Journal of Optoelectronics and Advanced Materials Vol. 7, No. 2, April 2005, p. 1065 - 1071

SPREAD SPECTRUM COLOUR VIDEO WATERMARKING IN THE DCT DOMAIN

M. Mitrea[a,b*], F. Prêteux[b], A Vlad[a,c]

[a]Faculty of Electronics and Telecommunications, Politehnica University, Bucharest–Romania
[b]ARTEMIS Project Unit, GET / INT, Evry–France
[c]The Research Institute for Artificial Intelligence, Romanian Academy

This paper presents a new robust watermarking algorithm devoted to colour video protection. This algorithm is developed out of generalising and adapting the principles we derived for still image sequences. The mark consists of a 64 bit message modulated by means of a spread spectrum technique and is embedded into the 2D-DCT hierarchy. The right room for the mark to be embedded into this hierarchy was established according to our previous study on statistical video modelling. The method thus obtained features *robustness* (against the general attacks, Stirmark included), *transparency* (*fidelity* and *quality*) and *low probability of false alarm* (lower than $2\times10^{-16} \cong 16^{-14}$). The experimental data consist of 100 colour video sequences of about 40 s each.

(Received December 7, 2004; accepted March 23, 2005)

Keywords: Colour video watermarking, Spread spectrum, 2D-DCT hierarchy

1. Introduction

Emerging from a two thousand year tradition of military and diplomatic applications (well mentioned in the literature, from Herodotus' Histories up to Shannon's works) the watermarking has been recently identified as the framework under which a large variety of problems directly connected to image/video/music distribution across Internet can be solved [1], [2].

Let us take the example of a digital video product which is **A**lice's property. When **B**ob buys this product, he can copy it and he might try to sell it again to somebody else, *e.g.* to **C**arol. Of course, this situation can be repeated as many times as there is an interest in the considered video: note that in the digital world the copies are exact (without any loss of quality as they were, for instance, in the case of the VHS cassettes). In such a scenario, although Alice may be aware of the unauthorised distribution of her video, she cannot do anything in order to protect her property rights. First, Carol can just pretend that the video is her property. Secondly, Bob can decline any involvement in such a distribution.

Under the watermarking framework, in order to protect her property rights, Alice is to embed an extra information (referred to as *mark* or *watermark*) in her video sequence. This mark is generally obtained starting from an information which may by public (a name, a logo, a serial number, *etc.*) and from a secret information (a *key*). In this scenario, when Alice find in Carol's video the information she embedded into Bob's video, she can both proof her property and identify Bob as the copy maker.

When the embedded information cannot be perceived by the human observer (that is, when a human observer cannot make any distinction between the marked and the original video), the watermarking procedure features *fidelity* [1]. When there are some artefacts in the watermarked product which do not disturb a human observer, the watermarked product features *quality* [1]. Note that for the *quality* definition, it does not matter whether the artefacts were induced by the watermarking procedure or whether they had already been present in the original video. The *fidelity*

* Corresponding author: mihai.mitrea@int-evry.fr

and the *quality* are the two components of the *transparency* property. Hence, by a *transparent* watermarking technique, Alice can obtain a marked video product which has the same commercial value as the original one, the copies of which can be tracked-down.

On the other hand, Carol may know that Alice protected her video sequence. She may also know the public information and the way in which the mark was embedded, but she does not know the *key*. Hence she may try to process the marked video she bought in order to turn the Alice mark undetectable. Such a transform applied by Carol is referred to as *attack*. For instance, Carol may try to change the video file format (*e.g.* from MPEG to DivX), to resample, to reduce the colour depth, to change the frame sizes, *etc*. Of course, Bob is also interested in attacking Alice's mark, in order to be able to sell the copies he can make. From this point of view, Stirmark [3] is a very dangerous attack: it was found out that very few watermarking methods reported in the literature can face it. A watermarking procedure which can face various attacks is referred to as *robust*.

Finally, a third requirement is stated: the probability of detecting a mark into an unmarked product (*i.e.* the *probability of false alarm*) should be lower than a certain value, *e.g.* $P_f \leq 10^{-10}$.

