Spread Spectrum Watermarking for Low Rate Video

MIHAI PETRESCU[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: - Music, video, and 3D characters are just some content examples that imposed themselves as a very important component of data distribution to mobile terminals. Hence, to reliably ascertain the related property rights is nowadays a crucial issue. This paper presents a study devoted to robust video watermarking in mobile networks: it reconsiders a method developed for regular networks and re-evaluates it under the mobile constraints. Experiments were carried out in cooperation with the SFR wireless service provider in France. The obtained results fulfil properties as: *robustness* (with respect to common attacks), *transparency*, *obliviousness*, and *low probability of false alarm*.

Key-words: - robust video watermarking, mobile networks, spread spectrum, DCT, DWT.

1 Introduction

The recent years have testified an amazing evolvement of the mobile networks: the cover area, the user number and the service variety exceeded any *a priori* expectation. The latest mobile-networks like GPRS - **G**eneral **P**acket **R**adio **S**ervice [1], or UMTS - **U**niversal **M**obile **T**elecommunications **S**ervices [2], can afford bit rate connections of about 192 kbit/s, thus turning the digital media distribution for mobile terminals into a common part of the daily life. Hence, to protect digital media property rights (music, video, 3D characters for *on-line* gaming) also becomes a crucial issue in mobile networks. The extensive research studies [3], [4], [5] on traditional networks (*e.g.* Internet) pointed out to the capability of robust watermarking techniques to solve such problems.

This paper presents a study dedicated to robust video watermarking in mobile networks. It reconsiders the method developed in [6] and re-evaluates it under the mobile network constraints.

The paper is structured as follows. The remaining of this section is devoted to the video watermarking main definitions (Section 1.1) and to the state-of-the-art in mobile video watermarking (Section 1.2). Section 2 describes the method under consideration. Section 3 details the experimental results while Section 4 concludes the paper and opens perspectives for future work.

1.1 Video watermarking key-words

In its largest acceptation, *video watermarking* stands for the practice of imperceptibly altering a video in order to embed a message.

This embedded message is referred to as *mark* or *watermark*. Generally, it is to convey copyright information (*e.g.* the video owner, the number of allowed copies, the time when that video was sold) 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 may vary. While for the first watermarking methods the payload was just 1 bit (a marked/unmarked decision), it is now required for it to have at least 60 bits [7].

When the embedded message does not alter the visual quality of the considered video, 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, temporal cropping, colour reduction, *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 video. In this respect, StirMark [8] can be considered as the most harmful attack.

When the unmarked video is not required during the detection procedure, the method is *oblivious*. As for most of the applications, the access to the original video is neither allowed nor recommended, oblivious watermarking techniques are preferred.

The probability of false alarm expresses the probability of taking an unmarked video for a marked one. The upper limit for this probability is application dependent. Just for illustration, it might be considered 10^{-10}.

Summarising these four watermarking 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.

For instance, when considering video distribution over Internet or DVD, the video quality is the key issue. The video is uncompressed (or, at least, very slightly compressed) and the transparency becomes a constraint which should be strictly observed to.

On the contrary, in video distribution for mobile phones, the original video itself has a very poor quality. This is a consequence of the transfer rate in mobile networks which is, in fact, the rate at which the video should be coded. The study in the present considers several video rates: 64kbit/s (the reference rate in telephony), 128 kbit/s, 192 kbit/s (the GPRS rate) and 256 kbit/s.

On the one hand, from the watermarking point of view, this low rate somehow means less room for the mark to be embedded and, therefore, harder robustness constraint. On the other hand, such a rate alleviates the transparency constraint: the watermarking artefacts can be somehow hidden by the artefacts already existing in the original video.

This is, in fact, the main objective of the present research study: to find out whether and how a method developed for the Internet/DVD [6] watermarking can be adapted to mobile networks.

1.2 Video watermarking in mobile networks

In contrast to the huge amount of bibliography on watermarking for generic applications (see the 3 sound books [3], [4], [5]), when considering just the mobile field, very few references can be found [9], [10], [11].

The paper in [9] is, in fact, a high-level review of the progresses in the standardisation processes. The authors present the principles of digital media rights management for mobile commerce (the electronic commerce for mobile terminals). The conclusion is that watermarking can and should be integrated as a component in such a copyright protection system. Without presenting any particular scheme, the non-triviality of the extension from Internet to mobile application is pointed out. It is also hinted that in mobile networks, there is some additional information which can accurately identify the user.

