Smart selective encryption of CAVLC for H. 264/AVC video

Article · November 2011

DOI: 10.1109/WIFS.2011.6123130

3 authors:

Loïc Dubois
L'Agence nationale de la recherche
7 PUBLICATIONS **33** CITATIONS
SEE PROFILE

William Puech
Université de Montpellier
264 PUBLICATIONS **1,346** CITATIONS
SEE PROFILE

Jacques Blanc-Talon
Université Paris-Sud 11
99 PUBLICATIONS **450** CITATIONS
SEE PROFILE

Smart Selective Encryption of CAVLC for H.264/AVC Video

Loïc Dubois[1], William Puech[1] and Jacques Blanc-Talon[2]

[1] *LIRMM Laboratory, UMR 5506 CNRS, University of Montpellier II*
161, rue Ada, 34392 MONTPELLIER CEDEX 05, FRANCE
loic.dubois@lirmm.fr, william.puech@lirmm.fr

[2] *DGA, Bagneux*
FRANCE
jacques.blanc-talon@dga.defense.gouv.fr

***Abstract*—In this paper we present a new approach of selective encryption (SE) of video. SE is currently applied to video codec H.264/AVC in order to perform confidentiality and preserve bitrate and data size of the underlying video sequence. Our approach analyzes, each macro-block of a given video sequence, separately before its entropy encoding. In this way, the system decides whether it is necessary to or not to encrypt the current macro-block. On the former case, the selected macro-block is encrypted using the SE-CAVLC in the entropy encoder. This approach allows to encrypt a smaller part of the macro-blocks without disturbing the level of confidentiality.**

I. Introduction

With the rapid growth in processing powers and in network bandwidths, digital videos are commonplace and their number is increasing exponentially. The phenomenal increase in the amount of transmitted and archived data requires its protection while ensuring high transparency. Data security and network security are two common solutions to resolve these issues. Generally, data security is preferred because it allows a better control over the processing time and data size. Further, multimedia data require to be compressed and encrypted in order to reduce the transmission time. In video processing, full encryption of data is rarely used because processing time is twice that a compression. That is why most of the applications use selective encryption (SE) which guarantees a good level of data confidentiality to protect the spatial resolution without any data inflation.

This paper presents a smart SE method of videos. We have developed a new method based on the SE-CAVLC [1]; encryption is performed according to a measure of quality, during the entropy encoding step. This approach is possible, thanks to the prediction error method in H.264/AVC where each MB is predicted from a neighboring MB that has been encoded previously. The encryption decision is taken according to the PSNR as measure of quality. A wide range of results highlights the efficiency with various PSNR thresholds. Moreover, our scheme ensures reduced encryption ratios (ERs) and adaptivity in the encryption.

In section II, H.264/AVC codec and previous work on SE of video are presented. In section III, we present our proposed smart selective encryption of CAVLC (SSE-CAVLC) approach and develop its scheme and implementation. In section IV, experimental results are given which show that the proposed SSE-CAVLC reduces the ER with respect to SE-CAVLC [1]. Finally, in section V, concluding remarks and future perspectives about the proposed scheme are being discussed.

II. State of Art

H.264/AVC (also know as MPEG-4 Part 10) is the video coding standard of ITU-T and ISO/IEC [2]. In H.264/AVC, each frame is divided into MBs of 16x16 pixels. These MBs are encoded separately; the encoding method is composed of an Integer Transform (IT) followed by a quantization of the errors of prediction of a MB (prediction between MBs in *intra* (I frame) or *inter* (P and B frames)) and an entropy encoding using either run length coding (CAVLC) or arithmetic coding (CABAC) as presented in Fig. 1. In I frame, the current MB is predicted spatially from the neighboring MBs that have been previously encoded and reconstructed (MBs at top and left). The purpose of the reconstruction in the encoder is to ensure that both the encoder and decoder use identical reference frame to create the predictions.