From the communication theory point of view, a watermarking procedure may be modelled as a noisy channel [1,4]. The Alice mark stands for the message to be transmitted. The video sequence and the attacks play the role of the noise (they alter the information during the transmission). For the human observer, the embedding procedure *a priori* leads to some artefacts in the marked video. Hence, the transparency requirement means a very low power for the mark. It is known [5], [6] that the spread spectrum - SS - modulation techniques provide the best framework for low power signal transmission.

When trying to take advantage of such an SS technique in a watermarking application, a limitation is found: generally, we do not dispose of a *large* spectrum suitable to convey the mark. The method we designed surmounts this difficulty by building up a *large* spectrum composed of some large 2D-DCT coefficients (the DCs - the direct components - are not taken into consideration) of each and every frame in the video sequence. As these coefficients stand for salient characteristics of the video, they might be preserved by all the transformations which keep the same quality of the video (included the attacks). On the other hand, as the SS techniques allow for a very low power signal to be detected (that is, a very low power mark to be recovered), the mark may be embedded into the largest 2D-DCT coefficients without altering the video quality.

For our method, the public message Alice wants to embed is represented on 64 bits, according to the general requirement concerning a message length [2] - between 60 and 70 bits. The key is represented by the coefficients of a primitive polynomial of 21 degree [7]. The SS embedding procedure we designed combines and generalises the principles in [6,8,9]. The mark thus obtained is embedded into the 2D-DCT hierarchy. Note that in contrast to the method devoted to still image sequences [9], this time we were not allowed to embed the mark into the highest 2D-DCT coefficients. In order to find out the rank interval in the 2D-DCT coefficient hierarchy which affords the trade off between *transparency* and *robustness* we considered the results we obtained on statistical video modelling [10].

The method we advance was applied to 100 colour video sequences of about 40 s each. These sequences were inhomogeneous, containing both indoor and outdoor scenes, low and high motion activities, *etc*. We consider several types of natural video sequences: art, documentary, news, amateurs.

The experimental results we obtained pointed to *robustness* (with respect to resampling, change of format, reducing colour depth, linear filtering, noise addition, small rotations and Stirmark attack), *transparency* and *low probability of false alarm* (lower than $2\times10^{-16} \cong 16^{-14}$).

2. The method presentation

We shall further present our algorithm. For clarity sake, the parameters will be set to the numerical values which were involved in our experiments. These numerical values can be adapted so as to meet the requirements of any particular application (*e.g.* for video surveying applications there are lower constraints regarding the *transparency* while for cartoons such constraints are increased). On the other hand, note that these values do not depend on the video sequence to be marked: we validated them on 100 video sequences of different types.

Be there a colour video sequence of $L = 1024$ frames (about 40 s). Each frame is represented in the HSV (hue-saturation-value) system, as implemented by the MPEG–7 colour space

descriptor [11]. The 2D-DCT is applied to each frame and the coefficients thus obtained are sorted in a decreasing order; note that we did not take into consideration the direct components. Be I a rank interval from this hierarchy which has a 64 rank length. In our experiments we shall individually consider three such intervals, *i.e.* either $I_1=[1;64]$, either $I_2=[50;113]$, or $I_3=[129;192]$. Be v the vector of coefficients (corresponding to one of the three intervals) and be l the vector of the corresponding locations; hence, these two vectors are both of $N=64\times1024$ length. (Note: this $N=64\times1024$ numerical value is large enough so as to allow the spread spectrum techniques to properly work).

The public information, *i.e.* the information Alice wants to insert into her video, is represented on 64 bit, *i.e.* 16 digits in hexadecimal. Be these hexadecimal digits denoted by $s_1,s_2,...,s_{16}$. This message is modulated by means of an SS - CDMA (Spread Spectrum - Code Division Multiple Access) technique, starting from the suggestions in [6,9].

