WATERMARKING VIDEO, HIERARCHICAL EMBEDDING IN MOTION VECTORS

Yann Bodo, Nathalie Laurent[*]

France Telecom R&D - DIH/HDM
4, rue du Clos Courtel - 35512
Cesson Svign Cedex - FRANCE

Jean-Luc Dugelay[†]

Institut EURECOM, Multimedia Com. Dept.
2229 Rte des Cretes-BP 193 F-06904
Sophia Antipolis Cedex, FRANCE

1. ABSTRACT

This paper proposes a new video watermarking scheme based on a hierarchical motion analysis, that is robust against classical video processing (filtering, lossy compression...). This model works in the uncompressed domain and disturbs the motion vectors computed by an exhaustive BMA on blocks of size n*n in order to insert an invisible watermark enabling a protection against most attacks. Finally, to increase the robustness, our system inserts the mark by generating a hierarchy of motion vectors of size N*N with N=k*n to spread the mark on the lower level associated to block n*n.

2. INTRODUCTION

With the development of communication systems, in particular the growth of the internet with adsl and cable, the exchange of multimedia content between people has increased. In this context, it is necessary to be able to control the life of value multimedia content. Many systems have emerged over the last few years, like the cryptographic system. However it does not protect against unauthorized copying after the content has been successfully transmitted and decrypted. This is the kind of protection that can be handled by watermarking. A digital watermark is in fact a piece of information inserted and hidden in the media content. This information is imperceptible to a human observer but can be easily detected by a computer. Generally the secret is based on a key and there are no a posteriori protection (for example the css system for DVD [1] has not resisted for a long time). In this context, watermarking algorithms appear to be a complementary solution which allows full protection. The watermarking is well-suited for the protection of multimedia content. Its applications are various and numerous, they concern copyright application, fingerprinting, "smart content" application, etc. In [4], the authors propose a good introduction on the watermarking world and in [5], the authors present a good mathematical formalization of the watermarking problem. Until now, most of the video watermarking systems purposed are based on the extension of still image algorithms. However they suffer from a lack of robustness. Another way is to use the intrinsic notion of the video i.e. the temporal information. Consequently, we can classify video watermarking schemes in two main categories: Still image based techniques and video-adapted techniques. Finally, few papers focus their interest on marking dynamic areas. Indeed, marking motion vectors has been first introduced by Kutter & al. in [2] in which the authors select a set of motion vectors over which they apply a parity rule to embed the mark. Zhang & al. in [8] used this principle to adapt the insertion rule by selecting the vector components that have the greatest magnitude. However, both methods suffer from serious drawbacks. Effectively a simple filter can destroy the parity of the motion vector components. In [6], the authors deal with marking motion vectors to be robust against transmission error on noisy and bandwidth limited channels. The goal of this article is not to protect media content but to provide better video quality. Thus the goal of this article consists in finding an insertion rule based on motion vectors able to resist to manipulation usually performed on video in real application (e.g. compression, filtering, etc.). Our system represents a semi-blind watermarking system. In fact, a marked vector must be maintained in a local space determined by the insertion rule to support attacks that will displace it around its initial position. This approach make it possible to resist filtering, compression and so on, that slightly displace marked motion vectors around their initial position. In section 3, we describe in detail our system by explaining the embedding and the retrieval processes. Then, in section 4 we present some results and we conclude in section 5.

3. TECHNICAL DESCRIPTION

A video stream is composed of a succession of still images. Thus, the watermarking of video sequences can be seen as an extension of image algorithm. One of the well-known techniques proposed in the video watermarking topic is the JAWS algorithm developed by T. Kalker & al.[3]. JAWS is based on simple operations allowing you to reach the real time requirement and in which the video is considered as a succession of still images. It is possible to watermark only the intra images of video sequence, that could explain the great number of watermarking scheme developed for still image adapted to video.

Nowadays, literature has only provided few watermarking algorithms that consider temporal information as a key advantage. Nor propose robust solution. Indeed, it seems natural to consider that the robustness of a mark can be greatly improved by considering the following two video properties:

- information amount (a video denotes a larger information than still images);
- motion information.

However, the insertion of the mark is also constrained by time complexity, compression constraints and new class of attacks such as collusion [1].