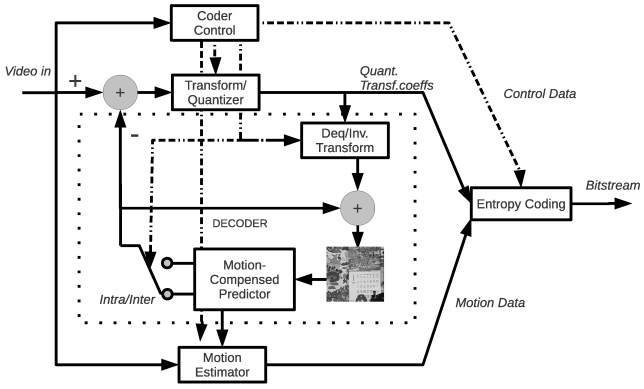

Fig. 1. H.264/AVC encoder scheme.

In the literature, several methods of SE of video have been proposed. SE, also known as partial encryption, is an encryption strategy which aims at saving computation time and enabling new system functionalities. In SE, only a portion of

978-1-4577-1019-3/11/$26.00 ©2011 IEEE

the compressed bitstream is encrypted while still providing adequate data security [3]. Moreover, SE fulfills the main tasks of video encryption, namely visual confidentiality and data protection. These tasks are performed by applying a SE in certain segments of the bitstream with respect to a total encryption which encrypts the whole bitstream. A challenge in SE is that both encrypted and non-encrypted informations should be appropriately identified and displayed [4] for bitstream compliance to standard H.264/AVC video format.

In the field of video, different techniques of SE have been developed which include permutation, the Data Encryption Standard (DES) and the Advanced Encryption Standard (AES) [5]. Based on the location of the encryption stage in the video codec, these SE techniques can be divided into five broad categories. These are named the spatial, the video codec structure based, transform based, entropy encoding based and bitstream. Encryption during the entropy encoding module is often efficient and has been adopted by several authors. The use of Huffman entropy coder, as encryption cipher, has been studied in [6]. Despite providing a compliant video bitstream, the scheme suffers from bitrate increase which makes it less fit limits real-time applications. Another method [7] has been introduced for performing encryption by using various Huffman tables, but it is vulnerable to plain-text attacks [8]. Moreover, a SE of MPEG-4 video standard has been studied in [6] wherein DES was used in order to encrypt fixed length and variable length codes. In this approach, the encrypted bitstream is fully compliant with the MPEG-4 bitstream format but the bitstream size is increased. This particularity is a current observation in SE of video. Furthermore, data security in *intra mode* is improved in [9] where each frame receives a specific and synchronized encryption key. Moreover, each type of MB is encrypted differently with chaotic sequences in order to improve the protection against plain-text or Friedman attacks. Perceptual encryption has also been presented in [10] where encryption is done with an alternative transform of the DCT coefficients.

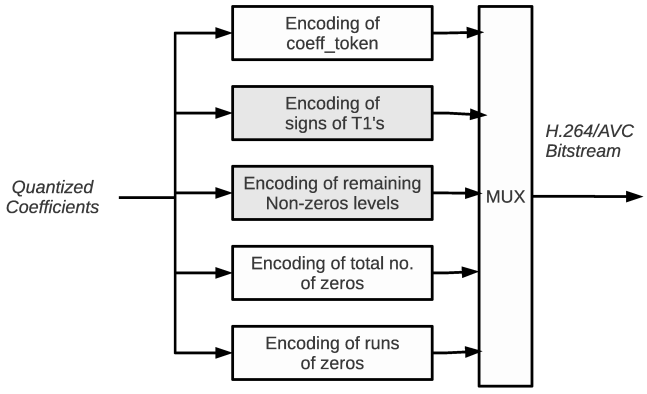

Fig. 2. Syntax elements used in CAVLC entropy encoder of H.264/AVC. In gray are represented the encrypted coefficients in SE-CAVLC [1].

AES has also been used in SE-CAVLC [11], [1] while encrypting only a part of the quantized coefficients in various VLC tables. SE-CAVLC [1] is performed by using the Advanced Encryption Standard (AES) algorithm in the Cipher Feedback (CFB) mode on a subset of codewords/bin-strings. The data information is selectively encrypted for each MB, whereas header information is never encrypted because it is used for the prediction of the next MBs. In the entropy encoder, the SE is performed in multiple VLC tables used in CAVLC. In CAVLC, five syntax elements are used to code levels and runs: *coeff_token, signs_of_trailing ones, remaining_non-zero_levels, total_number_of_zeros and runs_of_zeros* as shown in Fig. 2. Only *signs_of_trailing_ones* and *remaining_non-zeros_levels* are encrypted in order to keep the bitstream compliant. The encrypted space is the VLC codes which keep the same code lengths as a standard compression.