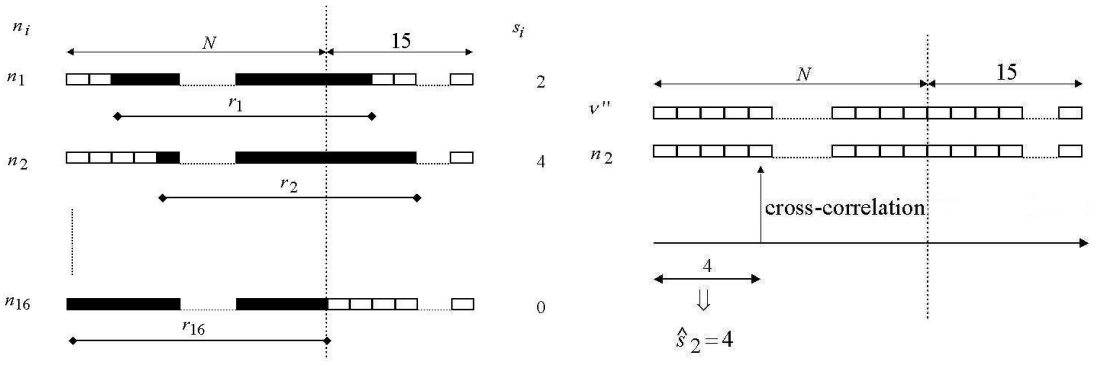
The procedure starts by considering 16 bipolar $(-1/+1)$ pseudo-noise sequences (one sequence for each hexadecimal digit in the message) of an $N+15$ length. These sequences, denoted by $n_i, i=\overline{1,16}$, should be orthogonal so as to afford an optimal detection: the performances of any SS technique depend upon the quality of the pseudo-random number generator it involves in [5,12,6]. In our experiments, we considered a generator implemented by means of an LFSR - Linear Feedback Shift Register [7, 13]; of course, other generators may be considered, as those based upon the Gold codes [12] for instance. Note that the output of an LFSR generator takes the 0 and 1 values. In order to obtain the $-1/+1$ bipolar values, the 0 value is just replaced by -1.

The LFSR we consider is characterised by a primitive polynomial of 21 degree. Hence, its output has a 2^{21} - 1 period [7,13]. The 16 n_i sequences of $N+15$ length were cut out from such a sequence. As the N value is large enough and as the polynomial is primitive, the 16 sequences thus obtained (denoted by n_i) are orthogonal. On the other hand, note that the 22 coefficients of the polynomial stand for the key in our watermarking method (the secret information Bob and Carol should not know).

Further on, each s_i symbol, $i=\overline{1,16}$, is put into correspondence with an r_i sequence. These 16 sequences are denoted by r_i and have an $N=1024\times64$ length and are extracted from the corresponding n_i sequence, $i=\overline{1,16}$, according to (1):

$$s_i \leftrightarrow r_i=[n_{i,s_i}, n_{i,s_i+1},...,n_{i,s_i+N-1}] \tag{1}$$

Fig. 1.a exemplifies how the r_i sequences are obtained. For instance, be there $s_1=2, s_2=4,..., s_{16}=0$. Eq. (1) means that the r_1 sequence is built up from the symbols which have the indices $2,3,...,N+1$ in the n_1 sequence. Following the same equation, the r_2 sequence is build up from the symbols which have the indices $4,5,...,N+3$ in the n_2, and so on. (The first component into a vector has the index equal to 0.)



(a) (b)

Fig. 1. The modulation (a) and the demodulation (b) technique involved in our watermarking method.

Note that as the s_i symbol can take a value between 0 and 15, the n_i sequences should have the above mentioned $N+15$ length in order to allow the r_i sequences to be obtained.

The mark to be embedded is a vector denoted by x which is the sum of the 16 r_i vectors, $i=\overline{1,16}$, *cf.* Eq. (2); hence, the x mark has the same $N=64\times1024$ length as any r_i vector:

$$x_j=\sum_{i=1}^{16} r_{i,j}, \forall j=\overline{0,N-1} \tag{2}$$

where x_j represents the j index component of the x vector while $r_{i,j}$ stands for the j index component of the r_i vector.

