Reduced selective encryption of intra and inter frames of H.264/AVC using psychovisual metrics

Loïc Dubois, William Puech, Jacques Blanc-Talon

REDUCED SELECTIVE ENCRYPTION OF INTRA AND INTER FRAMES OF H.264/AVC USING PSYCHOVISUAL METRICS

Loïc Dubois[1,2]*, 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 the field of video protection, selective encryption (SE) is a scheme which ensures a visual security of video by encrypting a small part of data. This paper presents a new SE algorithm for H.264/AVC in CAVLC mode. This algorithm controls the amount of encrypted alternative coefficients (AC) of the integer transform in the entropic encoder. Two visual quality measures, the peak signal-to-noise ratio (PSNR) and the structural similarity (SSIM), are used to measure the visual confidentiality level of each video frame and to regulate the amount of encrypted alternatives coefficients. This method can be applied on *intra* and *inter* frame video sequences.

1. INTRODUCTION

With the rapid evolution of digital media, growth of processing power and availability of network bandwidths, the digital videos are commonplace and their number rises exponentially. Consequently, archived and transmitted data require to be protected because they can be easily copied and modified. The data security for ensuring this problem is generally preferred to the network security thanks to a better optimization of processing time and data-size. Further, video data require to be compressed and encrypted in order to reduce the transmission time. In encryption process, Selective Encryption (SE) algorithms are usually advised, they guarantee data confidentiality and protection without data increase and with saving computation time.

This paper presents an analysis of SE of video combined with similarity measures. In H.264 codec, a SE based on SE-CAVLC [1] is spread in the *inter* frames while only encrypting the *intra* frames. Moreover, we use similarity measures like SSIM and PSNR to analyse the perceptual effect of this SE. Further, we present Reduced Selective Encryption (RSE) which decreases the amount of encrypted coefficients with respect to a good visual protection of each frame.

Section 2 presents the H.264/AVC codec and the main previous work on SE of video. In Section 3, we presented our approach and our analysis in detail. The two main objectives are to highlight the propagation of an encryption effect through predicted frames and the perceptual effect of this phenomenon. Further, we analyse a RSE which reduces the amount of encryped coefficient in the entropic coder with respect to the perceptual confidentiality of the video frames. In Section 4, the experimental results are given and discussed. In Section 5, concluding remarks and future perspectives about the proposed scheme are discussed.

2. STATE OF ART

H.264/AVC, also known as MPEG-4 Part 10, is the video coding standard of ITU-T and ISO/IEC. In H.264/AVC, each frame is divided in Macro-Blocks (MBs) of 16x16 pixels. These macro-blocks are encoded separately; the encoding method is an Entire Transform followed by a quantization of the MB, a prediction between MBs in *intra* (I frame) or *inter* (P and B frames), and an entropy coding using either run length coding (CAVLC) or arithmetic coding (CABAC). In *intra* frame, the current MB is predicted spatially from neighboring MBs which have been previously encoded and reconstructed. In *inter* frame, the current MB is predicted spatially and temporally from previous frames. The purpose of the reconstruction in the encoder is to ensure that both the encoder and the decoder use identical reference frame to create the predictions.

In the literature, SE of video is current and several methods have been proposed [2]. SE, also known as partial encryption, is a encryption strategy which aims at saving computation time or enabling new system functionalities. In SE, a small of the compressed bitstream is encrypted while still providing adequate data security [3] with respect to a total encryption which would encrypted the whole bitstream. Another challenge in SE is that both encrypted and non-encrypted informations should be appropriately identified and displayed in order to remain SE bitstream compliant as a standard H.264/AVC video. 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 [4]. Despite providing a compliant video bitstream, the scheme suffers from bitrate increase which makes it less appropriate and limits real-time applications. Moreover, a SE of MPEG-4 video standard has been studied in [4] 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 [5] 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 attacks. Perceptual encryption has also been presented in [6] where encryption is done with an alternative transform of the DCT coefficients.

Further, AES has also been used in SE-CAVLC [1] by encrypting only a part of the quantized coefficients in various VLC tables. SE-CAVLC [1] is performed by using the AES algorithm in the Cipher Feedback (CFB) mode on a subset of codewords/bin-strings. The data information is selective encrypted for each MB, header information is never encrypted because it is used for the prediction of the next MBs. In the entropy coder, the SE is performed in the the multiple VLC tables used in CAVLC. Only non-zeros coefficients 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.

