VISUAL QUALITY ASSESSMENT FOR MOTION VECTOR WATERMARKING IN THE MPEG-4 AVC DOMAIN

M. Hasnaoui, M. Mitrea, M. Belhaj, and F. Prêteux
Instiutut Télécom ; Télécom SudParis, ARTEMIS Department
9, rue Charles Fourier, 91011 Evry – France
{marwen.hasnaoui, mihai.mitrea, maher.belhaj_abdallah, francoise.preteux@it-sudparis.eu}

Abstract
Nowadays, to directly watermark a compressed video stream becomes a challenging research topic: low complexity, time saving and interoperability are the three main advantages of such an approach. The present paper deals with the MPEG-4 AVC motion vector watermarking and presents the first reference assessment for the visibility of the related artefacts. The experiments follow the most popular types of quality metrics in video processing: *pixel difference-based* (peak signal to noise ratio, absolute average difference, peak mean square error, and image fidelity), *correlation based* (structural content, normalized cross correlation, correlation quality, structural similarity metric), and *psycho-visual* (digital video quality). The numerical results correspond to a corpus of 5 video sequences of about 30 minutes each.
Key words: MPEG-4 AVC, *P* frames, motion vectors, watermarking, transparency, quality metrics.

I. Introduction

In order to enact piracy countering while keeping all liberties for a rightful user, watermarking solutions can be considered: an imperceptible (transparent) tracking message is inserted into the media to be protected [1-3]. Should the detection of this message be robust to mundane processing and malicious attacks, the pirate can be tracked down to the last legal buyer.
One of the nowadays video watermarking disadvantage is represented by its speed. On the one hand, highly performing watermarking algorithms require complex transformations on uncompressed media, thus becoming computationally complex and forbidding real time applications. On the other hand, services like video on demand depend on real time processing thus imposing this severe constraint on the related watermarking systems [4]. A solution to this deadlock may be represented by compressed domain watermarking systems. Their main advantage is related to their ability to insert the mark with minimal decoding of the compressed stream.
Unfortunately, watermarking and compression are two antagonistic desiderata. Compression eliminates the visual redundancy in the video sequence: as a limit, even the slightest alteration of a compressed stream would lead to visual artefacts. In its turn, watermarking exploits the visual redundancy in order to hide the mark.
The present study objectively establishes whether and at what extent the transparent additive watermarking can be achieved for the *P* frame motion vectors of the MPEG-4 AVC stream. From the watermarking point of view, the study is general: no particular *a priori* restrictions apply for the insertion rule or for the strategy followed when selecting the motion vectors to be marked.
The paper has the following structure. Section II outlines some basic syntax elements for the MPEG-4 AVC stream. Section III presents the visual quality metrics considered in the present study. The video corpus and the experimental work are described in Section IV. Section V brings into light the influence of the compression rate and of the video content on the artefact visibility while Section VI concludes the paper.

II. MPEG-4 AVC basic syntax elements

The MPEG-4 AVC compression standard, a.k.a. H.264, can reduce the redundancy existing in the natural video frames both in the temporal and spatial domains [5-7].
In the temporal domain, the redundancy among successive frames is reduced by motion compensation. In order to do this, the video is divided in GoPs (Group of Pictures). Each GoP contains at least one *I* frame, and a variable number (even 0) of *P* and *B* frames. The *I* frames are encoded by themselves, using image compression techniques. The *P* frames are motion-predicted from previous frames. The *B* frames are motion-predicted from both previous and subsequent frames. The *P* and *B* prediction relies only on *I* and *P* frames in the same GoP.
For the inter prediction, different sized blocks are first defined on the frame, according to its content (motion, level of detail, *etc*.). The block sizes, in pixels, can be 16×16, 16×8, 8×16, 8×8, 8×4, 4×8, and 4×4. All smaller blocks are obtained by splitting a parent block in two equal area blocks. Consequently, all blocks are subparts of 16×16 *macroblocks*, which are the basic

building blocks for the video encoding. Once its structure is defined, each block is assigned a reference frame and a motion vector giving a provenience offset in the reference frame. The difference between the pixels of the block and the pixels copied from the location pointed by the motion vector in the reference frame is then transformed by an integer DCT, quantized, and entropy coded. The set of motions vectors, reference frames, and transform-quantized differences can be used to rebuild the frame.