In [10], the authors reconsidered some known watermarking methods and test their effectiveness under two error types that frequently occur in mobile networks, namely burst and packet loss errors. The mark is represented on 15 bits and is recovered by an oblivious detection. However, no explicit discussion about the common attacks is made.

A watermarking related system which is devoted to advertising monitoring in TV broadcasting and 3G networks is introduced in [11]. For such an application, the false alarm probability plays the central role.

To conclude with, despite the huge potential impact of watermarking in mobile networks, at our best knowledge no reliable method to meet all the requirements was yet advanced.

2 Method presentation

The method under consideration in this paper was developed for video distribution on Internet and is detailed in [6]. Basically, it is a spread spectrum based technique [12], [13] in the DCT (Discrete Cosine Transform) domain. It embeds 64 bits into a video sequence of about 40s, depending on a key represented on 22 bits.

The method met all the requirements nowadays stated with respect to video watermarking on Internet: transparency (no visible difference between the original and marked frames), robustness (change of file format, compression, linear and nonlinear filtering, StirMark), obliviousness and low probability of false alarm (lower than $2 \times 10^{-16} \cong 16^{-14}$). The method extension in the DWT (Discrete Wavelet Transform) domain was presented in [13], [14]. This extension allowed a lower computation complexity and a lower probability of false alarm.

In the present paper, we considered both the DCT and the DWT representations for video. In the sequel, this method will be just summarised; for a complete description, see [6], [15].

Be there an original (unmarked) video sequence of L frames and be there 64 bits to be embedded in. The video frames are represented in the HSV (*H*ue – *S*aturation – *V*alue) colour space; the V component is normalised to the $[0,1]$ interval.

The watermarking procedure starts by applying the considered transform (either DCT or DWT) to each frame in the sequence and by recording R coefficients per frame alongside with their locations. The way these coefficients should be selected in order to ensure method optimality is demonstrated in [6] and [14]. Consequently, two $N = L \times R$ length vectors are obtained. The former is the recorded coefficient vector, denoted by v, and the latter is the location vector, denoted by l.

The mark is generated according to a CDMA (**C**ode **D**ivision **M**ultiple **A**ccess) technique [13], [6], [15]. Be $s_1, s_2, ..., s_{16}$ the 16 symbols in hexadecimal corresponding to the above mentioned 64 bits. 16 orthogonal bi-polar (+1/-1) sequences are randomly generated (by means of an LFSR – Linear Feedback Shift Register – characterised by a primitive polynomial of 21st degree; the corresponding 22 coefficients stand for the key). These sequences, denoted by n_i, have an $N+15$ length:

$$n_i = [n_{i,0}, n_{i,1}, ..., n_{i,N+14}],$$
$$i \in \{1,2,...,16\}. \quad (1)$$

An r_i sub-sequence, cut-out from the n_i sequence, is associated with each s_i:

$$s_i \leftrightarrow r_i = [n_{i,s_i}, n_{i,s_i+1}, ..., n_{i,s_i+N-1}],$$
$$i \in \{1,2,...,16\}. \quad (2)$$

The x mark is obtained by summing-up the r_i sequences:

$$x = \sum_{i=1}^{16} r_i \,. \quad (3)$$

The x mark is a pseudo-noise sequence, having the same length as the r_i sequences and zero mean (it is the sum of zero mean sequences); hence, a multiplication by a c value will multiply its variance by c^2 while keeping its 0 mean.

The embedding procedure adapts the principles in [11], [6], [14]. The v' vector of marked coefficients is obtained as follows:

$$v' = v \cdot (1 + \sigma \cdot x)\,. \quad (4)$$

The inverse transform (either IDCT – Inverse DCT - or IDWT –Inverse DWT) is computed, by considering at the l locations the v' values.

Finally, a post-processing transform is applied, with the following aims: (1) to minimise the artefacts induced by the embedding procedure and (2) to keep the marked V component in the $[0,1]$ interval, [6], [15].