[*] yann.bodo, nathalie.laurent@francetelecom.com
[†] jld@eurecom.fr
[1] http://www.dvd-copy.com/

We can classify the video watermarking schemes into 2 main categories, still image adapted techniques and video adapted techniques. Our scheme belongs to the second one.

3.1. Embedding

The main functionality of our watermark algorithm consists in providing robust authentication of a Copyright Owner. The copyright information, composed by 8 bits, carried by the watermark could be generated by a secret key owned by the copyright owner. Thus this information could be inserted in the video by using a psycho-visual mask or a pseudo-random process. In this paper, only the rule insertion of that watermark is described, a lot of preprocessing of the mark, like error correcting code or cryptographic technique could be added to robustify the process. In fact, an exhaustive BMA estimation is applied to block n*n, with n=4. Thus, we determine a hierarchical motion vectors pyramid by averaging the fourth vectors associated to the four concatenated blocks B_1, B_2, B_3, B_4 as explained in the Figure 1. The motion vector associated to the block8x8 represents the father block in our approach. Thus, the insertion rule is applied on the averaged motion vectors associated to the blocks 8*8.

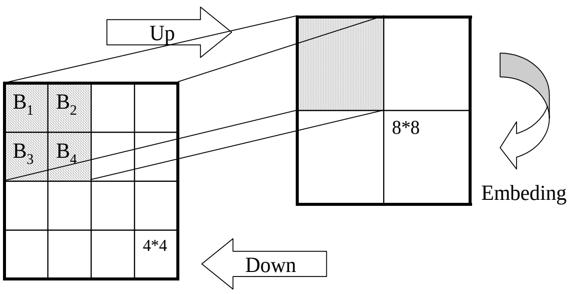

Fig. 1. Hierarchical scheme

For this, we select a subset of motion vectors to be marked. Today, this selection is a determinist one, however to improve the robustness of this approach and to minimize the visibility of the mark, we could use a pseudo-random selection or an adaptive selection based on a psycho-visual mask. In this approach, the embedding rule is defined by:

$$\forall \vec{d_f} = (d_f^x, d_f^y)^T \in \widetilde{V}_f, d_f^{\vec{W}} = \vec{d_f} + \widetilde{\Phi}(\alpha, \sigma_f(W), K_{\sigma_f(W)}) \quad (1)$$

where:

$$\widetilde{\Phi}(\alpha, \sigma_f(W), K_{\sigma_f(W)}) = \alpha \times \Upsilon(\sigma_f(W), K_{\sigma_f(W)}) \quad (2)$$

where $\widetilde{\Phi}$ and Υ are non reversible functions. To improve the robustness of this approach, the insertion rule must respect a spatial structure based on the construction of a reference grid $\mathcal{G}$ as illustrated on Figure 2. This rectangular grid is generated in the Cartesian space and is associated to a referential $(O, \vec{i}, \vec{j})$. It represents a block-based partitioning of the image compact support resulting in a set of block elements E, each of size $H \times K$ (in the case presented in this paper, $H = K = 7$) . Let us denote by R_i the intersection points between blocks that we call *reference points*.

Each selected motion vector of $\widetilde{V}_f$ is first projected on $\mathcal{G}$ and

