing Letters*, Vol. 9, No. 3, pp. 81-84, 2002.

[18] M. Mitrea, F. Prêteux, J. Nunez. *Procédé de Tatouage d'une Séquence Video*. French Patent Request No. 05 54132, deposed on December 29th, 2005, in the name of SFR and GET/INT.