III. Proposed Method

Our method, called the smart selective encryption of CAVLC (SSE-CAVLC), encrypt with SE-CAVLC [1] a limited part of the MBs for each frame. We apply our algorithm only in the *intra* mode of the CAVLC entropy encoder, as illustrated in Fig. 3. This limited part is selected while comparing each MB in terms of confidentiality on the basis of a measure of quality. For this paper, we have chosen to use the Peak Signal Noise Ratio (PSNR) as the measure of quality. Our approach, as SE-CAVLC [1], keeps bitstream compliant for a classic H.264/AVC decoder. The main objective is to limit encryption to a smaller portion of data in order to reduce load over the entropy encoder and support real-time applications. Our method consists of two main steps, namely, the memory feedback, which is discussed in Section III-A, and the quality measure based, which is presented in Section III-B.

A. The memory feedback of encryption

In the H.264/AVC codec, the prediction error is used in order to reduce the bitstream of video sequences. This prediction error is the difference between the current MB and a previous neighbor. A scan of each previously neighboring encoded MB is achieved in order to find the MB yielding the smallest prediction error. Thus, this method considerably reduces the quantity of the transmitted data. During the decoding step, a MB which has been decoded from an encrypted MB should be heavily distorted. We use this specificity in order to spread the SE-CAVLC [1] through each frame of a video sequence.

In order to decide whether a MB is encrypted enough, it has to be decoded and compared to the original compressed MB. On the one hand, the intra-frame prediction system of the H.264/AVC codec has a memory of the current frame which allows to optimize the prediction error. The original compressed MB may be extracted in this part of the H.264/AVC encoder. One the other hand, our system requires to create a similar memory of the previously encoded MBs. Indeed, in order to compare each MB, the prediction error must be decoded and added to the MB that has been used for the prediction, as illustrated in the area B' of Fig. 3. This second memory is mandatory in order to compare all the MBs. As the encryption of each MB is done in the entropy encoder,