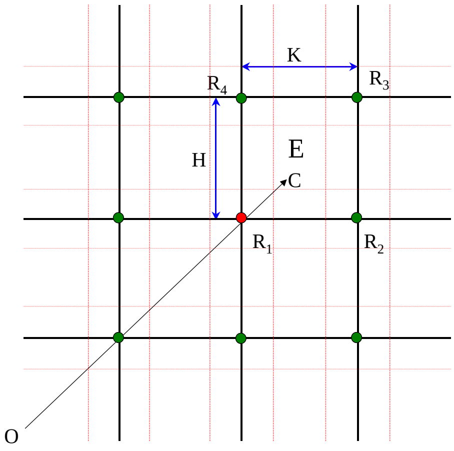

Fig. 2. Construction of a reference grid to embed a watermark on motion vectors

this projection serves to compute its associated reference point. The Figure 2 illustrates this process: the extremity of the projected motion vector $\vec{OC}$ belongs to a block E of $\mathcal{G}$ from which four intersection points R_1, R_2, R_3 and R_4 can be deduced. The reference point of the motion vector is the one which is the nearest to the extremity of the vector according to the L^2 distance. In the example of the Figure 2, the reference point of $\vec{d_f}$ is R_1.
Then, to embed the mark, the motion vector is modified (see Figure 3) by constructing in each block element E a rectangular element of size $h \times k$ (area Z_1), where $h = H - \delta_1$ and $k = K - \delta_2$ and δ_1 and δ_2 are chosen in order to have the same area covered by Z_1 and Z_2 and with $Z_1 \cup Z_2 = E$ (in the case presented here, $\delta_1 = \delta_2$ = 1). These two zones Z_1 and Z_2 drive the mark embedding rule: Z_1 is associated to the bit -1 and Z_2 to the bit $+1$.

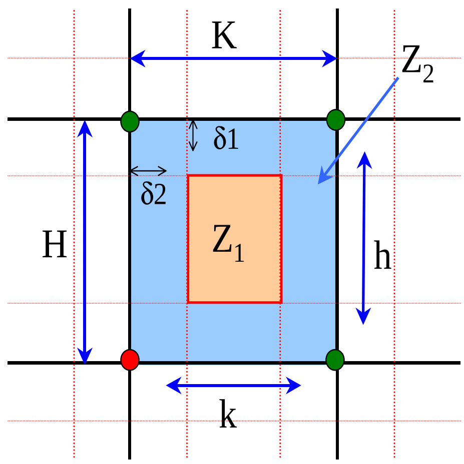

Fig. 3. Block element partitioning to embed the mark

Then, if we consider that $\vec{d_f} = \vec{OC}$ is the vector to be marked and W_i is the bit to be inserted, the marked vector $d_f^{\vec{W}}$ is computed

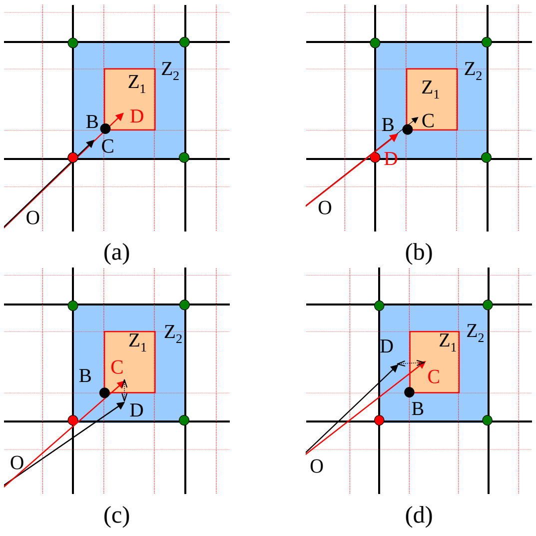

Fig. 4. Computation of the marked vector

as follows:

- if $W_i = -1$ and if $\vec{d_f}$ is in the right place (i.e. in the zone Z_1) then $d_f^{\vec{W}} = \vec{d_f}$; otherwise, a central symmetry of center B must be applied resulting in $d_f^{\vec{W}} = \overrightarrow{OD}$ (cf. Figure 4(a));
- if $W_i = +1$ and if $\vec{d_f}$ is in the right place (i.e. in the zone Z_2) then $d_f^{\vec{W}} = \vec{d_f}$; otherwise, as the Z_2 area is not compact, three possibilities can appear to compute $d_f^{\vec{W}}$:
  - $d_f^{\vec{W}}$ is given by a central symmetry of center B (cf. Figure 4(b));
  - $d_f^{\vec{W}}$ is given by an axial symmetry parallel to the y axis and going through B (cf. Figure 4(c));
  - $d_f^{\vec{W}}$ is given by an axial symmetry parallel to x axis and going through B (cf. Figure 4(d)).

The choice of symmetry must minimize the distortion of $\vec{d_f}$. In fact, after computing $d_x = C_x - B_x$ and $d_y = C_y - B_y$ (with $B = (B_x, B_y)^T$ and $C = (C_x, C_y)^T$), the symmetry is chosen as follows:

- if $d_x \leq \delta_2$ and $d_y \leq \delta_1$, the central symmetry is applied;
- if $d_x \leq \delta_2$ the axial symmetry parallel to the x axis is applied;
- if $d_y \leq \delta_1$, the axial symmetry parallel to the y axis is applied.

