THE PUBLISHING HOUSE
OF THE ROMANIAN ACADEMY

PROCEEDINGS OF THE ROMANIAN ACADEMY, Series A,
Volume 14, Special Issue 2013, pp. 328–337

SEMI-FRAGILE WATERMARKING BETWEEN THEORY AND PRACTICE

Mihai MITREA, Marwen HASNAOUI

Institut Mines-Telecom; Telcom SudParis, ARTEMIS Departement
9, rue Charles Fourrier, 91011 Evry France
Corresponding author: Mihai MITREA, E-mail: mihai.mitrea@telecom-sudparis.eu

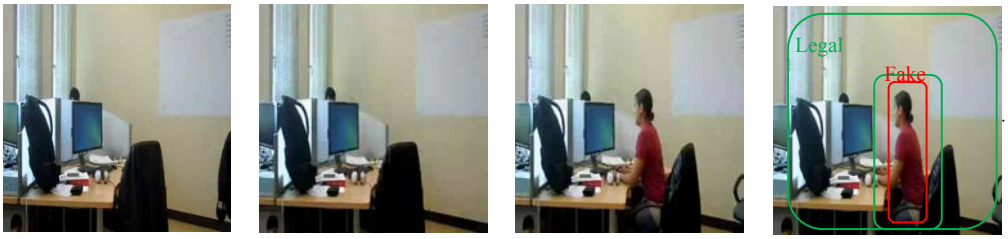
The present paper reports on a theoretical and experimental study on the possibility of achieving content-based MPEG-4 AVC video authentication. The authentication signature is extracted so as to both maximize the probability of correct decision and the mutual information under content-preserving attacks. The experiments (carried out on 1h of video surveillance corpus) demonstrate that the content-modification attacks can be identified with *Precision* and *Recall* rates larger than 0.9 while ensuring a 9s temporal accuracy and an 1/81 frame size spatial accuracy.

Key words: video surveillance, integrity, semi-fragile watermarking, MPEG-4 AVC.

1. INTRODUCTION

Nowadays, digital video is the core of various applications, ranging from home entertainment (*e.g.* video on demand) to civil/state security (*e.g.* video surveillance). Despite the particular type of application, the source/content authentication is always required. On the one hand, video on demand platforms distribute their content under strict copyright constraints: hence, the traceability of source/intermediate distribution points should be kept. On the other hand, video surveillance helps police investigations and may subsequently serve as a piece of evidence in courts: hence, both its source and content should be certified.

Two approaches can be considered in order to ensure video authenticity, namely data and content based. The former considers the data (binary) representation of the video and extracts an authentication signature (*e.g.* a hash function) which meets the uniqueness and sensitivity (fragility) requirements. This signature is further stored as metadata. The latter no longer targets the binary representation of the video but its visual/semantic content. In order to be effective, such a signature should be robust to content preserving alterations and sensitive to content changing alterations. Consider the example in Fig 1. Fig 1.a represents an original content while Fig 1.b shows its JPEG compressed version at quality factor $Q = 70$; Fig 1.c gives a content-modified version of the image in Fig. 1.b in which a person has been added. From the binary representation point of view, the three images in Figs. 1.a, 1.b and 1.c are completely different, hence their data-based signatures should be different. However, from the semantic content point of view, Fig 1.a and Fig 1.b are identical while differing from Fig 1.c. Consequently, the content-based signatures corresponding to Figs. 1.a and 1.b should be identical while differing from the signature extracted from Fig. 1.c.



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

Fig. 1 – The original frame (a) suffers a content preserving attack (a compression) (b) then a content alteration attack (the insertion of a person) (c). Signature based integrity verification should discriminate between the legal/fake areas (d).

Content-based authentication methods can be classified as passive or active [1].

Passive authentication, also called forensic analysis [2] attempts to verify whether the content has been altered or not by using a direct statistical analysis. This way, no a priori processing/storing operation is required for the original content. While very appealing by its functional principle, passive authentication is not always possible and doubts about its reliability and security are raised for some applications [2].

This is not the case of the active approach, where the integrity of digital content is ensured by the embedding of an authentication signal (or signature) in the content itself, before its sharing/distribution. The active authentication is also called watermarking based authentication. Consider again the example in Fig. 1 and assume that Fig. 1.a carries (in a visual imperceptible way) a content authentication signature. On the one hand, this signature should be robust to compression; hence it should be recovered from the unaltered content areas in Fig. 1.c. On the other hand, it should be fragile against content-modifications; hence, it should no longer be recovered from the areas where the person was inserted. This way, the legal/faked areas can be discriminated into a same video frame (see Fig. 1.d).