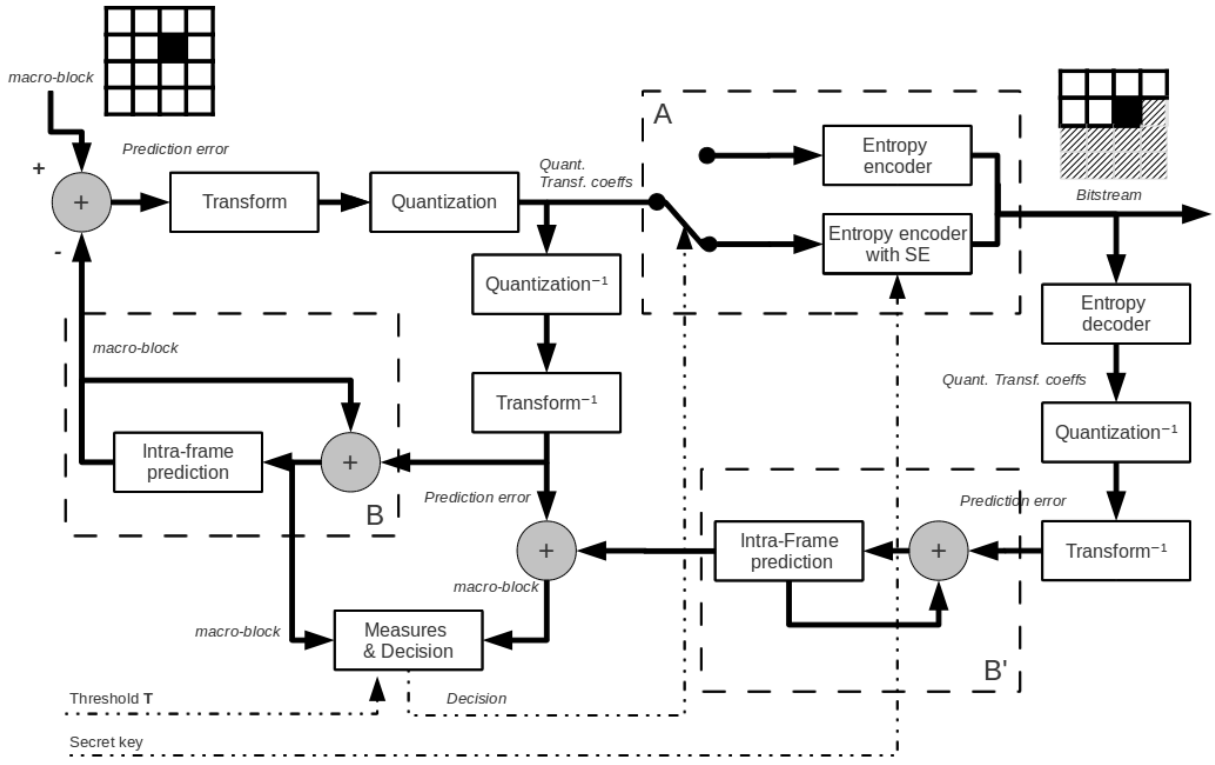

Fig. 3. Block diagram of the proposed SSE-CAVLC. The area A presents the block to encrypt or not a MB as a function of the decision. The area B is the standard intra-prediction system and the area B' is the memory feedback which is used for decision system.

we have to implement an entropy decoder to save each MB, this decoder being similar to a H.264/AVC decoder. Inverse-quantization and inverse-transform is applied to the MB before storing it. Moreover, in spite of the encryption of some MBs, the second intra-prediction is not affected by this encryption, because the header data of the MB is not encrypted. Finally, this approach implies that the first MB is always encrypted and this is the only block which does not use intra-prediction. Thus, the encryption of this MB has the greatest influence throughout the frame.

B. Measure of quality and decision

In Section III-A, we showed that our approach needs a second intra-prediction for the encrypted MBs. Thus, the comparison between the compressed and the encrypted MBs is possible and is presented in this section. In SSE-CAVLC, during the decoding step of the decoder, our measure system compares the current intra-predicted MB to its original compressed MB in the intra-prediction memory. If its measure of quality is below a threshold T, we consider that its confidentiality is sufficient and this MB is not encrypted in the entropy encoder, as illustrated in the area A of Fig. 3. Also, if its measure of quality is over this threshold, the MB is encrypted in the entropy encoder. The decision is sent to the entropy encoder which works in two modes: a H.264/AVC encoder for the non-encrypted MBs and a SE-CAVLC encoder for the encrypted MBs. Whatever the entropy encoder mode is, the bitstream of each MB is decoded and added in the second intra-prediction memory in order to compare the next MBs.

The decision of encryption is done while using the PSNR on the luminance. The PSNR is used as a threshold in the decision system, it can be tuned in order to have different results in terms of confidentiality. The threshold T decides if a MB needs to be encrypted, as presented in Fig. 3, but it cannot ensure if an encrypted MB is under this threshold. Furthermore, the PSNR threshold is preset and each MB is compared to the original compressed MB, if its PSNR is under this threshold the entropy encoder does not encrypt it. We only use the luminance in the encryption decision because it has the highest influence on the human visual system (HSV), but other ways like the color PSNR might be developed in oder to improve the method.

For the decoding step, each MB is decoded and decrypted only if it has been encrypted. In order to know if a MB has been encrypted, we can use various strategies. A simple solution consists to insert an encryption binary map in the header information.

IV. Experimental Results

We have used six benchmark video sequences in QCIF: *bridge-close, city, football, foreman, hall, mobile*. Each video shows different combinations of motion, color, contrast and objects. The results are presented with the most representative samples. In term of encryption, we consider a good confi-

dentiality if PSNR is approximately less than 13 dB. All of the videos have been compressed with a QP at 18, which represents a hight quality compression with a final PSNR of around 45 dB for a non-encrypted video.

For our experiments, we have applied different thresholds of PSNR for each video. These experiments allow to highlight the quantity of the encrypted MBs (EMBs) used for a partial encryption of the MBs. The thresholds have been chosen between 10 and 15 dB. This range represents well the limit of similarity in terms of visual confidentiality. In Fig. 4, we show the effects of SSE-CAVLC on the first frame of *city* video sequence and on three frames of *mobile* video sequence. Visually, the localization of the encrypted MBs, represented in black in Fig. 4, is different according to the videos and the frames as a function of the threshold T. However, the amount of encrypted MBs seems to be linked to the video sequence. Note that the amount decreases clearly with the elevation of the threshold.

