# Informed Watermarking for Low Rate Video

SORIN DUȚĂ[1,2], MIHAI MITREA[1,2], FRANÇOISE PRÊTEUX[1]
[1] ARTEMIS Project Unit,
GET/INT
9, Rue Charles Fourier, 91011, Evry-France
[2] Faculty of Electronics and Telecommunications,
POLITEHNICA University of Bucharest-Romania
http://www-artemis.int-evry.fr

*Abstract:* - Faced with the continuous increasing of processing and storage capabilities in the digital world, the intellectual right holders consider watermarking as an appropriate mean to protect their property. Emerging from the mobile network environments, the youngest member of the watermarking application family is the low rate video protection. This paper reconsiders an outstanding informed coding method developed for grey level still images and adapts and extends it for such colour video. The experiments were carried out in cooperation with the SFR wireless service provider in France and pointed to the significant improvements in some practical applications. These results also allow a comparison between the spread spectrum and informed watermarking techniques in the mobile networks.

*Key-words:* - informed embedding, robust video watermarking, mobile networks.

## 1 Introduction

In the early nineties, a *good* personal computer had an 8086 processor, a hard disk of 40MB, and a 3½ inch disk drive which could write 1.44MB on a floppy disk. Just 15 years later, a *good* personal computer has a Pentium IV processor, a hard disk of 120GB and disposes of a combo drive which can write 750MB on a CD or 4.5GB on a DVD.

This example has no advertising intensions! It is just to illustrate how explosively the computation power and the storage capabilities have increased. Each user becomes the master of a digital universe in which he/she can process (*e.g.* modify, compress, copy, transmit) any type of data: personal communication, audio, video, 3D, and text.

For the intellectual right holders, the main concerns are related to the copy and transmission facilities offered to a mundane user.

It should be emphasised that in the digital universe, the copies are perfect and *a priori* unlimited. Just an example: a VHS video cassette bought in the early nineties could not have been copied without any loss of quality. Hence, the value of the resulted copy was lower than the original. After a few copies (5 to 7), the quality dropped under any commercial threshold: nobody would have been interested in such a copy. On the contrary, any copy of a DVD is an exact replica of the original, thus preserving the original commercial value. Such a copy can be further spread over Internet. Under this framework, with no particular effort (either financial or technological), a film can be broadcasted on the Internet, thus frustrating its intellectual owner from the deserved commercial benefits.

A lot of research studies [1-4] pointed out to the watermarking technique capabilities to prevent such a scenario and a lot of sound theoretical and technical solutions have been advanced in this respect.

However, when approaching property right protection in mobile networks, very few studies can be found [5-8].

This paper addresses the challenging issue of video protection in mobile networks. The method presented in [9] for grey level still images is here reconsidered and extended for low rate video watermarking.

The paper structure is the following. Section 2 describes the watermarking framework for mobile networks. Section 3 is devoted to the watermarking method considered in this paper while Section 4 presents the experimental results. Section 5 concludes the paper and opens perspectives for the future works.

## 2 Problem statement

*Video watermarking* stands for the method of imperceptibly altering that video in order to embed a message, called *mark*. There are four requirements for a watermarking system:

- *Transparency* – the embedded message should not alter the video quality.
- *Robustness* – the message should not be affected either by mundane transformations applied to the video, or by any malicious attack intended to destroy the mark.
- *Obliviousness* – the detection algorithm should not require the original, unmarked video.
- *Low probability of false alarm* – a mark should not be detected in an unmarked video.

On the one hand, the role of each requirement depends on the considered application. For example, in medical data protection, the transparency is the strongest constraint. In contrast, for video distribution protection, the robustness is the key issue.

On the other hand, these requirements are somewhat contradictory. For instance, a better transparency generally comes across with a lower robustness.

Under the mobile framework, because of the small size of the terminal screens and of the restrictions imposed by the low bit-rate during transmission, the video quality is already low; hence, the constraints on transparency are somewhat alleviated. However, the same low bit-rate that creates the slack in transparency increases the restrictions on robustness.

From the communication theory point of view, a watermarking method can be modelled as a noisy channel: the message to be transmitted is the mark and the "noise" is represented by the original video itself, the transformations it may support and the malicious attacks, Fig. 1. Under this framework, the transparency constraint is expressed by a power constraint on the message to be transmitted. In order to ensure a reliable transmission two types of watermarking methods have been advanced: spread spectrum and informed watermarking.