From the applicative point of view, the main semi-fragile watermarking difficulties are the choice of the signature and of the embedding technique. The signature must ensure the semi-fragility property by being both sensitive to content changing alterations (frame cropping, object removing, …) and robust to content preserving alterations (transcoding, resizing, ...). As video sequences are preponderantly stored and distributed in compressed format, the signature should be both generated and inserted directly from/in the compressed domain, thus avoiding the computational overhead required by the decompression/compression.

In this paper, we focus on the MPEG-4 AVC (Advanced Video Coding) [3] integrity verification. First, the MPEG-4 AVC syntax elements that jointly reflect the semantic content and ensure the semi-fragility propriety are identified. The theoretical supported is granted by information theory tools. Secondly, the *m*-QIM (*m*ultiple symbols Quantization Index Modulation) insertion technique (whose optimality in proved in [4]) is considered in order to insert the mark. The experiments are carried out on 1 hour of video surveillance content and show that the proposed video watermarking based video integrity verification system is able to distinguish between content preserving and video content changing alterations.

2. RELATED WORK

Watermarking based integrity verification was already the subject of several studies [5-10]. Several insertion domains are considered: still images [5], MPEG-1/2 [6, 8, 9] and MPEG-4 AVC [7, 10].

Titman [5] and Queue [6] use image edges and corners to generate the authentication signature. The signature is embedded according to additional modification rules of overlaying 8x8 blocks. They show that these features are sensitive to content changing alteration. Beside this advantage, the method is still fragile against compression and scaling. Moreover, such deployed signatures require complex generation operations and remain of large size, thus imposing particular constraints on the watermarking insertion method.

In [7], the signature generation is based on a chaotic system. Temporal authentication information (frame index and GOP index – Group of Pictures) is used to compute the signature for each *I* (Intra) frame. The experiments carried out on a video sequence of 795 frames proved the detection of temporal changes, but the properties of spatial detection of alterations have not been evaluated. The robustness to compression (up to 30%) was also shown.

S. Thiemert *et al.* [8] calculate the authentication signature in the uncompressed domain, from the points of interest obtained through the Moravec operator [11]. A binary mask is generated for each frame *I*, then embedded into the high-frequency DCT coefficients belonging to adjacent frames. The method detects content changing alterations (object removing/inserting) while being robust to content preserving manipulation (compression up to 50%, scaling). However, the authentication signature calculation increases the method complexity. The study in [9] resumes and extends these principles. The difference relies on the use of the entropy of gray level in groups of blocks to generate the binary signature. The advanced method is robust against compression down to 50% and detects content changing alterations (object removing). Nevertheless, the signature increases the complexity of the video integrity verification system.

K. Ait Saadi *et al.* [10] consider a signature generated from low frequency quantized DCT coefficients. For each *I* frame, the low frequency DCT coefficients are collected in a buffer to be further hashed using an

MD5 function. This results into a 128 bit binary signature. The obtained signature is embedded in the *P* (Predicted) and *B* (Bidirectional) motion vectors. Experiments show that this system remains fragile to all manipulations. Moreover, the signature generation requires an MPEG-4 AVC entropic decoding.

The state-of-the art brings to light that the trade-off between fragility, robustness and complexity is not yet reached in the compressed domain. Moreover, the signature is heuristically generated, without any theoretical support.

In order to minimize the overhead induced by the decoding/encoding operations, the present study generates the signature directly from the MPEG-4 AVC syntax elements. From the theoretical point of view, the optimal syntax elements are identified by a study carried out on the basis of information theory, *cf.* Section 3. From the applicative point of view, a semi-fragile system based on this optimal signature and the *m*-QIM insertion rule is assessed within the SPY ITEA2 project, *cf.* Section 4. Conclusions are drawn and perspectives are opened in Section 5.

3. THEORETICAL INVESTIGATION ON THE AUTHENTICATION SIGNATURE

This section investigates the MPEG-4 AVC compressed stream, in order to build a syntax element based signature. The targeted signature must meet the trade-off between the robustness against content preserving and the sensitivity to content changing alterations.

3.1. Syntax elements identification

MPEG-4 AVC [3] video sequences are structured into group of pictures (GOP). A GOP is constructed by fixed number of successive images of three main types (*I*, *P*, and *B*) as illustrated in Fig 2.

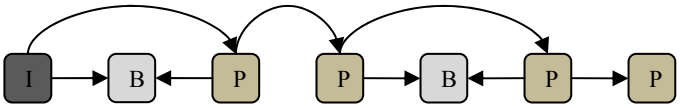


Fig. 2 – A GOP structure example.