3. PROPOSED METHOD

3.1. SE propagation analysis

In the H.264/AVC codec, the prediction error is used in order to reduce the bitstream size of video sequences. This prediction error is the difference between the current MB and a previous neighbor MB. A scan of each previously neighboring encoded MB is achieved in order to find the MB yielding the smallest prediction error. Moreover this prediction is used in spatial domain in order to encode *inter* frames. 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 encryption through each *inter* frame of a video sequence, the *intra* frames are encrypted with SE-CAVLC and the *inter* frames not. In order to know the efficient range of this technique of SE, we use a group of similarity measures.

The similarity measures are mathematic tools which allow to compare a recorded image to the original measures. The Peak Signal-to-Noise Ratio (PSNR) is:

$$PSNR = 10\,log_{10}\left(\frac{d^2}{MSE}\right), \qquad (1)$$

where d is the dynamic range, generally 255, MSE is the mean squared error pixel to pixel between the original frame and the recorded one.

Usually, the PSNR is used to evaluate the confidentiality of an encrypted image. But, nowadays, new methods with perceptual contents are more correlated with the Human Visual System (HVS). The Structural SIMilarity (SSIM) [7] is one of the best of these methods:

$$SSIM = \frac{(\,2\,\mu_x\,\mu_y\,+\,c_1\,)\,(\,2\,cov_{xy}\,+\,c_2\,)}{(\,{\mu_x}^2\,+\,{\mu_y}^2\,+\,c_1\,)(\,{\sigma_x}^2\,+\,{\sigma_y}^2\,+\,c2\,)}\,, \qquad (2)$$

where *x* is the original frame, *y* the recorded one, μ the average, σ the variance, *cov* the covariance, and, c_1 and c_2 two constants which stabilize the division.

SSIM uses the covariance coupled with the average and the variance while PSNR only uses the Mean Square Error (MSE).

We propose to use another measure in order to analyse the blinking between two neighbor, we call this measure the Blink Measure (BM):

$$BM = \frac{\left|\sum_{i=1}^{n-1}\sum_{k=1}^{m}(x_{i,k}-x_{i+1,k})^2 - \sum_{i=1}^{n-1}\sum_{k=1}^{m}(y_{i,k}-y_{i+1,k})^2\right|}{(n-1)m}, \qquad (3)$$

where *x* is the original frame, *y* the recorded one, *n* the sequence length, *m* the frame resolution. BM is the difference of the mean squared errors of the two successive original frames and the same successive frames of the encrypted video sequence. With the BM, we can analyze the blink feeling when one is watching the encrypted video sequence.

3.2. Decreasing of the amount of encrypted coefficients

In SE of H.264, the encryption of non-zeros coefficients is generally sufficient to protected the video. One main way to improve this method is to continue to decrease this amount of coefficients. In the proposed method in this paper, we use the SE-CAVLC algorithm while encrypting just a part of the non-zeros coefficients, and we analyze if the visual confidentiality keep efficient. Non-zeros AC levels of lower frequencies are encrypted in prior according to the selected number of non-encrypted coefficients. An overview of this method is presented Fig. 1. Moreover, during the encoding, we include a similarity measure (SSIM) between each Group Of Pictures (GOP), this is a trigger of selection, its scheme is presented Fig. 2. The SSIM measures the quality of each GOP frame and if one of them is upper from a threshold the next GOP will be encrypted with more encrypted coefficient ($C_{i+1} = C_i + 1$), or, if it is under another threshold we decrease the number of encrypted coefficients ($C_{i+1} = C_i - 1$). In terms of cryptanalysis, our scheme lightens the number of

coefficients, so the number of encrypted bits by the AES algorithm, and this may weaken the security against plaintext attacks.

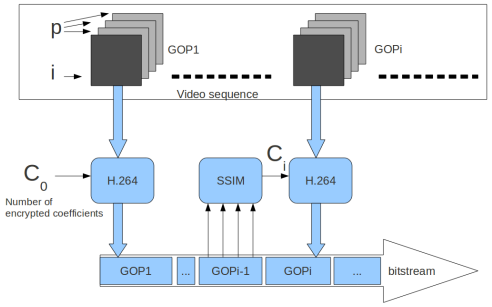

Fig. 1. Overview of the proposed Reduced Selective Encryption method.