The former [1], [10] is based on spread spectrum techniques inherited from military and mobile applications. Following this paradigm, in [11-12], the authors already advanced watermarking methods for good quality video: a serial number represented on 64 bits is inserted in each 40s of video. These methods feature good properties concerning the four above mentioned requirements. They have also been reconsidered and adapted for video watermarking in mobile networks [8]. However, for some applications, a larger amount of inserted bits may be necessary.

The latter type (informed watermarking), is based on the fact that the main noise component for the channel in Fig. 1 is the original video itself, which is known during the embedding procedure. Hence, the *informed watermarking* paradigm tries to take advantage on this side information, according to the principles presented in [13], [14]. A very interesting watermarking method belonging to this approach is presented in [9]. This method allows a very long message (about 1300 bits) to be inserted in the DCT (Discrete Cosine Transform) coefficients corresponding to an individual still image of about $250 \times 360$ pixels.

The present paper tries to benefit from such an outstanding method by adapting it for colour video watermarking in mobile networks.

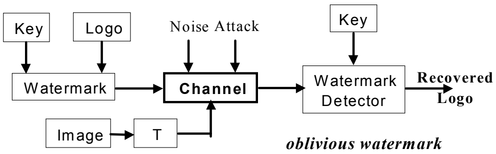


**Fig. 1 The watermarking method as a noisy channel.**

## 3 Method presentation

At a glance, the watermark embedding is seen in [9] as a three stage process:

1. *the message coding*: the original digital message (*i.e.* the bits to be embedded) is encoded into a signal by means of a modified trellis encoder [15], [9];
2. *the modification stage*: the $8 \times 8$ DCT is applied to the original image and a vector of

salient characteristics is extracted and processed so as to be matched to the trellis encoded signal;

3. *the post-processing stage*: the modified vector is considered instead of the original salient characteristic vector and the $8 \times 8$ IDCT (Inverse DCT) is computed.

The watermark detection is just a Viterbi decoding algorithm [15] applied to the $8 \times 8$ DCT coefficients in the marked image.

It should be emphasised that the vector to be embedded in the last step is computed - step 2 - starting from both the original message and the original image; hence this method does belong to the informed watermarking paradigm.

The three steps are further detailed. 1380 bits are to be embedded into a grey level still image of $240 \times 368$ pixels.

*Step 1*: The 1380 bits will be encoded with a trellis code which has 8 states and two arcs exiting each state. Consequently, each transition encodes 1 bit. Each arc has a label of 12 randomly generated numbers. The same set of labels is used at each trellis step. The result of this step is a vector denoted by $g$ which has $1380 \times 12$ components.

*Step 2:* The $8 \times 8$ DCT is applied to the original image and a vector denoted by $c_W$ is obtained by selecting some low frequency coefficients in each and every block, Fig. 2. Note that each block provides 12 coefficients, in concordance with the modified trellis label length. Further on, a vector denoted by $b$ is computed by first applying the Viterbi decoder to $c_W$ and then the trellis encoder to the obtained bits.

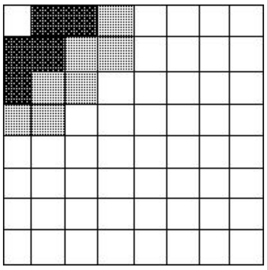


**Fig. 2 Low frequency coefficients to be marked in each DCT block.**

The $c_W$ vector is now modified according to (1):

$$c_W \leftarrow c_W + \alpha \cdot (g-b)/\left|g-b\right| . \qquad (1)$$

In (1), $|.|$ stands for the vector Euclidian norm (the square root of the sum of squared components), while the $\alpha$ scalar value is computed as follows:

$$\alpha = R_t - R(g,b,c_W) , \qquad (2)$$

where $R(g,b,c_W) = c_W \cdot (g-b)/\left|g-b\right|$ and $R_t$ is a scalar; no explicit suggestion concerning its value is found in [9]. The dot product between the $c_W$ and the $(g-b)$ vectors is the correlation coefficient.

The loop of $b$ computation and $c_W$ modification is repeated until $R(g,b,c_W) \geq R_t$ .