An *I* frame describes a full image coded independently, containing only references to itself. Secondly, the unidirectional predicted frames *P* use one or more previously encoded/decoded frames as reference for picture encoding/decoding. Finally, bidirectional predicted frames *B* consider in their computation both forward and backward reference frames. According to the coding principles, *I* frames preserve more information and require more bits for encoding than the other two types.

Our study focuses on the *I* frames. The *I* frames contain the salient visual/semantic information which is also exploited by the *P* and *B* frames in that GOP. Consequently, extracting the signature from the *I* frames has two main *a priori* advantages: the signature can be related to the video semantic and represent the whole GOP. Fig 3 details the *I* frame encoding block diagram. MPEG-4 AVC transforms the uncompressed data in a classical compression chain: prediction P, transformation T, quantization Q and entropic coding E.

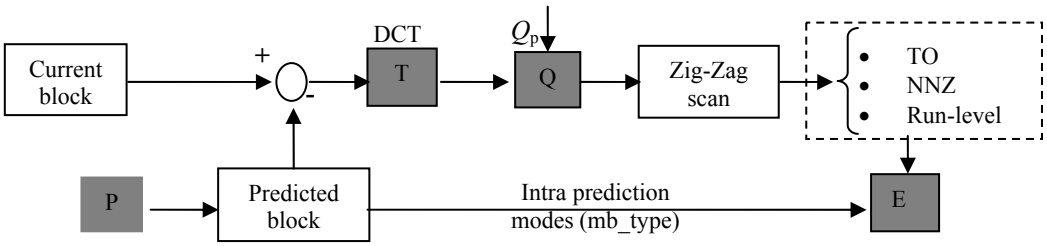


Fig. 3 – Intra frame coding diagram.

I frames are encoded according to Intra prediction modes which exploit the spatial redundancy to enhance the compression efficiency. The MPEG-4 AVC standard offers 13 directional intra prediction modes. For each current block, a predicted block is constructed from the boundary pixels of the neighboring

blocks which are previously encoded. MPEG-4 AVC offers two intra prediction modes classes: *I*16x16 (with four prediction modes) for smoothed regions and *I*4x4 (with 9 prediction modes) for the textured regions. For each block, the prediction mode minimizing the rate-distortion cost is selected.

The residual block computing the difference between the current block and the predicted block is transformed by using a DCT and a quantizer. Each quantized transformed residual block is further mapped into a 15 coefficients vector. In the baseline profile, the resulting vector is encoded by the CAVLC (Context Adaptive Variable Coding Length) encoder [3]:

– CAVLC uses the run-level entropic encoding to represent compact zero sequences.
– The non-zero values taken by coefficients are often expressed in ± 1 sequences. CAVLC encodes the ±1 number of the large compact succession of ±1 (Trailing-Ones).
– The number of non-zero coefficients (NNZ) is encoded using a lock-up table. The encoding table is determined based on the NNZ of neighboring blocks.
– Levels values of nonzero coefficients in low frequencies are larger than those in high frequencies. CAVLC takes advantage of this behavior by adapting the choice of the encoding table (VLC look-up table). The standard provides four encoding tables, depending on the NNZ value, see Tab 1.

The *I* frames contain four main syntax elements per intra block: mb_type, TO (Trailing Ones), NNZ and run-level. The extraction of these elements is performed through a syntax parser. This study will not consider the run-level. In fact, to be effective, the signature should have a limited alphabet size; this is not the case of run-level element. Tab 1 illustrates the syntax elements that will be investigated in this study.

Table 1

MPEG-4AVC syntax elements

Syntax elements	Description
mb_type	Intra prediction mode class type : (*I*4x4 or *I*16x16).
TO (Trailing Ones)	It takes a value from 0 to 3. If there is a succession of more than three ±1, only the last three are considered.
NNZ (Number of Non Zero coefficients)	NNZ encoded according to 4 tables are considered based on its value: (Table 1 (0, 1), Table 2 (2, 3), Table 3 (4, 5, 6, 7) and Table 4 (8 or more)).

3.2. Syntax elements behavior to content preserving attacks

3.2.1. Experimental protocol

This section investigates the possibility of building an authentication signature from the 3 syntax elements identified above (mb_type, TO, NNZ). To do this, the behavior of each of these elements to content preserving attacks (transcoding, Gaussian filtering, sharpening and scaling) is first investigated. Then, the considered attacks are applied according to two scenarios: (S1) the MPEG-4 AVC encoder is allowed to choose the quantization parameter *Qp* and (S2) the quantization parameter *Qp* is set to 31. Tab 2 shows the steps involved in both scenarios.