The modification realized on the highest level is then applied to the lowest level (down step). This approach allows us to create redundancy in the insertion phase, the watermarking scheme is consequently more robust. We realize a hierarchic spreading of the mark. To end, we perform a motion compensation. This step can be either performed on all of the blocks or either only on marked blocks and completed by original blocks. The second approach make it possible to avoid artifacts generated by only exploiting motion estimator, and in the same time increased the robustness of the detection process.

3.2. Retrieval process

The retrieval process correspond to the dual of the embedding process. Thus to detect the mark, we only had to apply algorithm1.

Algorithm 1 Mark detection algorithm

for f= 1 **to** N { //N denotes the video frame number
  for i = 1 **to** k_f {
    if $\vec{d_f^i} \in Z_1$ **then** $\sigma_f^i(W) = -1$;
    else if $\vec{d_f^i} \in Z_2$ **then** $\sigma_f^i(W) = +1$;}}

Once a candidate mark $\widetilde{W}$ is detected by the Algorithm 1, we must decide if it corresponds to the real embedded mark W. For this purpose, we compute the correlation C_f at frame f between $\widetilde{W}$ and W by the following recursion:

$$C_f = \frac{C_{f-1} \times (f-1) + (1 - \frac{d(\widetilde{W}, W)}{N})}{L} \qquad (3)$$

where L denotes the number of marked frames, $d(\widetilde{W}, W)$ denotes the Hamming distance between $\widetilde{W}$ and W and N is the mark length.
If $C_f \geq \theta$, where θ is a pre-defined correlation threshold, $\widetilde{W}$ is considered to correspond to W.

4. RESULTS

The above watermarking system has been tested on various video. In this section, we give some results obtained on the well-known sequences ”stefan” (250 frames) and ”ping-pong” (250 frames). These sequences are in the YUV format and their size are 288*352 (CIF format). We have conducted some experiments to check the robustness of our proposed watermarking scheme. For this purpose, we have performed some of the classical sequence manipulations including Divx[2] lossy compression, blurring with a uniform kernel and rotation. Additionally, we also have controlled the robustness of our algorithm to the new codec that appears nowadays, namely H264 [7]. The H264 retained model was IBP with quantization steps of 10 and 20. Moreover, all optimization parameters (eighth pixel motion estimation, five reference images, etc.) have been used. The compression ratio used for the sequence ”stefan” (”ping-pong” respectively) was 1:60 (1:60) for the Divx codec, 1:28 (1:40) for H264 IBP10 and 1:123 (1:154) for H264 IBP20. We have not tested the algorithm against frame dropping and rate changes yet.
The correlation results obtained with these attacks on the ”stefan” sequence are plotted on Figure 5 and on Figure 6 for the ”ping-pong” sequence. Let us recall that the correlation level for a frame index f tells us if the mark has been detected in f. On this figure, the correlation threshold θ has been set to $\theta = 0.875$. The average PSNR for the sequence ”stefan” is $43.7db$ and for the sequence ”ping-pong” is $44.9db$. These results show that the mark is quickly detected whatever the transform applied onto the sequence. The results obtained on the sequence ”ping-pong” are slightly worst than for the sequence ”stefan”. Indeed, we can note that for the Divx compression and for the ”IBP20” compression the threshold is not reached, but the curves seems to converge to the right

[2]http://www.divx.com/

mark. By analyzing these results, we can conclude that our watermark process is particularly robust to all tested attacks and few sequence images are needed to detect the embedded mark.