The study reported in the present paper evaluates the visual artefacts induced in the original (uncompressed) video by watermarking the motion vectors of the *P* frames. In other words, we shall evaluate by means of some popular quality metrics the differences between the original (uncompressed) video and the decompressed video obtained after some predefined alterations of the inter motion vector values.

III. Transparency quality metrics

Regardless the MPEG-4 stream modification strategy, the human observer will always watch a video content displayed (after decoding) on some device (computer screen, TV set, video projector …).

The MPEG-4 AVC considers frames represented in the YC_bC_r 4:2:0 format. However, the artefact effects are here investigated only in the Y (luminance) component, by considering several objective measures [8-9]; these measures are divided into three categories [10]:

- *pixel difference-based measures* (peak signal to noise ratio - PSNR, absolute average difference - AAD, peak mean square error - PMSE, and image fidelity - IF);
- *correlation based measures* (structural content - SC, normalized cross correlation - NCC, correlation quality - CQ, and structural similarity metric - SSIM);
- *psycho-visual measures* (digital video quality - DVQ).

Note that, except for the DVQ [11], all the other measures were adapted from image quality assessment, by computing an average over all video frames.

The pixel difference based measures seek a perfect equality between the tested and the reference video sequences. That is, even if the differences are not very high (for example, some white noise added on the pixels) they will be shown by these measures. However, these measures cannot discriminate between sparse and powerful distortions (which are visible) and frequent small distortions (which are invisible). The correlation based measures seek a statistical fit of the tested video on the reference. That is, if the tested signal resembles the original, even with some noise, the correlation based measures will reflect this resemblance. The DVQ is a sophisticated measure, taking into account both the filtering and the masking properties of the human visual system. First, both the reference and the tested videos are spatially and temporally filtered. Then, their difference is computed and the reference video is passed through a second filter in order to obtain the locations where the artefacts would be masked by the original data themselves.

In the present study, all these measure have employed the parameters and the working manner detailed in [12].

IV. Experimental setup

IV.1. Video corpus

The results reported in this study have been obtained by processing a corpus of 5 video sequences of about 30 minutes each.

The corresponding video content is heterogeneous, combining natural and synthetic scenes, movies, news, and sports [4]; examples of original and marked frames are presented in Figure 1.

Fig. 1. Original (left) and marked (right) frames.

The frame size is 352x288 pixels. These sequences were coded at three different rates:

- 64 kbit/s (low quality video, according to the mobile network usages);
- 512 kbit/s (medium quality video, corresponding to the Internet exchanges);
- 1024 kbit/s (high quality video).

The sequences were coded according to the *baseline* profile (no *B* frames), with 20 frames per GoP; hence, the GoP structure is $IP_1P_2\ldots P_{19}$.

In the experiments, only motion vectors corresponding to *P* frames (inter frame prediction) were considered. Consequently, these vectors are computed by a median predictor [5-7].

The experiments started by processing a particular video sequence coded at medium quality (Tables 1-4) and followed by establishing the generality of the results with respect to the compression rate (low and high quality video results are illustrated in Figs 2-5) and to the video content (the rest of the sequences are considered in the experiments reported in Figs. 6-9).

IV.2. Quantitative results

The study focuses on the artefact visibility corresponding to any additive watermarking schemes. In other words, it is assumed that a random value (the mark) is added to an arbitrarily chosen number of motion vectors corresponding to a *P* frame. No *a priori* restriction applies to the watermarked vector position in the *P* frame or to the considered *P* frame order in the GoP. Consequently, the visibility of the artefacts is a function of several variables:

- the amplitude of the modification, *cf.* Table 1;
- the number/position of the modified macroblocks into an individual *P* frame, *cf.* Tables 2 and 3;
- the position of the *P* frame in the GoP, *cf.* Table 4.