Table 2

Test scenarios

S1	S2
(1) Encode the video at variable *Qp*	(1) Encode the video at *Qp* = 31
(2) Extract the syntax elements and store them	(2) Extract the syntax elements and store them
(3) Attack the video and re-encode it according to (1)	(3) Attack the video and re-encode it according to (1)
(4) Extract the syntax elements and compare them to those extracted at (2)	(4) Extract the syntax elements and compare them to those extracted at (2)

First, the syntax elements (mb_type, TO, NNZ) are extracted using a syntax parser and stored as signatures. Secondly, the video sequence is attacked (transcoded, filtered, scaled, …) and further re-encoded according to the same initial configuration. Finally, the syntax elements are extracted from the attacked video and compared to those extracted previously from the original video. The syntax elements extracted from each 4x4 block intra are coded as follows:

$$\text{mb\_type} = \begin{cases} 0 & if\ I4\times 4 \\ 1 & if\ I16\times 16 \end{cases} \qquad \text{TO} \in \{0,1,2,3\}$$

$$\text{NNZ} = \begin{cases} 0 & if\ \ \text{NNZ} \le 1 \\ 1 & if\ \ 1 < \text{NNZ} \le 3 \\ 2 & if\ \ 1 < \text{NNZ} \le 7 \\ 3 & if\ \ \text{NNZ} > 7 \end{cases}.$$

3.2.2. Corpus

The experiments were carried out on a video surveillance corpus composed of 6 sequences of 10 minutes each, downloaded from internet [12] or recorded under the framework of the SPY project. This corpus is encoded in MPEG-4 AVC in Baseline Profile at 512 kbps, 640x480 pixel frames; the GOP size is set to 8.

3.2.3. Robustness against content preserving attacks

The initial value of a syntax element is likely to change after an attack on a random basis, given by both the attack and the content itself; hence we can investigate these modifications by modeling the attack with noise matrices. In such a matrix, the lines correspond to the values of original elements, while the columns to the values after attacks. An element in a matrix is the corresponding conditional probability, estimated on the corpus. These noise matrices are estimated by successively applying four content preserving attacks (transcoding, sharpening, Gaussian filtering and scaling) according to the two scenarios presented above. The size of the corpus was large enough so as to ensure the statistical relevance of the results: for each syntax element, 95% confidence limits are computed with relative error $\hat{\varepsilon}_r < 0.005$.

By further computing the probability of correct detection (P_c) and the mutual information (I) the related decision can be made on the optimal syntax element.

$$P_c = \sum_{i=1}^{N} P_{ii} \times P_i \,, \qquad I = \sum_{i=1}^{N}\sum_{j=1}^{N} P_{i,j} \times P_i \log(P_{i,j}/P_j) \,,$$

where $P_{i,j}$ presents the noise matrix element of coordinates i and j, P_i is the average probability that the syntax element takes the value i in the original video and P_j is the average probability that the syntax element takes the value j in the attacked video. N is the size of the corresponding alphabet (*i.e.* $N = 2$ for mb_type and $N = 4$ for NNZ and TO).

Tab 3 and 4 illustrate the distribution probability and the noise matrices for mb_type. The P_c and I values for each syntax element and for each tested attack according to the two scenarios are plotted in Figs 4 and 5, respectively. The analysis of these results brings to light two main conclusions:

– The probability of correct detection and the mutual information values show that the syntax element (mb_type) remains more robust than the two other syntax elements (TO and NNZ) against all the investigated attacks. The obtained averaged P_c over the 4 attacks and the 2 scenarios are 0.92, 0.76 and 0.47 for mb_type, NNZ and TO, respectively. I values feature an average of 0.55, 0.44 and 0.15 for mb_type, NNZ and TO, respectively;

– A better robustness is achieved for (S2). A fixed quantization step increases the mb_type correct detection probability by 0.06, 0.02, 0.04 and 0.03 and the mb_type mutual information by 0.21, 0.13, 0.01 and 0.02 for transcoding, sharpening, Gaussian filtering and scaling, respectively.