In order to detect the mark, the transform is computed on possible corrupted/attacked frames and the coefficients corresponding to the l locations are recorded; be v'' the obtained vector. The embedded symbols are recovered by computing the cross-correlation function between the v'' coefficients and the n_i, $i \in \{1,2,...,16\}$, sequences. The peak position in such a cross-correlation function is the $\hat{s}_i$ recovered symbol:

$$\hat{s}_i = \underset{t \in \{0,1,...,15\}}{\arg\max}\; R_{v''n_i}(t)\,, \quad (5)$$

where $R_{uw}(\cdot)$ stands for the cross-correlation function between two u and w discrete sequences.

3 Experimental results

This section starts by presenting the parameter numerical values which ensure the trade-off between robustness and transparency. However, these numerical values can be adapted according any particular requirements.

The video sequences consist of $L = 1000$ frames corresponding to 40s (at a 25fps frame rate).

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) and 256 kbit/s. The frame sizes are 192×160 pixels, corresponding to a Motorola V550 cell phone.

Each frame provided $R = 64$ coefficients to be marked.

Concerning the σ parameter in Eq. (4), several numerical values have been considered, so as to lead to a mark power of $1/512$, $1/256$, $1/128$ and $1/64$.

When considering the DWT, a (9,7) bi-orthogonal transform [15], [16] has been applied at 3 and 4 resolution levels.

The *transparency* was evaluated by both subjective and objective means.

During the subjective evaluation, 10 human observers of different ages could not make any distinction between the marked and unmarked video sequences. These results were obtained in the DWT domain for a mark power of $1/128$ while in the DCT domain a mark power of $1/256$ was required. (The transparency is more restrictive in the DCT than in the DWT).

In order to objectively evaluate the transparency, the universal image quality index [17] was computed on each frame in the video sequence and then averaged; the corresponding numerical values are filled in Table 1. By its very definition [17], this index ranges between -1 and 1, the upper limit being reached if and only if the two images are identical. As all the numerical values in Table 1 are very close to 1, they also support method transparency.

Samples from the original video sequence are presented in Fig. 1, while the marked frames in the DWT domain at a 3 and 4 resolution level are under display in Figs. 2 and 3, respectively. Fig. 4 stands for the video sequence marked in the DCT domain. In Figs. 2, 3, 4, the mark power was set to 1/128. Note that the artefacts induced by the watermarking method in the DCT domain cannot be distinguished when inspecting individual frames but become obvious when the whole video sequence is watched.

In order to check up whether the method features robustness, several types of transforms were applied to marked video: change of file format (from avi to MPEG), temporal & spatial cropping, and small rotations. Each and every time the mark was successfully recovered, both in the DCT and DWT domains.

Finally, the robustness against the StirMark attack was taken into consideration. The attack was individually applied at its standard parameters [8] to each frame in the sequence. This time, several hexadecimal symbols have been erred out (of the 16 embedded hexadecimal symbols). Fig. 5 synoptically displays the average number of these errors: Figs. 5a and 5.b correspond to 3 and 4 resolution levels the DWT domain, respectively while Fig. 5c corresponds to the DCT domain. For each situation, the StirMark attack was applied 30 times and the error number was averaged. Out of inspecting Fig. 5, the following conclusion can be stated: the DWT domain can feature StirMark robustness only for a 4 resolution level and for a $1/128$ mark power while in the DCT domain, the StirMark robustness can not be reached.

All the experimental results were resumed on 20 different video sequences.

Table 1: The universal image quality index. The values correspond to the DWT applied at a 3 (line 1) and a 4 resolution levels (line 2), and to the DCT (line 3). Two mark power values have been considered: 1/256 (column 1) and 1/128 (column 2).

Mark power	1/256	1/128
DWT, 3	0.99993	0.99986
DWT, 4	0.99986	0.99971
DCT	0.99989	0.99989

4 Conclusion and perspectives

The present paper addressed the challenging issue of the robust video watermarking for mobile. By reconsidering and adapting a method devoted to video watermarking on Internet, the robustness, transparency, and obliviousness requirements were jointly met for the mobile terminals.

The method discussed in this paper is spread spectrum based. In the future work, the informed embedding paradigm will be taken into consideration in order to increase the size of the embedded message.

(a) (b) (c) (d)

Fig. 1. Frames sampled from the orginal video sequences, coded at different bit rates: 64 kbit/s in (a), 128 kbit/s in (b), 192 kbit/s in (c), and 256 kbit/s in (d).