Fig. 5 and Fig. 6 present *mobile* and *city* video sequences which are encrypted using SSE-CAVLC on twenty frames and sixty frames respectively. In these figures are given the evolution of the mean PSNR, its standard deviation and the evolution of the amount of encrypted EMBs as a function of the PSNR threshold T used for the decision. As we might predict it, the PSNR and the EMBs evolve in reverse. When the threshold rises, the amount of EMBs decreases and the color PSNR also rises. Furthermore, the standard deviation grows according to the threshold. Its increase is probably due to the variation of the encryption which depends of the secret key. Moreover, when the threshold is higher there is a greater bandwidth of PSNR variation.

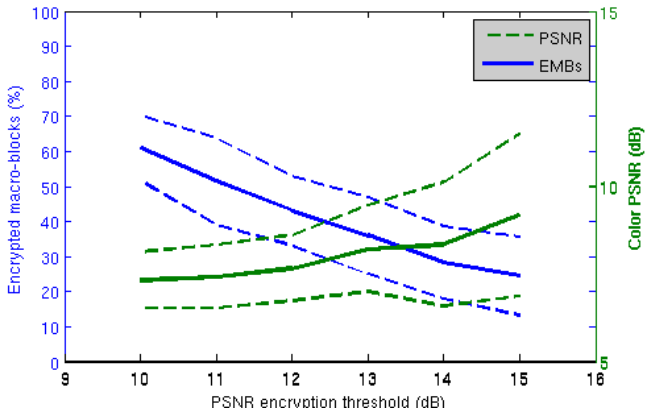

Fig. 5. Variation of the global color PSNR (mean in straight line and standard deviation in dashed line) as a function of the PSNR threshold T on the video sequence *mobile* for 20 frames.

Tab. I shows the results on the six other video sequences in term of mean PSNR, standard deviation and EMBs [1]. As of the two previous examples, the color PSNR and the standard deviation increased with the threshold T. However, mean color

[1]Visual results on these video sequences are available at: http://www.lirmm.fr/~dubois/videos

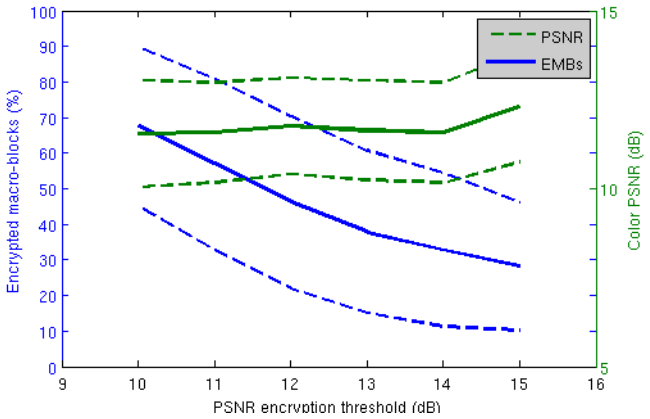

Fig. 6. Variation of the global color PSNR (mean in straight line and standard deviation in dashed line) as a function of the PSNR threshold T on the video sequence *city* for 60 frames.

PSNR is always under 15 dB with an amount of encrypted MBs relatively low. For some video sequences only a few MBs are encrypted which represent less than 5% of the total data size of these videos.

T (dB)	Stats	*bridge-close*	*football*	*foreman*	*hall*
10	Mean (dB)	10.39	10.54	8.47	9.20
	Std (dB)	1.08	2.03	2.11	0.80
	EMBs (%)	69.34	63.38	55.05	48.18
11	Mean (dB)	10.74	10.10	8.61	9.19
	Std (dB)	0.93	1.22	2.05	0.82
	EMBs (%)	49.24	60.61	50.70	36.97
12	Mean (dB)	11.11	10.18	9.00	9.28
	Std (dB)	0.87	1.21	2.29	0.95
	EMBs (%)	38.18	52.27	45.60	26.87
13	Mean (dB)	11.20	10.45	8.75	9.77
	Std (dB)	0.98	1.28	2.33	1.08
	EMBs (%)	39.59	48.73	38.58	20.45
14	Mean (dB)	11.98	10.59	8.95	10.24
	Std (dB)	1.01	1.80	2.48	1.21
	EMBs (%)	34.24	39.54	34.84	12.88
15	Mean (dB)	11.70	11.09	8.55	10.64
	Std (dB)	0.97	1.99	2.20	1.29
	EMBs (%)	30.91	34.79	27.52	9.49