Table 3

mb_type distribution probability

	S1		***S2***	
	0	***1***	***0***	***1***
P	0.62	0.38	0.54	0.46

Table 4

mb_type transition matrix

	Transcoding				Sharpening				Gaussian filtering				Scaling			
	S1		*S2*		*S1*		*S2*		*S1*		*S2*		*S1*		*S2*	
	0	***1***	***0***	***1***	***0***	***1***	***0***	***1***	***0***	***1***	***0***	***1***	***0***	***1***	***0***	***1***
0	0.95	0.04	0.98	0.02	0.78	0.22	0.82	0.18	0.94	0.06	0.95	0.05	0.92	0.08	0.93	0.07
1	0.11	0.88	0.03	0.96	0.06	0.94	0.03	0.96	0.15	0.85	0.05	0.94	0.12	0.88	0.07	0.93

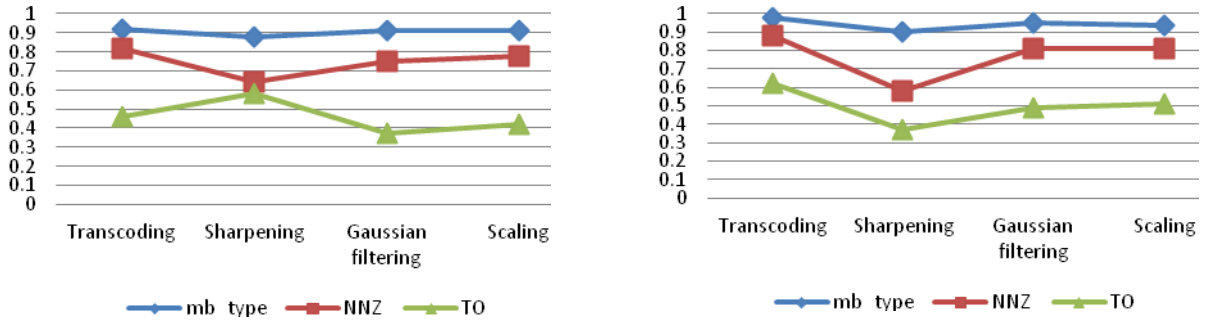


Fig. 4 – Probability of correct detection, for (S1) – left and (S2) - right.

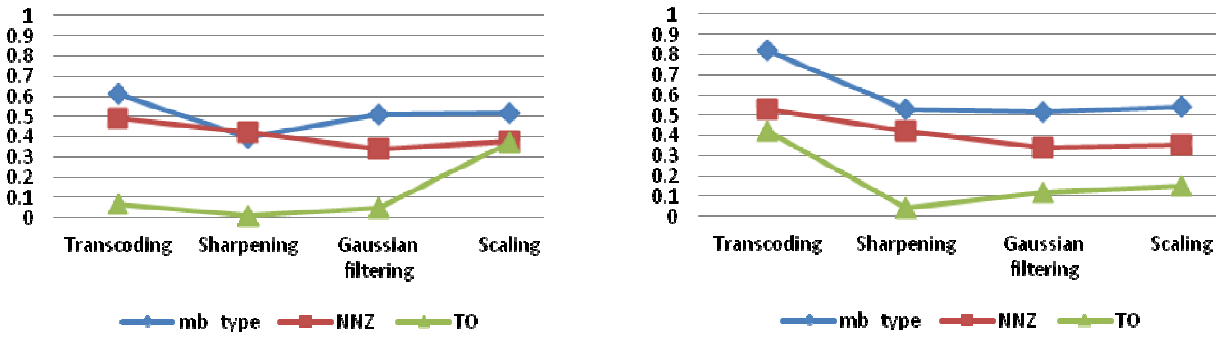


Fig. 5 – Mutual information, for (S1) – left and (S2) - right.

These results prove that the mb_type is the single element able to satisfy the constraint of robustness over the three investigated syntax elements. Thus, the rest of the study will concern only the mb_type elements.

3.3. Sensitivity to content changing alterations

This section investigates the sensitivity of mb_type syntax element to content changing alterations. In this study, the content changing alteration attack is performed by removing a person as illustrated in Fig 6-a and Fig 6-b. Just for illustration, such an attack can be performed by any non-professional user on a home PC, with software available on Internet; it takes about 10 minutes to process 1 second of video.

First *I* frame of original video (a) First *I* frame of attacked video (b) Content alterations detection (c)

Fig. 6 – Content changing alterations.