1. *The artefact visibility as a function of the modification amplitude* (*i.e.* the mark value)

Be there an arbitrarily chosen *P* frame position in the GoP (the 4th *P* frame in the numerical illustrations) and a single watermarked motion vector, corresponding to the macroblock located in the centre of this frame. The motion vector value was altered by addition with a binary random generated sequence, with different amplitudes: {-1, 1}, {-2, 2}, {-3, 3}, {-4, 4}, {-5, 5}.

2. *The artefact visibility as a function of the number of modified macroblocks in a P frame*

Be there an arbitrarily chosen *P* frame position in the GoP (the 4th *P* frame in the numerical illustrations) and consider a fixed {-2, 2} amplitude modification of a variable number of motion vectors. Three cases are illustrated in Table 2: (a) a single motion vector, corresponding to the central macroblock; (b) two motion vectors, corresponding to the upper-left corner and central macroblocks; (c) three motion vectors, corresponding to the upper-left, bottom-left and central macroblocks.

3. *The artefact visibility as a function of the position of the "watermarked" macroblock in the frame*

Be there a P frame position arbitrarily chosen (in Table 3, the 4th *P* frame in the GoP) and assume a modification of fixed amplitude {-2, 2} is applied to a single macroblock. The impact of the position of this macroblock in the artefacts is evaluated for 5 cases, namely the macroblocks corresponding to the four corners and to the centre.

4. *The artefact visibility as a function of the position of the watermarked frame in the GoP*

Be there a modification of fixed amplitude {-2, 2} and assume this modification is done on the macroblock located in the centre of the frame. The dependency of the artefacts with the frame *P* position in the GoP is illustrated for five positions, namely the first, the 4th, the 9th, the 14th, and the last (the 19th).

When inspecting the numerical values reported in Tables 1-4, it can be noticed that:

- all the considered watermarking scenarios (all mark amplitudes, all numbers/positions of watermarked macroblocks and all *P* frames positions in the GoP) resulted in transparent artefacts, regardless the type of quality metric;
- the pixel difference-based measures showed themselves more discriminative than the correlation-based measures (which were quite invariant); this is a consequence of the fact that the motion vector modifications lead to small and spread errors in the decoded video;
- as expected, the most sensitive quality metric is the DVQ: as the artefacts in the *P* motion-vector watermarking are induced by the motion content alteration, they are more likely to be spotted out by measures including temporal filtering than by "static" measures.

V. The generality of the results

As a general conclusion on the numerical values presented in the previous section, it can be said that each and every time watermarking is possible: all the objective measures have values corresponding to transparency.

In the sequel, the generality of these results will be investigated with respect to the compression rate and to the video content, *cf.* Figs. 2-9. Actually, the 4 above-presented experiments will be resumed on different video sequences (*i.e.* a different compression rate or a different visual content) and the relative variations (denoted by ε) with respect to the reference values presented in Tables 1-4 will be computed according to:

$$\varepsilon = \left| VQ_{reference} - VQ_{test} \right| / VQ_{reference},$$

where $VQ_{reference}$ represents the value for any of the investigated quality metrics reported in Tables 1-4 while VQ_{test} stands for the corresponding value computed on the new video sequence.

1. *The artefact visibility dependency with the compression rate (Figs. 2-5)*

The experiments 1-4 are resumed for the same visual content, this time coded at high ($r_1 = 1024\ kbit/s$) and low qualities ($r_2 = 64\ kbit/s$). The corresponding ε variations are illustrated in Figs. 2-5. Note that as the relative variations of the correlation-based measures and of the Image Fidelity were each and every time lower than 1%, they were not displayed in these figures.

2. *The artefact visibility dependency with the visual content (Figs. 6-9)*

The experiments 1-4 are repeated for the rest of 4 video sequences in the corpus. Figs. 6-9 illustrate the ε variations of the visual metrics with the visual content for two new video sequences (denoted by v1 and v2, respectively), coded at medium quality (512 kbit/s). Here again, the correlation-based measures and the Image Fidelity were not illustrated as they resulted in values lower than 1%.