TABLE I
STATISTICAL RESULTS FOR FOUR OTHER VIDEOS IN QCIF ON TWENTY FRAMES WITH QP = 18. T REPRESENTS THE USED THRESHOLD.

Fig. 7 and Fig. 8 show the repartition of the final color PSNR with respect to the amount of the encrypted MBs. This repartition reflects the entire frames and the entire PSNR thresholds of each video. These figures present clearly the variation of the standard deviation when the amount of encrypted MBs is low. However, although the percentage of encrypted MBs decreases, the global color PSNR rises rarely beyond 15 dB which ensures a good confidentiality level. In terms of encrypted data, SE-CAVLC [1] processes about 20% of the total bits. When using our proposed SSE-CAVLC, the amount of encrypted bits is decreased to around 10% of the bitstream size while still achieving a adequate security and a good confidentiality.

Furthermore, even with only I frames, the high color varia-

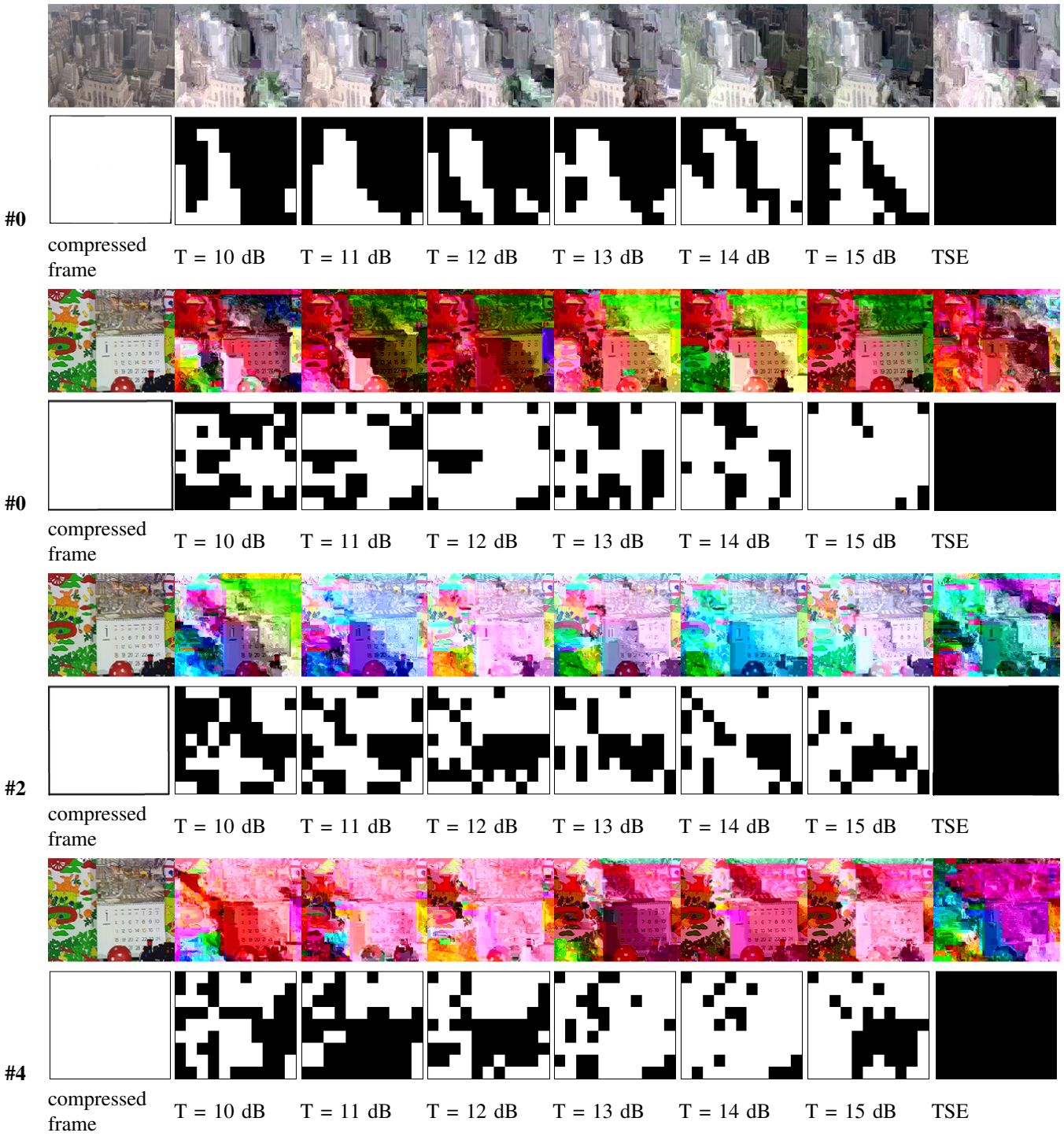

Fig. 4. Results for different thresholds T of PSNR on video sequences *city* and *mobile*. The first frame of *city* and the frames #0,2,4 of *mobile* are presented. On left is presented the original compressed frame with QP = 18. On right is shown the frame with a total selective encryption (TSE) with