This mark is embedded into the selected v coefficients by means of a weighted addition [8,9]. The embedding procedure is described by Eq. (3):

$$v'_j=v_j\cdot(1+\sigma x_j) \tag{3}$$

where the lower index j denotes the component of the respective vector ($j=\overline{0,N-1}$), v' denotes the marked coefficients while σ stands for a constant. Note that this multiplication by a σ constant adjusts the power of the x mark. The larger the σ value, the greater the robustness but the worst the transparency. Out of our experiments, we found out that a $1/512$ mark power reaches the trade off between the robustness and the transparency desiderata.

Further, the v original coefficients are exchanged for the v' marked coefficients and $L=1024$ 2D-IDCTs (2D Inverse DCT) are computed, thus obtaining the marked V components.

Finally, a post-processing transform is applied in order to reduce the visibility of the artefacts induced by the watermarking procedure. We empirically designed this non-linear transform, which ensures that the V (luminance) component has the same mean value before and after marking procedure, see relations (4) and (5). In order to afford such a property, the minimum value of the marked V component is first substracted from each value in this V matrix; the obtained V component is denoted by $\tilde{V}$:

$$\begin{aligned} &m=\min(marked\_V\_component);\\ &\tilde{V}\_component=marked\_V\_component-m. \end{aligned} \tag{4}$$

At this stage, some values in $\tilde{V}$ may be larger than the maximal value in the unmarked V component; we limit such values to this maximal value. Finally, this limited $\tilde{V}$ component is divided by the ratio of its mean value to the mean value of the unmarked V component:

$$\begin{aligned} &\boldsymbol{r}=\mathbf{mean}(limited\_\tilde{V}\_component)/\mathbf{mean}(unmarked\_V\_component)\\ &final\_marked\_V\_component=limited\_\tilde{V}\_component/\boldsymbol{r} \end{aligned} \tag{5}$$

The watermarking procedure is thus achieved.

Let us now suppose that Alice finds a video sequence she thinks is her in Carol's possession. In order to prove her property rights on this copy, Alice computes the 2D-DCTs on each frame in the sequence. Note that Carol's sequence generally differs from the marked sequence Alice sold to Bob. This difference has several reasons, *e.g.* Bob might try to attack the watermark or Carol just wanted a compressed version of the video. In any situation, Alice records the coefficients which correspond to the l recorded locations (see the first step of the embedding procedure); the vector obtained by concatenating these coefficients is denoted by v''.

The public message is recovered [6,9] by means of $R_{v''n_i}(.)$ cross–correlation functions, computed between v'' and each n_i , $i=\overline{1,16}$ sub–sequence. Note that there is no need for these n_i

sequences to be recorded: Alice simply computes them again, starting form the polynomial coefficients.

The peak position in such a cross–correlation functions represents the $\hat{s}_i$ recovered symbol. That is, the $\hat{s}_i$ recovered symbol is implicitly involved in (6), see Fig. 1.b:

$$R_{V''n_i}(\hat{s}_i) = \max_t R_{V''n_i}(t), \; t = \{0,1,...,15\} \tag{6}$$

Eq. (6) means, in fact, to determine the $\hat{s}_i$ value which maximises the cross-correlation function.

Although these cross-correlation functions are not delta-like because of the noise and of the attacks, the peak position might keep its original position, as a consequence of the fact that the $n_i, i = \overline{1,16}$ sequences were orthogonal. However, when the peak position change its position, an error is encountered.

3. Experimental results

The watermarking method we designed was applied to 100 video sequences of about 40 s each. These sequences were inhomogeneous, containing both indoor and outdoor scenes, low and high motion activities, etc. We considered several types of natural video sequences: art, documentary, news, amateurs. We also considered several values for the frame sizes, from 200×250 to 450×750.

We shall first present the results obtained when the mark is embedded into the $I_3 = [129;192]$ interval in the 2D-DCT coefficient hierarchy and when considering a σ value in (3) which corresponds to a $1/512$ mark power. Then we shall include three tables regarding both the $I_1 = [1;64]$ and $I_2 = [50;113]$ intervals and other mark power values.