Table 1: Artefacts visibility as a function of the modification amplitude (the modification is made on the motion vector corresponding to the central macroblock in the P_4 frame in GoP).

	PSNR (dB)	*AAD*	*PMSE (micro)*	*IF (mili)*	*SC (milli)*	*NCC*	*CQ*	*SSIM (mili)*	*DVQ*
{-1, 1}	66.52	0.01	3.06	999.90	999.91	0.99	77.13	999.89	37.94
{-2, 2}	63.37	0.02	5.50	999.98	999.98	0.99	77.12	999.86	43.21
{-3, 3}	61.82	0.02	7.30	999.84	1000.06	0.99	77.12	999.85	47.73
{-4, 4}	60.71	0.03	11.95	999.84	999.89	0.99	77.12	999.80	52.85
{-5, 5}	59.82	0.04	16.71	999.83	999.73	0.99	77.12	999.73	52.01

Table 2: Artefacts visibility as a function of the number of modified macroblocks (the modification has the amplitude {-2, 2} and is made on the motion vectors the P_4 frame in GoP).

	PSNR (dB)	*AAD*	*PMSE (micro)*	*IF (mili)*	*SC (milli)*	*NCC*	*CQ*	*SSIM (milli)*	*DVQ*
1 block	63.37	0.02	5.50	999.89	999.98	0.99	77.12	999.86	43.21
2 blocks	54.68	0.46	110.59	999.25	999.67	0.99	77.15	998.24	80.71
3 blocks	53.92	0.47	112.64	999.19	999.53	0.99	77.10	998.18	82.74

Table 3: Artefacts visibility as a function of the position of the macroblock in the frame (the modification has the amplitude {-2, 2} and is made on the P_4 frame).

	PSNR (dB)	*AAD*	*PMSE (micro)*	*IF (mili)*	*SC (milli)*	*NCC*	*CQ*	*SSIM (mili)*	*DVQ*
upper-left	58.08	0.44	110.27	999.19	1000.10	0.99	77.07	998.24	77.05
bottom-left	63.37	0.02	5.50	999.89	999.98	0.99	77.12	999.87	43.21
central	65.50	0.02	11.58	999.85	999.99	0.99	77.12	999.86	45.69
upper-right	67.40	0.005	1.00	999.98	1000.00	0.99	77.12	999.92	36.04
bottom-right	68.86	0.002	0.29	999.98	1000.00	0.99	77.13	999.97	35.11

Table 4: Artefacts visibility as a function of the position of the *P* frame in the GoP (the modification has the amplitude {-2, 2} and is made on the motion vector located at the centre of the frame).

	PSNR (dB)	*AAD*	*PMSE (micro)*	*IF (mili)*	*SC (milli)*	*NCC*	*CQ*	*SSIM (mili)*	*DVQ*
First	62.67	0.02	5.41	999.88	999.88	1.00	77.13	999.88	41.31
4th	63.37	0.02	5.50	999.98	999.98	0.99	77.12	999.86	43.21
9th	64.80	0.02	6.03	999.7	999.79	0.99	77.12	999.87	38.51
14th	67.25	0.01	2.56	1000.01	1000.01	0.99	77.12	999.91	36.62
19th	69.43	0.001	0.43	1000.01	1000.01	0.99	77.13	999.96	34.96

The values represented in Figs. 2-9 show that:

- the artefact transparency was each and every time validated;
- the visual quality metrics are more sensitive to the visual content than to the compression rate; this behaviour would suggest that the practical watermarking methods may work equally good for any compression rate but some parameters would be required to be set according to the visual content type;
- the position of the modified macro-blocks into an individual *P* frame lead to artefacts more unpredictable than the modification of the mark amplitude and or the position of the *P* frame in the GoP; for practical watermarking schemes, these results point to the fact that the *P* frame position and the mark amplitude can be randomly selected while some restrictions apply to the positions of the watermarked macroblocks in successive GoPs.