In order to spatially localize the content manipulations, an alteration detection image is generated. When the mb_type is detected as changed, the pixel value corresponding to it is set to red. The resulting detection image is illustrated in Fig 7. To spatially locate alterations, positions reported as manipulated by the detection matrix are further projected onto the inspected image, see Fig 6-c. It can be noticed the presence of false alarms and that the red block are denser in the altered area than other areas. To enhance the alteration detection and avoid the false alarms, the following filter is applied to the detection image:

$$M(i,j)=\begin{cases}1 \ \ if \ \ \sum_{i_0}^{a}\sum_{j_0}^{b} M(i+i_0, j+j_0) > s \\ 0 \qquad\qquad\quad otherwise\end{cases},$$

where M is the alteration detection image, $a \times b$ presents neighborhood window filter size and s is the decision threshold. i_0 and j_0 present the line/column indexes in the filter window. Fig 7 illustrates the evolution of the detection image as function of the decision threshold and Fig 8 reports the obtained false alarm probability as a function of $s = 3$. We note that from a given threshold ($s = 3$ in our case) false alarms disappear and the altered area is surrounded. We also notice that the optimization of the parameter s may be mandatory to improve the spatial accuracy of the alteration detection according to the targeted application.

M for *s* = 0 (a) *M* for *s* = 1 (b) *M* for *s* = 2 (c) *M* for *s* = 3 (d)

Fig. 7 – Alteration detection matrix.

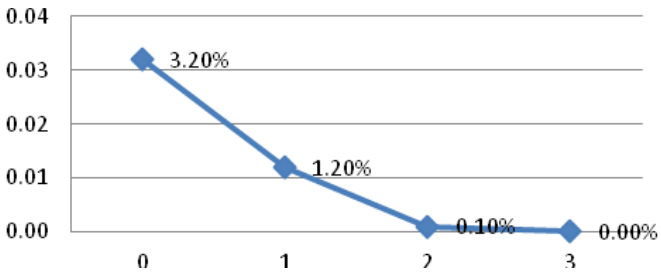


Fig. 8 – False alarm as function of s.

3.4. Conclusion

This section investigates the MPEG-4 AVC syntax elements with respect to their potential usage as authentication signature robust to content preserving attacks and sensitive to content changing. After an *a prior* study, 3 syntax elements (mb_type, NNZ and TO) are identified and tested by using information theory based entities. First, noise matrices, P_c and I demonstrate that mb_type is the optimal choice for content preserving attacks. Secondly, mb_type proved to be able to result into probability of false alarms equal to zero under content changing attacks. The next section will present a semi-fragile watermarking method whose inserted signature is based on mb_type syntax element.

4. APPLICATION VALIDATION OF THE SEMI FRAGILE WATERMARKING

This section details the deployment of mb_type based signature for building an MPEG-4 AVC video integrity verification system and subsequently evaluates its performances under the SPY[1] project framework.

4.1. Method presentation

The authentication signature is generated by encoding mb_type to an *m*-ary alphabet.

The obtained signature is further inserted by an *m*-QIM watermarking technique [4] in the video content. The low complexity requirement can be met when such a signature is extracted and inserted directly from/in the MPEG-4 AVC syntax elements, with minimal decoding/re-encoding operations. To meet this requirement, we consider individual groups of *i* successive *I* frames (further referred to as *I*-Group) sampled from an MPEG-4 AVC video sequence. The signature is computed from the first I_0 frame in such an *I*-Group. The obtained signature conveying information about the content of the first frame is further inserted into the rest $i-1$ *I* frames of that *I*-Group by the *m*-QIM embedding method. The shorter the *I*-Group, the more accurate the temporal localization of altered content.

In order to verify the integrity of an attacked *I*-Group, the authentication signature of the attacked video is computed from its first I_0 frame. This signature is compared to the mark extracted from the rest of the *I*-Group frames. In our experiments, we consider that an area in a frame is altered when at least s (*cf.* Section 3.3) elements of the attacked signature elements belonging to that area do not match with the corresponding extracted watermark elements.

4.2. Functional evaluation

4.2.1. Robustness

In videosurveillance context, transcoding is the main harmless authorized attack. While the Section 3.2.3 investigated the effects of this attack at the signature feature level, we are now assessing the global effectiveness of the video integrity system. In this respect, the watermarked sequence was subject to a transcoding attack applied according to the S2 scenario described above. In order to identify the spatial content alterations, the detection procedure was applied on areas obtained by partitioning the I_0 frames with a 9x9 equidistant rectangular grid (see Fig 9).

This set-up allows the robustness to be objectively assessed by the probabilities of missed detection (*i.e.* the probability of not detecting a watermark from an initially marked area), and false alarm (*i.e.* the probability of detecting a mark in an initially un-watermarked area).

Our experiments showed that the re-encoding from 512 kbps down to 128 kbps resulted in no content modification, thus demonstrating the robustness of the proposed system, with ideal values for the probabilities of missed detection and of false alarm (P_m , P_f).