In order to check up the transparency, 25 human observers were involved in our experiments. They were of different ages (from 20 to 75 years old) and professions: 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. All the video sequences we considered were coded into the DivX format, good quality. We found out that the method features fidelity. This transparency property was exhibited by all the 100 natural video sequences. On the other hand, when considering cartoons, where there are large areas with the same colour, some artefacts became obvious (some artificial shadows on the backgrounds).

In order to investigate the robustness, we first changed the file format. That is, the marked video was represented into the avi format and then converted into the MPEG 4 and DivX formats; the detection procedure was successfully in both cases.

The method also features robustness against all the attacks included into the Stirmark software, applied at their standard parameters [3]: linear filters (convolution, median, and FMLR), JPEG compression, colour quantisation, row and column removal, image resizing. Regarding the rotations, without a registration procedure, the method we propose can face only small angles (lower than $2.5°$). The method also presents robustness against the StirMark attack.

Note that these results depend on the mark power and on the rank interval where the mark is to be embedded. Tables 1, 2 and 3 present a comparison among the three above mentioned rank intervals: $I_1 = [1;64]$, $I_2 = [50;113]$ and $I_3 = [129;192]$. The values in these three tables correspond to the following three mark powers: $1/16$ - Table 1, $1/256$ - Table 2 and $1/1024$ - Table 3.

From inspecting the information in Tables 1, 2 and 3, it may be stated that $I_3 = [129;192]$ and a $1/512$ power afford the best performances.

At a first glance, this experimental result is quite un-expected. Note that the I_1 and I_2 intervals contain larger coefficients and hence, they represents more accurately the video sequence; consequently they are *a priori* susceptible to be less altered by the attacks and to afford a better

detection. However, a sound theoretical support in this respect is afforded by the statistical video modelling we carried out [10]. In [10] we presented an original statistical approach which fusions the results obtained on multiple data sets sampled from the same video sequence and combines four types of statistical tests, namely: the Chi-square test on concordance, the Ro test on correlation, the Fisher F test on equality between two variances and the Student T test on equality between two means. The final results we obtained state that the Gaussian law can accurately model only the values taken by the ranks belonging to the [6;15] and [121;192] intervals. On the other hand, note that the detection rule (6) is, in fact, a matched filter to the embedded mark. Such a filter is optimal only when the noise (*i.e.* the 2D-DCT coefficients and the attacks) is Gaussian distributed. Consequently, when embedding the mark into coefficients which obey to the Gaussian law - *i.e.* $I_3 = [129;192]$ - the (6) detection rule approaches its optimality. Hence, these results on video watermarking can be considered as a first validation of the theoretical results reported in [10].

Tabel 1. Method performances corresponding to a mark power of $1/16$. The interval in the rank interval containing the mark is presented in the left column.

Interval	Transparency	Robustness					
		File format	Linear filter	Row&Colum removal	Frame resizing	Colour quantisation	StirMark
$I_1 = [1;64]$	No	Yes	Yes	Yes	Yes	Yes	No
$I_2 = [50;113]$	No	Yes	Yes	Yes	Yes	Yes	No
$I_3 = [129;192]$	No	Yes	Yes	Yes	Yes	Yes	Yes

Tabel 2. Method performances corresponding to a mark power of $1/512$.

Interval	Transparency	Robustness					
		File format	Linear filter	Row&Colum removal	Frame resizing	Colour quantisation	StirMark
$I_1 = [1;64]$	No	Yes	Yes	Yes	Yes	Yes	No
$I_2 = [50;113]$	Quality	Yes	Yes	Yes	Yes	Yes	No
$I_3 = [129;192]$	Fidelity	Yes	Yes	Yes	Yes	Yes	Yes

Tabel 3. Method performances corresponding to a mark power of $1/1024$.