SE-CAVLC [1]. In center are the encrypted frames with SSE-CAVLC as a function of the threshold T (between 10 dB and 15 dB). Under each frame is presented its encryption map where the black blocks are the encrypted MBs.

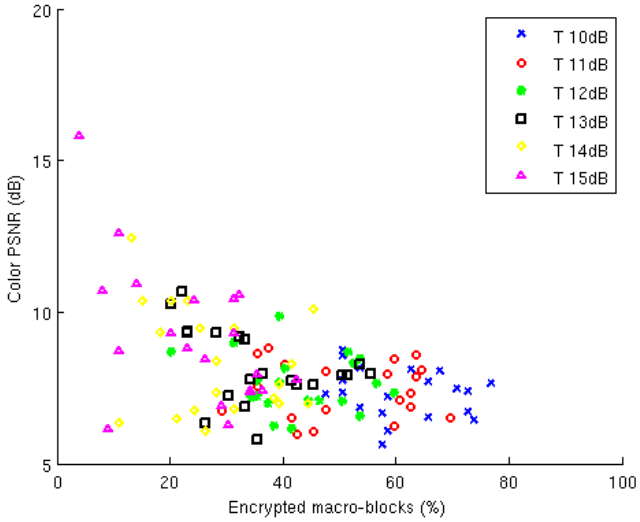

Fig. 7. Variation of the global color PSNR of each frame and each PSNR threshold with respect to their percentage of EMBs for the video sequence *mobile* on 20 frames. Each PSNR threshold is represented with a different color and symbol.

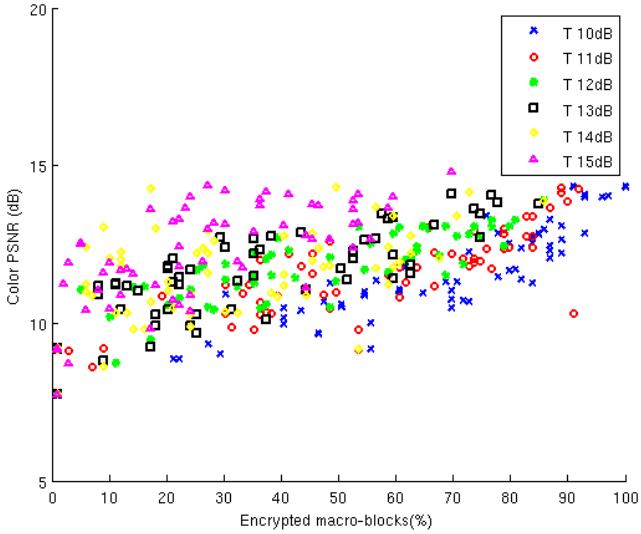

Fig. 8. Variation of the global color PSNR of each frame and each PSNR threshold with respect to their percentage of EMBs for the video sequence *city* on 60 frames. Each PSNR threshold is represented with a different color and symbol.

tions between two neighbor frames, due to the SSE-CAVLC, is a additional advantage in term of visual confidentiality. Indeed the video sequences are hard to watch. The amount of encrypted MBs varies depending on the video being processed, and it showed that SSE-CAVLC is adaptive. However, in some cases, some encrypted frames do not reach an adequate level of confidentiality which is certainly due to the PSNR which is a measure of quality and not a measure of similarity.