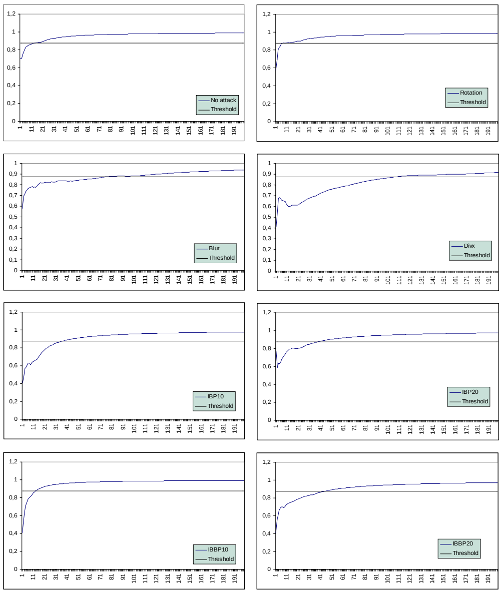

Fig. 5. Correlation results on the ”stefan” sequence

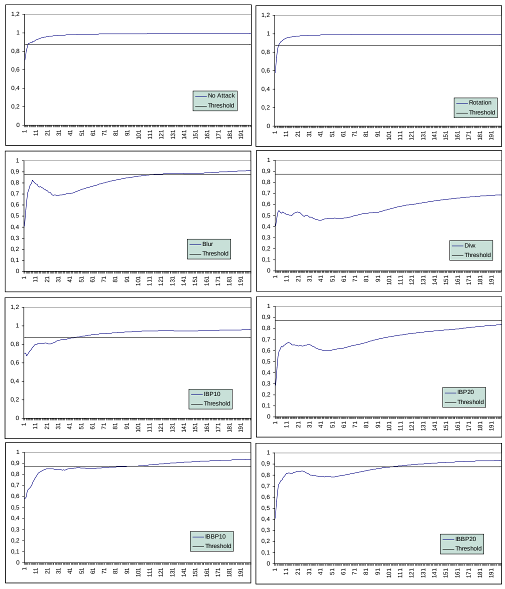

Fig. 6. Correlation results on the ”ping-Pong” sequence

5. CONCLUSION

In this paper, we have presented a new video watermarking algorithm that is robust to classical attacks. Effectively, preliminary experiments have shown the effectiveness of this algorithm. This approach allows us to exploit the notion of temporal axis, by inserting a watermark on motion vectors. The chosen insertion rule has proven that we are able to increase the algorithm robustness. In order to validate our approach, more tests must be done on various sequences and other attacks. Future works will deal with improving the detection criteria in order to reach more flexibility. Indeed sometimes the detector converges to the right mark but the correlation score does not reach the threshold fast enough. Finally, some improvements concerning the robustness had to be done as pretreatment on the mark (like cryptography rules or by using error correcting codes). Moreover due to the determinist approach, we actually have some visible artifacts on a few videos, thus currently we are working on HVS (Human Visual System mask) to exclude them.

6. REFERENCES

[1] Cox I.J. and Linnartz J.P.M.G. Some General Methods for Tampering with Watermarks. *IEEE Journal on Selected Areas of Communication*, 16(4):587–593, May 1998.

[2] Kutter M. Jordan F. and Ebrahimi T. Proposal of a Watermarking Technique for Hiding/Retrieving Datas in Compressed and Decompressed Video. In *ISO/IEC/JTC1/SC29/WG11/MPEG97/M2281*, 1997.

[3] Kalker T., Depovere G., Haitsma J., and Maes M. A Video Watermarking System for Broadcast Monitoring. In *Proceedings of the SPIE Conference on Security and Watermarking of Multimedia Content*, pages 103–112, San Jose, January 1999.

[4] Katzenbeisser S. and Petitcolas F.A.P. *Information Hiding*. Artech House, London, 2000.

[5] Pitas I. and Voyatzis G. The use of watermarks in the protection of digital multimedia products. *Proceedings of the IEEE*, 87:1197–1207, 1999.

[6] Song J. and Liu K.J.R. A data embedding scheme for H.263 compatible video coding. *Proceedings of the 1999 IEEE International Symposium on Circuits and Systems*, 4:390–393, July 1999.

[7] Wiegand T. Joint Final Committee Draft of Joint Video Specification. Standard, $(ITU-TRec.H.264|ISO/IEC14496-10AVC)$, 2002.

[8] Li J. Zhang J. and Zhang L. Video Watermark Technique in Motion Vector. In *XIV Brazilian Symp. on Computer Graphics and Image Processing*, pages 179–182, Florianopolis - Brazil, October 2001.