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, the loop is continued, but with a different way of computing the $b$ vector. The Viterbi decoder is now applied to a vector $(c_W + n)$ and then the trellis encoder is applied to the obtained bits ($n$ is a Gaussian noise sequence of 0 mean and $\sigma$ standard deviation). Each time the equality between the $b$ and $g$ vectors is obtained, the $\sigma$ value is incremented with a fixed quantity $\delta = 0.1$. This processes of incrementing the $\sigma$ value is stopped when $R(g,b,c_W) \geq R_t$. From now on, $\sigma$ is decremented with the same quantity $\delta$ and the whole Step 2 is repeated 100 times, on condition that $R(g,b,c_W) \geq R_t$. Should this condition be false, the process of $\sigma$ increasing is resumed.

*Step 3:* The values of the coefficients from the original $8 \times 8$ DCT are replaced with the values from the $c_W$ vector and the $8 \times 8$ IDCT is then computed.

In order to improve the method performances, in [9] two additional issues are considered: *informed coding* (*i.e.*, in *Step 1* the trellis is modified so as to obtain several code vectors for the same message) and *perceptual shaping* (the embedding procedure also takes into account the human visual system peculiarities [16]).

The final result consists in a watermarking method with the following properties: *transparency* (although the artefacts are noticeable in the marked image, they are not disturbing), *obliviousness*, and *robustness* (against low pass filtering, noise addition, changes in contrast, and JPEG compression). However, its main advantage is the very large amount of inserted data.

This method is now to be re-evaluated under the low rate video constraints, see Section 4. For a video watermarking, the transparency property is more restrictive than for the still image watermarking: some artefacts which may be unnoticeable in still images become obvious and disturbing in video sequences. Moreover, the class of attacks the method can face should be larger (for a real application, the StirMark should not be neglected [17]).

## 4 Experimental results

The experimental data consists of 20 video sequences, each of them having 1000 frames (40s). The frame sizes are $192 \times 160$ pixels, corresponding to a Motorola V550 cell phone.

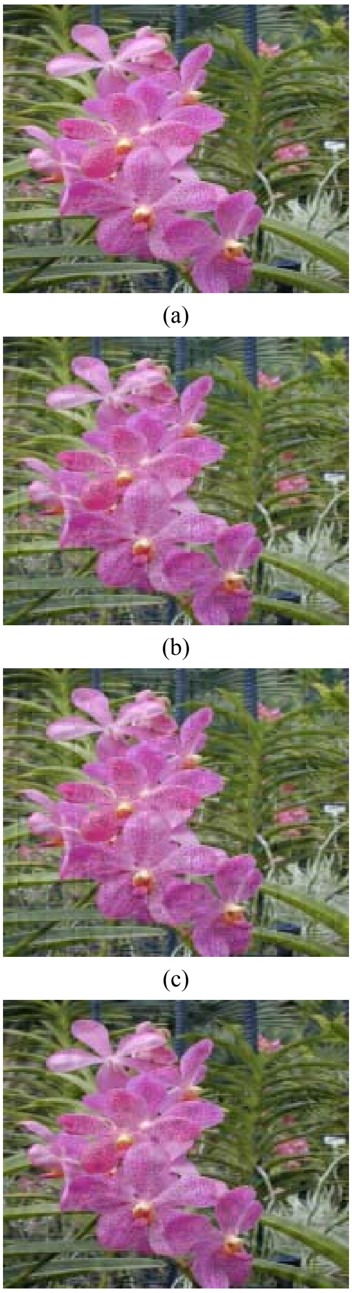
Concerning the video bit rate, four values have been considered, namely: 64kbit/s (the reference rate in telephony), 128 kbit/s, 192 kbit/s (the GPRS rate - **G**eneral **P**acket **R**adio **S**ervice) and 256 kbit/s. A frame from the original video coded at 192 kbit/s rate is represented in Fig. 3.a.

The video frames are represented in HSV (*H*ue – *S*aturation – *V*alue) colour space; the $V$ component is normalised to the $[0,1]$ interval.

The experiments were structured in several stages: first the method was applied in its basic form [9] and then successive adaptations have been considered.