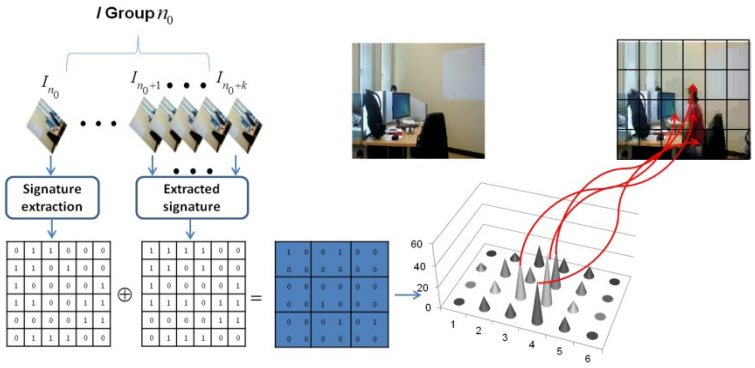


Fig. 9 – Spatial alteration detection.

[1] The ITEA 2 SPY (Surveillance imProved sYstem) European project aims at creating a new automated intelligent surveillance and rescue framework adapted for mobile environments.

4.2.2. Fragility

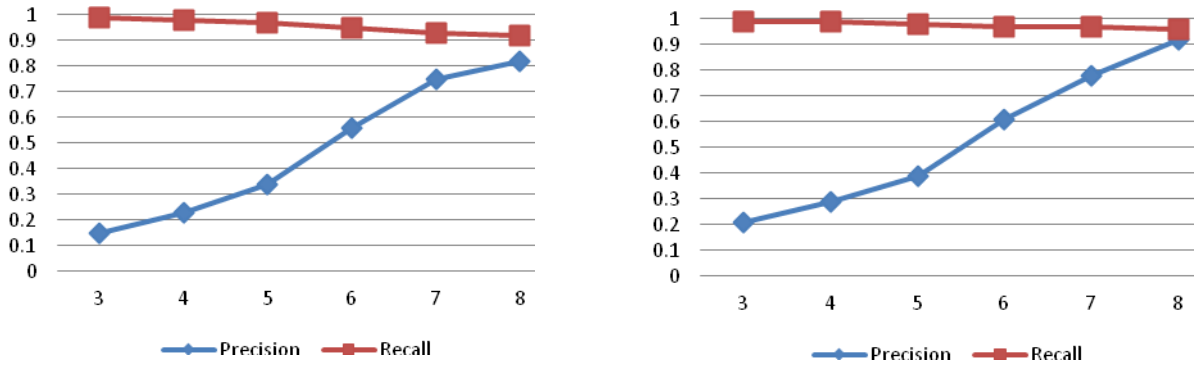
This Section investigates the sensitivity to content changing attack. The content is considered as being attacked when one object is moved, deleted or substituted. To simulate this attack, we used a piece of code that tampers the videos by arbitrarily changing 1/16 of the frame content. For each video sequence in the corpus, we applied such an attack to sequences of successive frames (between 9s and up to 3min).

In order to spatially locate alterations, we kept the same conditions as in Section 4.2.1: the I_0 frames in each *I*-Group are portioned in 81 areas, according to 9x9 equidistant rectangular grid. From the fragility point of view, an ideal watermarking method will fail in detecting the mark from each and every area which was subject to content alterations. While such a behavior can be also expressed in terms of probability of missed detection and false alarm, the literature bring to light two more detailed measures, namely the precision and the recall ratios, defined as follows [13]:

$$Precision = \frac{tp}{tp + fp}, \quad Recall = \frac{tp}{tp + fn},$$

where tp is the number of true positive (*i.e.* the number of content modified areas which do not allow the mark to be recovered), fp is the false positive (*i.e.* the number of content preserved areas which do not allowed the mark to be recovered) and fn is the false negative number (*i.e.* the number of content modified areas which allowed the mark to be detected).

Fig 10-a illustrates the obtained precision and recall average values as a function of the threshold s (*cf.* Section 3.3). The experiments exhibit $Precision = 0.81$ and $Recall = 0.92$ at $s = 8$ (*i.e.* more than 50% of area syntax element size). As these average measures are quite far from the ideal cases ($Precision = Recall = 1$), we went further in our investigation. Fig 11 illustrates the temporal detection of alterations: the abscissa corresponds to the *I-Group* index while the ordinate is set to 1 for the *I-Groups* identified by the system as being modified. The content attacked sub-sequences are circled in blue.



(a) (b)

Fig. 10 – Precision and recall as function of *s*: initial method (a) and method with post processing (b).