V. Conclusion

In this paper, we have presented a new approach for the SE of video. The error prediction system of H.264/AVC allows to encrypt neighbor MBs of a previous encrypted MB indirectly. In SSE-CAVLC, we have added a prediction system of the decoded frame with an implementation of measures which decides whether to encrypt a given MB or not during the entropy encoding. In this way, only the essential MBs are encrypted while keeping a global final PSNR under the predefined threshold. Indeed, this approach allows to encrypt only a small percentage of the bitstream. However, this method needs some additional computations in the H.264/AVC encoder in order to analyze the confidentiality of the current MB. To sum up, SSE-CAVLC is a solution to the key problems which are data protection and adequate confidentiality. Further, SSE-CAVLC optimizes the reduction of the encrypted bits in the H.264/AVC bitstream.

In this work, the threshold used to measure the quality is the PSNR and that too on the basis of the luminance only. This system is not the most efficient in term of similarity. The proposed SSE-CAVLC has some perspectives in these ways. That is why an improvement of the proposed SSE-CAVLC will be to use other measures of quality in color, and also a measure of similarity like the SSIM [12]. This approach should provide a better precision of the encrypted MBs and should be more efficient in terms of the final amount of encrypted bits.

References

[1] Z. Shahid, M. Chaumont, and W. Puech, "Fast Protection of H.264/AVC by Selective Encryption of CAVLC and CABAC for I & P frames," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 5, pp. 565–576, May 2011.

[2] Joint Video Team, "Draft ITU-T Recommendation and Final Draft INternational Standard of Joint VIdeo Specification (ITU-T Rec. H.264 / ISO/IEC 14496-10 AVC)," *Doc. JVT-G050*, vol. Tech. Rep., March 2003.

[3] T. Lookabaugh and D. Sicker, "Selective Encryption for Consumer Applications," *IEEE Communications Magazine*, vol. 42, no. 5, pp. 124–129, May 2004.

[4] H. Chen and X. Li, "Partial Encryption of Compressed Images and Videos," *IEEE Transactions on Signal Processing*, vol. 48, no. 8, pp. 2439–2445, August 2000.

[5] A. Ulh and A. Pommer, *Image and Video Encryption - From digital Rights Management to Secured Personal Communication*. Springer, 2005.

[6] J. Wen, M. Severa, W. Zeng, M. Luttrell, and W. Jin, "A Format-Compliant Configurable Encryption Framework for Acess Control of Video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 6, pp. 545–557, June 2002.

[7] C. Wu and C. Kuo, "Design of Integrated Multimedia Compression and Encryption Systems," *IEEE Transactions on Multimedia*, vol. 7, pp. 828–839, October 2005.

[8] G. Jakimoski and K. Subbalakshmi, "Cryptanalysis of Some Multimedia Encryption Schemes," *IEEE Transactions on Multimedia*, vol. 10, no. 3, pp. 330–338, April 2008.

[9] Y. Liu, Z. Su, G. Zhang, and S. Xing, "An Improved Selective Encryption for H.264 Video based on Intra Prediction Mode Scrambling," *Journal of Multimedia*, vol. 5, pp. 464–472, 2010.

[10] S.-K. Au Yeung, S. Zhu, and B. Zeng, "Perceptual Video Encryption using multiple 8x8 transforms in H.264 and MPEG-4," *IEEE ICASSP*, pp. 2436–2439, May 2011.

[11] Z. Shahid, M. Chaumont, and W. Puech, "Fast protection of H.264/AVC by selective encryption of CABAC for I & P frames," *EUSIPCO*, pp. 2201–2205, 2009.

[12] Z. Wang, A. C. Bovik, and E. P. Simoncelli, "Multi-scale Structural Similarity for Image Quality Assessment," *IEEE Asilomar Conference Signals, Systems and Computers*, pp. 1398–1402, 2003.