(a) (b) (c) (d)

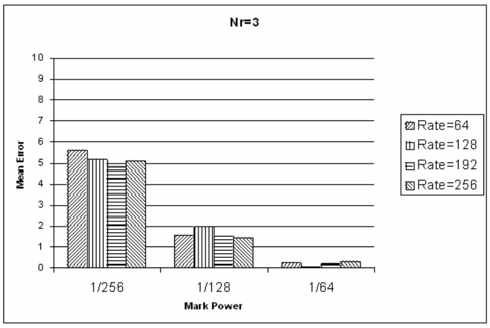
Fig. 2. Frames sampled from the marked video sequences, in the DWT domain, at a 3 resolution level. The video was coded at different bit rates: 64 kbit/s in (a), 128 kbit/s in (b), 192 kbit/s in (c), and 256 kbit/s in (d).

(a) (b) (c) (d)

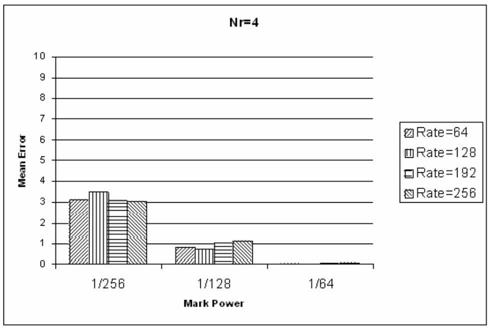
Fig. 3. Frames sampled from the marked video sequences, in the DWT domain, at a 4 resolution level. The video was coded at different bit rates: 64 kbit/s in (a), 128 kbit/s in (b), 192 kbit/s in (c), and 256 kbit/s in (d).

(a) (b) (c) (d)

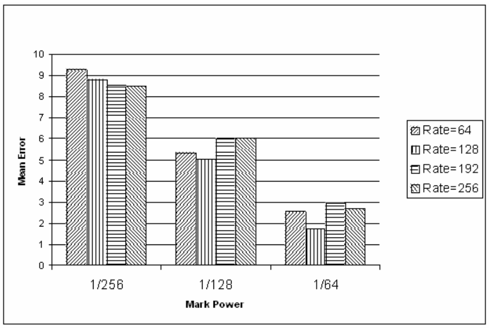
Fig. 4. Frames sampled from the marked video sequences, in the DCT domain. The video was coded at different bit rates: 64 kbit/s in (a), 128 kbit/s in (b), 192 kbit/s in (c), and 256 kbit/s in (d).



(a)



(b)



(c)

Fig. 5. The average number of errors after StirMark attack: the plots (a) and (b) correspond to the DWT applied to a 3 and 4 resolution levels, respectively while the plot (c) corresponds to the DCT. In each plot, the mark power is represented on the abscissa, while the ordinate stands for the number of erred symbols. For the same mark power, four bars are plotted, one for each video rate (from left to right): 64kb/s, 128kb/s, 192kb/s, and 256kb/s.

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] www.gsmworld.com/technology/gprs/intro.shtml.
[2] www.umtsworld.com/technology/overview.htm.
[3] I. Cox, M. Miller, J. Bloom, *Digital Watermarking*, Morgan Kaufmann Publishers, 2002.
[4] S. Katenbeisser, F. Petitcolas, *Informartion Hiding – Techniques for Steganography and Digital Watermarking*, Artech House, 2000.
[5] M. Arnold, M. Schmucker, S. Wolthusen, *Techniques and Applications of Digital Watermarking and Content Protection*, Artech House, 2003.
[6] 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.
[7] 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.
[8] 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.
[9] 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.
[10] 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.
[11] 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.
[12] 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.
[13] J. Ó Ruanaidh, T. Pun, *"Rotation, scale and translation invariant digital image watermarking"*, *Signal Processing*, Vol. **66**, No. **3**, 1998, pp. 303-317.
[14] M. Mitrea, F. Prêteux, A. Vlad *"Watermarking Oriented Video Modelling in the Wavelet Domain"*, WSEAS Transactions on Mathematics, Vol. **3**, Issue **1**, 2004, pp. 282-287.
[15] 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.
[16] 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.
[17] Z. Wang, A. Bovik, *"A Universal Image Quality Index"*, IEEE Signal Processing Letters, Vol. **9**, No. **3**, 2002, pp. 81-84.