*Stage 1*: The method is applied as presented in the three steps in Section 3 to each and every frame in the video sequence. The only difference comes from the frame sizes: the $240 \times 368$ numerical value is replaced by $192 \times 160$. Hence, each frame allows $24 \times 20 = 480$ DCT blocks, which represents, in fact, the number of bits to be inserted. The marked frames thus obtained have visible artefacts, see Fig. 3.b. In absence of any type of attack, about 3% of the bits are erred during the detection. The method features no robustness: even a mild low pass filter can destroy the mark.

*Stage 2*: In order to improve the transparency, the number of marked coefficients is decreased, *i.e.* the $c_W$ vector (Step 2, Section 3) is obtained by recording from each block only 5 coefficients (the dark grey shadowed area in Fig. 2) instead of 12. A marked frame is presented in Fig. 3.c: an acceptable image is now obtained, but the robustness is still out of question.



(a)

(b)

(c)

(d)

**Fig. 3 Frames sampled from the original (Fig. 3.a) and marked sequences**: Fig. 3.b corresponds to Stage 1, Fig. 3.c corresponds to Stage 2, and Fig. 3.d corresponds to Stage 4.

*Stage 3:* In order to grant some robustness, the data payload should be reduced, *i.e.* a redundancy should be induced among the 480 inserted bits. That means, in fact, an error correcting code. Just for illustration, in the experiments, a repetition code has been considered. Actually, in each frame, each bit was repeated 25 times and the same mark was embedded into 25 successive frames. In such a scenario, there are no more errors left after the detection procedure, even when the marked sequences are low pass filtered. However, even a mild JPEG compression (at a quality factor of 80) can destroy the mark.

*Stage 4:* The final stage in the experiments was to replace the DCT by a bi-orthogonal (9,7) DWT (Discrete Wavelet Transform) [18]. It should be mentioned that a previous study on spread spectrum watermarking [11], [12], [8] pointed out that DWT can afford better results than the DCT.

In fact, the DWT was applied to each frame at a 3 resolution level – see Fig. 4 - and some low frequency coefficients have been considered in order to obtain the $c_w$ vector defined in Section 3, Step 2.

The watermarked frame corresponding to the original in Fig. 3.a is depicted in Fig. 3.d.

The method thus obtained features robustness with respect to low pass filtering, noise addition, median filtering and JPEG compression (up to a $Q = 80$ quality factor).

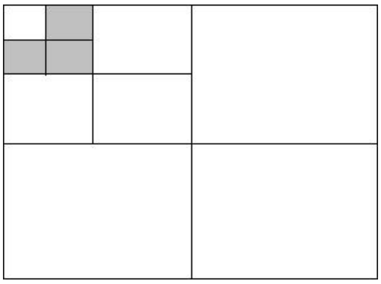


**Fig. 4 Low frequency coefficients to be marked in the image DWT decomposition.**

A synoptic view of the above mentioned results on robustness is presented in Fig. 5. The bits to be inserted represent a binary image – the logo of the ARTEMIS Project Unit, Fig. 5.a. This logo has a $40 \times 20$ pixel size and is subject to the following transformations:

- inserted according to the *Stage 2* and detected in the absence of any type of attack; the result is illustrated in Fig. 5.b;
- inserted according to *Stage 3* and recovered either without any attack – Fig. 5.c, or with a JPEG compression at a $Q = 80$ – Fig. 5.d ;
- inserted according to *Stage 4* (*i.e.* in the DWT domain) and recovered either without any attack – Fig. 5.e, or with a JPEG compression at a $Q = 80$ – Fig. 5.f ;

Note that such results were obtained for each of the 20 video sequences considered in the experiments.

Moreover, they were also invariant with respect to the considered video rates: 64kbit/s, 128 kbit/s, 192 kbit/s and 256 kbit/s and with the $192 \times 160$ frame size. When considering a better quality video, with larger frames, the results are expected to be improved.

Note that the logos in Figs. 5.d and 5.f have quite the same quality but for the one in Fig. 5.d an error correcting code was used.

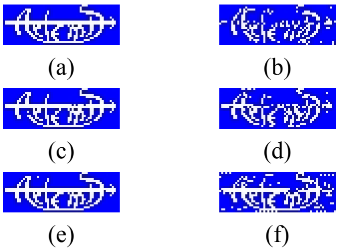


**Fig. 5 Synopsis of the robustness.** The original in Fig. 5.a is subject to various types of transformations: Fig. 5.b - Stage 2 embedding, no attack; Fig. 5.c - Stage 3 embedding, no attack; Fig. 5.d - Stage 3 embedding, mild JPEG compression; Fig. 5.e - Stage 4 embedding, no attack; Fig. 5.f - Stage 4 embedding, mild JPEG compression.