Interval	Transparency	Robustness					
		File format	Linear filter	Row&Colum Removal	Frame resizing	Colour quantisation	StirMark
$I_1 = [1;64]$	Quality	Yes	Yes	Yes	Yes	Yes	No
$I_2 = [50;113]$	Fidelity	Yes	Yes	Yes	Yes	Yes	No
$I_3 = [129;192]$	Fidelity	Yes	Yes	Yes	Yes	Yes	No

The probability of false alarm was evaluated as being lower than $2 \times 10^{-16} \cong 16^{-14}$. Note that the detection procedure provides a 16 symbols recovered message $\hat{s}_i$, $i \in \{1,2,...,16\}$, no matter if the investigated sequence was marked or not; the probability that the 16 $\hat{s}_i$ recovered symbols would be identically with the 16 s_i considered symbols by chance is lower than 16^{-16}. However, as a consequence of the attacks previously described, in case of 1024 frames sequences (*i.e.* about 40 seconds of video), one or even two $\hat{s}_i$ recovered symbols might not match the corresponding s_i. It can be still considered that the video was marked, with a probability of false alarm lower than 16^{-14}.

When considering longer video sequences (*e.g.* 5000 frames or about 3 minutes) and when the detection procedure, independently applied to each sub-sequence of 1024 images is followed by a majority decision, there were no more errors left.

4. Conclusion

To conclude with, this paper presents a new watermarking algorithm for colour video protection. This method can afford *robustness*, *transparency* and *low probability of false alarm*. As the recovered watermark is represented on 64 bits, this procedure also makes it possible to identify a copy maker or, at least, a group of presumptive copy makers. The robustness is a consequence of the new way in which the mark is embedded into the 2D-DCT hierarchy and of the locations we found out as being optimal to bear the mark. The transparency results from both the very low power of the embedded mark and of non-linear post-processing transform we designed.

Note that such results were obtained on 100 different colour video sequences, arbitrarily chosen. They belong to different fields: art, news, amateur, *etc.* Hence, the numerical values involved in the method presentation feature a certain degree of generality.

On the other hand, we may consider that the post-processing transform may be of a larger interest: it can be involved not only in any watermarking scheme but in any video processing as well. We emphasise that although this transform is heuristically designed, it is compulsory to be applied: when skipping it over, strong artefacts are visible in the marked video.

As a final remark, we emphasise that the performances we reported were obtained by the procedure itself. We do not consider either error correcting codes or special counterattacks.

References

[1] I. Cox, M. Miller, J. Bloom, Digital Watermarking, Academic Press, 2002, ISBN 1-55860-714-5.
[2] G. Langelaar, I. Setyawan, R. Lagendijk, IEEE Signal Processing Magazine **17**(5), 20 (2000).
[3] F. Petitcolas, R. Anderson, M. Kuhn, Proc of the Second workshop on information hiding, David Aucsmith Ed., in Lecture Notes in Computer Science, Vol. 1525 Portland, USA, 1998.
[4] P. Moulin, Signal Processing, **81**(6), 1121 (2001).
[5] R. Pickholtz, D. Schilling, L. Milstein, IEEE Trans. on Communications **30** (5), 855(1982).
[6] J. Ó Ruanaidh, T. Pun, in Signal Processing **66**(3), 303 (1998).
[7] Al. Spătaru, Fondements de la théorie de la transmission de l'information, Presses Polytechniques Romandes, Lausanne 1987, ISBN 2-88074-133-0.
[8] I. Cox, et. al., IEEE Trans. on Image Processing **6**(12), 1673 (1997).
[9] M. Mitrea, F. Prêteux, A. Vlad, Proc. of The Tyrrhenian International Workshop on Digital Communications: Advanced Methods for Multimedia Signal Processing - IWDC 2002, Sept. 2002, Capri-Italy, pp. 311–318.
[10] M. Mitrea, F. Prêteux, A. Vlad, C. Fetita, Journ. Optoelectron. Adv. Mater. **6**(1), 95 (2004).
[11] 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.
[12] R. Prasad, CDMA for wireless personal communications, Artech House, 1996, ISBN 0-89006-571-3.
[13] W. Peterson, E. Weldon, Error correcting codes, MIT Press, 1972, ISBN 0-262-16039-0.