VI. Conclusion

This paper presents a general study (valuable for any additive watermarking technique) establishing that the motion vector watermarking in MPEG-4 AVC is possible, at least from the transparency point of view. These results are obtained on a large and heterogeneous video corpus and are validated for three representative compression rates (64kbit/s, 512kbit/s and 1024kbit/s).

The evaluation was based on nine intensively considered objective measures (PSNR, AAD, PMSE, IF, CQ, SC, NCC, SSIM, DVQ) and considered four modification strategies.

These results can be considered as a first step towards a proof of concepts for motion vector based watermarking: it proved that in each and every investigated case the transparency is supported by all the considered measures. However, the result presented in this paper should be followed by an investigation on the robustness.

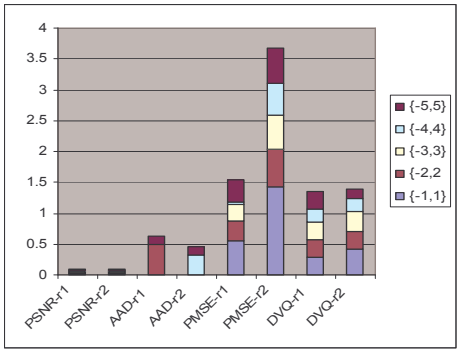

Fig. 2. Artefacts visibility variation as a function of the modification amplitude for a video sequence compressed at different rates.

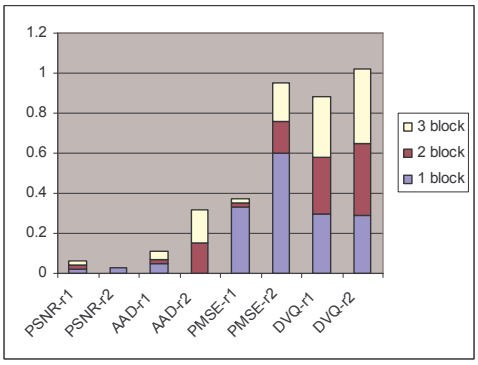

Fig. 3. Artefacts visibility variation as a function of the number of modified macroblocks for a video sequence compressed at different rates.

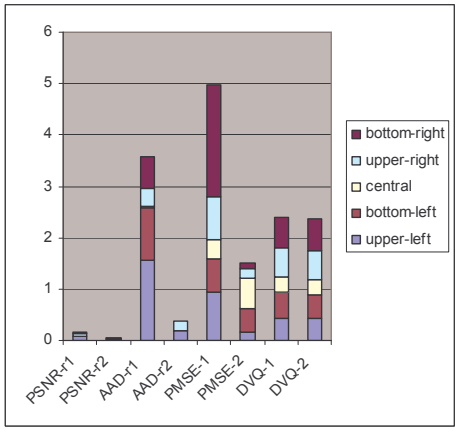

Fig. 4. Artefacts visibility variation as a function of the position of the macroblock in the frame for a video sequence compressed at different rates.

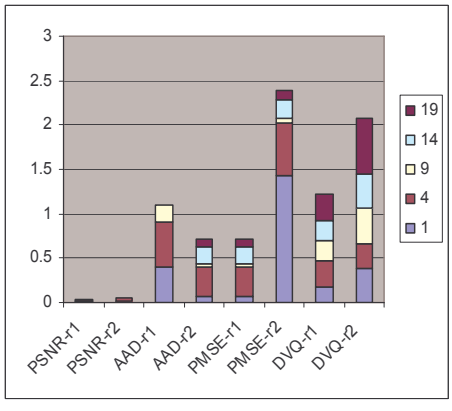

Fig. 5. Artefacts visibility variation as a function of the position of the P frame in the GoP for a video sequence compressed at different rates.