## 5 Conclusion and perspectives

This paper focuses on the opportunity of informed watermarking paradigm when protecting low rate colour video in mobile networks.

By reconsidering a method advanced for grey level still images and by gradually improving it, a large amount of data can be transparently embedded into a colour video sequence (*e.g.* 8000 bits for a 40s video). The drawback is represented by the relative lack of robustness: in fact, the mark can survive common transformations but just vanishes when malicious attacks are performed. In contrast to these results, the spread spectrum based method featured a very good transparency, a strong robustness but allowed just a 64 bit message to be embedded into 40s of video.

The results presented in this paper should be considered as a first step. Note that in the algorithms in Section 3 there are several numerical values ($R_t, \delta$, the trellis code itself) which seem to be application dependent (they were heuristically set). A first direction of future work is to find some supports enabling an automate assignation for these parameters.

Additionally, the informed coding and perceptual shaping are to be considered. However, this study identifies the need for a technique based on both spread spectrum and informed watermarking.

*Acknowledgement* This study is part of the TAMUSO (**TA**touage **M**ultimédia et ses **U**sages dans les ré**S**eaux m**O**biles) R&D contract between GET/INT and the SFR wireless service provider in France.

*References*

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

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

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

[4] G. Langelaar, I. Setyawan, and R. Langedijk, *"Watermarking Digital Image and Video Data. A State-of-the-art Overview"*, IEEE Signal Processing Magazine, Vol. **17**, No. **5**, 2000, pp. 20-46.

[5] F. Hartung, F. Ramme, *"Digital rights management and watermarking of multimedia content for m-commerce application"*, IEEE Comm. Magazine, Vol. **38**, Issue **11**, 2000, pp. 78-84.

[6] N. Checcacci, M. Barni, F. Bartolini, S. Basagni, *"Robust video watermarking for wireless multimedia communications"*, in Proc. of the IEEE Wireless Communications and Networking Conference, Chicago, Sept. 2000, pp. 1530-1535.

[7] J. G. Gomes, M. de Farias, S. Mitra, M. Carli, "*An Accurate Billing Mechanism for Multimedia Communications"*, in Proc. of IEEE ICME 2003, Vol. III, Baltimore, July 2003, pp. 93-96.

[8] M. Petrescu, M. Mitrea, F. Prêteux, *"Spread Spectrum Watermarking for Low Rate Video"*, submitted to WSEAS Intl. Conf. on Communications, Vouliagmeni-Greece, July 2005.

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

[10] I. Cox, J. Kilian, T. Leighton, T. Shamoon, "Secure Spread Spectrum Watermarking for Multimedia", *IEEE Trans. on Image Processing*, Vol. **6**, No. **12**, 1997, pp. 1673-1687.

[11] M. Mitrea, F. Prêteux, A Vlad, *"Spread spectrum colour video watermarking in the DCT domain"*, JOAM (Journal of Optoelectronics and Advanced Materials), Vol. **7**, No. **2**, 2005, pp. 1065-1072.

[12] M. Mitrea, T. Zaharia, F. Prêteux, A. Vlad, "*Video watermarking based on spread spectrum and wavelet decomposition*", Proc. SPIE Vol. 5607, 2004, pp. 156-164.

[13] C.E. Shannon, *"Channels with Side Information at the Transmitter"*, IBM Journal, October 1958, pp. 289-293.

[14] M. Costa, *"Writing on the dirty paper"* IEEE Trans. on Information Theory, Vol. **29**, 1983, pp. 439-441.

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

[16] A.B. Watson, *"DCT quantisation matrices optimised for individual images"*, Proc. SPIE Vol. 1913, pp. 202-216, 1993.

[17] F. Petitcolas, R. Anderson, and M. Kuhn, "*Attacks on copyright marking systems*" in Proc of the Second workshop on information hiding, David Aucsmith Ed., Lecture Notes in Computer Science Vol. **1525** Portland, USA, 1998.

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