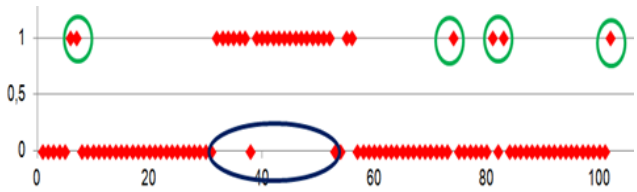


Fig. 11 – Temporal alteration detection.

Fig 11 shows that almost all altered *I*-Groups have been detected; however, some content-preserved *I*-Groups (circled in green) were also detected. When inspecting the entire corpus on the temporal axis, it was noticed that such errors are sparse. From the result interpretation point of view, feed-back provided by professional under the framework of the SPY project brought to light that when an *I*-Group is maliciously

altered, it is very likely to have other successive blocks altered. Consequently we included in our decision procedure a post-processing rule of the type: one *I*-Group is considered as altered if at least two *I*-Groups that succeed it or precede it are detected as altered. This way, the new values for precision and recall ratios are $Precision1 = 0.97$ and $Recall1 = 0.97$ at $s = 8$ (see Fig 10-b). Of course, this increase in the statistical performances was obtained at the expense of decreasing the time accuracy: content modifications shorter than 9s cannot be detected.

4.3. Conclusion

This section shows that coupling mb_type based signature and *m*-QIM insertion can ensure practical relevant results for semi-fragile watermarking based integrity verification. Experiments conducted under SPY project corpus showed that the advanced method is robust against content preserving attacks and able to locate content changing modification with a *Precision* value of 0.92 and *Recall* value of 0.97. The spatial and temporal accuracies are evaluated at 1/81 frame size and 9s, respectively.

5. CONCLUSION

With the present study, the MPEG-4 AVC syntax elements which can optimally serve the needs of content-based video authentication are first identified by carrying out an information-theory based investigation. Secondly, a semi-fragile video watermarking software is implemented and assessed in terms of robustness (against content preserving attacks) and precision in identifying the content changing attacks.

Future work is scheduled on extending this study for ensuring the authentication of the emerging HEVC (High Efficiency Video Coding) standard.

REFERENCES

1. S. Upadhyay and K. Singh, *Video authetication – an overview,* International Journal of Computer Science & Engineering Survey (IJCSES), vol. **2**, No. 4, pp. 75–96, November 2011.
2. Z. J. Geradts and J. Bijhold, *Forensic video investigation with real time digitized uncompressed video image sequences*, Proc. SPIE, vol. **3576**, p. 154–164, 1999.
3. I. Richardson, *H.264 and MPEG-4 video compression*, The Robert Gordon University, Aberdeen, 2003.
4. M. Hasnaoui, M. belhaj, M. Mitrea and F. Préteux, *MPEG-4 AVC stream watermarking by m-QIM technique*, Proc SPIE. **7881**, pp. OL1-OL8, 2011.
5. J. Titman, A. Steinmetz and R. Steinmetz, *Content based digital signature for motion pictures authentication and content fragile watermarking*, Multimedia Computing and Systems, IEEE International Conference, vol. **2**, pp. 209–213, 1999.
6. P. M. Queue, *Toward robust content based techniques for image authentication*, Multimedia Signal Processing, IEEE Second Workshop, pp. 297–302, 1998.
7. S. Chen and H. Leung, *Chaotic watermarking for video authentication in surveillance applications*, IEEE Trans On Circuits and Systems For Video Technology, vol. **18**, pp. 704–709, 2008.
8. S. Thiemert, H. Sahbi and M. Steinebach, *Applying interest operators in semi-fragile video watermarking,* Proc. SPIE, vol. **5681**, pp. 353–362, 2005.
9. S. Thiemert, H. Sahbi and M. Steinebach, *Using entropy for image and video authentication watermarks*, Proc. SPIE, vol. **6072**, pp. 18–1 – 18–10, 2006.
10. K. AIT. Saadi, A. Bouridane, and A. Guessoum, *Combined fragile wateramrking and digital signature for H.264/ AVC video authentication*, 5th Symposium on Comminucation and Mobile Network, pp. 1–4, 2010.
11. H. P. Moravec, *Towards automatic visual obstacle avoidance*, Morgan Kaufmann Publishers, 5th Int. Joint Conference on Artificial Intelligence, Vol. **2**, 1977.
12. www.webcamendirect.net
13. M. Buckland, F. Gey, *The relationship between recall and precision*, Jornal of the American Society for Information Science, vol. **45**, pp. 12–19, 1994.

Received July 1, 2013