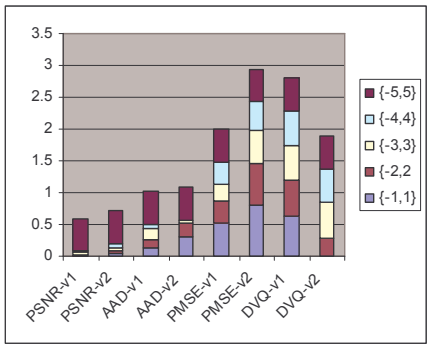

Fig. 6. Artefacts visibility variation as a function of the modification amplitude for different visual contents.

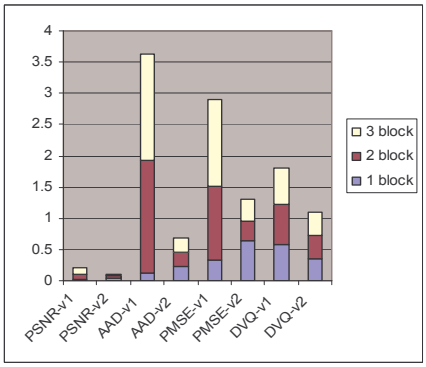

Fig. 7. Artefacts visibility variation as a function of the number of modified macroblocks for different visual contents.

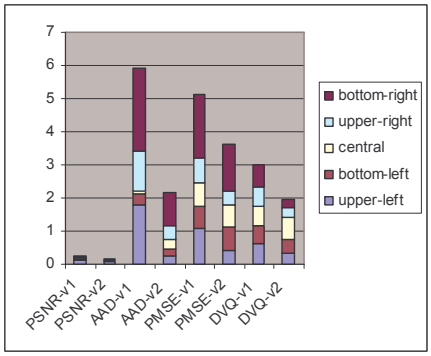

Fig. 8. Artefacts visibility variation as a function of the position of the macroblock in the frame for different visual contents.

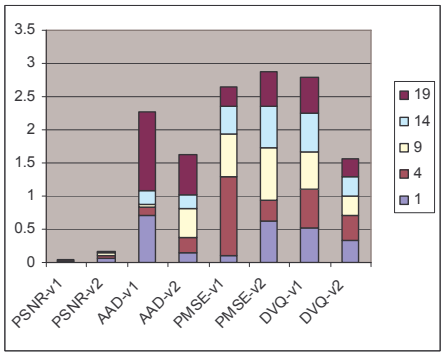

Fig. 9. Artefacts visibility variation as a function of the position of the P frame in the GoP for different viosual contents.

Acknowledgment

This study was supported by the MEDIEVALS project (*Marquage et Embrouillage pour la DIffusion et les Echanges Vidéos et Audios Légalisés et Sécurisés*) granted by the French National Research Agency (ANR).

References

[1] I. Cox, M. Miller, J. Bloom, *Digital Watermarking*, Academic Press, 2003.

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

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

[4] French National Project MEDIEVALS, http://www.medievals.org/

[5] ITU-R BT.601

[6] ISO/IEC 14496-10 (2005)

[7] T. Wiegand, G.J. Sullivan, G. Bjontegaard, A. Luthra, "Overview of the H.264/AVC Video Coding Standard", IEEE Trans. on Circuits and Systems for Video Technology, 13(7), 560-576, 2003.

[8] J. Zhang, H. Maitre, "Watermarking in the MPEG1-2 video sequences", Research report no. 2001D006, ENST Paris, 2001.

[9] Z. Wang, A.C. Bovik, H.R. Sheikh and E.P. Simoncelli, "Image quality assessment: From error visibility to structural similarity", IEEE Transactions on Image Processing, 13(4), 600-612, 2004.

[10] I. Avcibas, Image Quality Statistics and Their Use in Steganalysis and Compression, PhD thesis, Bogazici University, 2001.

[11] A.B. Watson, J. Hu, J.F. McGowan, "Digital Video Quality Metric Based on Human Vision", Journal of Electronic Imaging, 10(1), 20-29, 2001.

[12] S. Duta, M. Mitrea, F. Prêteux, M. Belhaj, "The MPEG-4 AVC domain watermarking transparency" Proc. SPIE 6982, p. 69820F, 2008.