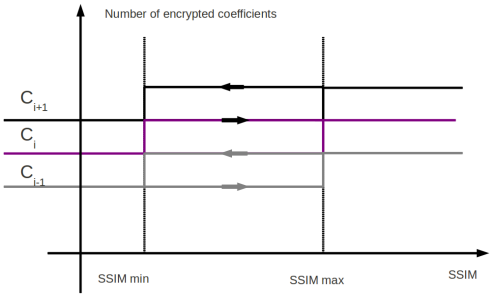

Fig. 2. Trigger of selection for the amount of encrypted coefficients. C_i is the number of encrypted coefficients for the GOP number i.

4. EXPERIMENTAL RESULTS

We have used four video sequences in QCIF. We have compressed 120 video frames for each video. The results are presented with the most representative samples. The results for the similarity measures are based on the luminance. In term of encryption, we consider a good confidentiality if the PSNR is approximately less than 13 dB and less than 0.6 for the SSIM. All of the videos have been compressed with a QP of 32, which represents a moderate compression with a final PSNR around 35 dB for a non-encrypted video sequence. In Section 4.1, we present an analysis of the impact to only SE *intra* frames. In Section 4.2, the results of the proposed RSE are discussed.

4.1. Analysis of the encryption propagation through *inter* frames.

In this section we analyze the efficient range of SE through the non-encrypted *inter* frames. Fig. 3 and 4 show that if the GOP is too long, the visual protection decreases, and becomes ineffective. As we can notice, the maximum size of GOP is between four and ten frames, after this threshold, the confidentiality is affected. Moreover, the BM has also to be analyzed, indeed if the blink effect decreases with time, this improves the content reading of the video. The Fig. 5 shows this evolution. We assume that the BM has to keep upper than 1000 and according to the results, a GOP of ten frames is the allowed maximum.

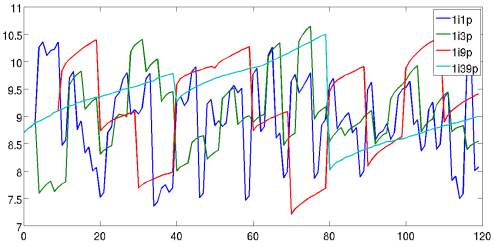

Fig. 3. Evolution of the PSNR of the encrypted video of *mobile* as a function of the GOP length. Blue: GOP with 1 intra frame and 1 inter frame. Cyan: GOP with 1 intra frame and 39 inter frames.

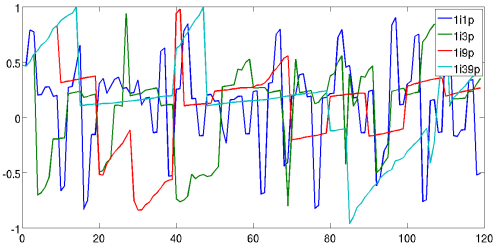

Fig. 4. Evolution of the SSIM of the encrypted video of *mobile* as a function of the GOP length. Blue: GOP with 1 intra frame and 1 inter frame. Cyan: GOP with 1 intra frame and 39 inter frames.

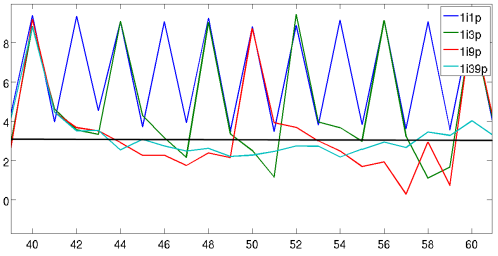

Fig. 5. Logarithmic evolution of the BM of the encrypted video of *mobile* as a function of the GOP length. Blue: GOP with 1 intra frame and 1 inter frame. Cyan: GOP with 1 intra frame and 39 inter frames. In black line represented the threshold at 1000 of BM.

In Fig. 3 and Fig. 4, we can notice that PSNR is always under 13 dB but SSIM reach near 1 which points out that the visual confidentiality is not preserved in these cases. That is why we will use SSIM rather than PSNR in our RSE scheme presented in Section 4.

4.2. Reduced Selective Encryption on the non-zeros coefficients

Tab. 1 presents the results of RSE with a decrease of the amount of encrypted coefficients. We conclude that a lighter SE can be applied to a video while keeping an adequate visual confidentiality. In Fig. 7, only 5.71% of the bistream is encrypted and SSIM is under 0.6. Moreover, we can underline, that the number of encrypted coefficients can be decreased because the main part of the SSIM is under 0.4, that is why we propose a scheme which uses a similarity measure to improve the encryption. Fig. 6 shows an application of our scheme presented in Section 3.2.

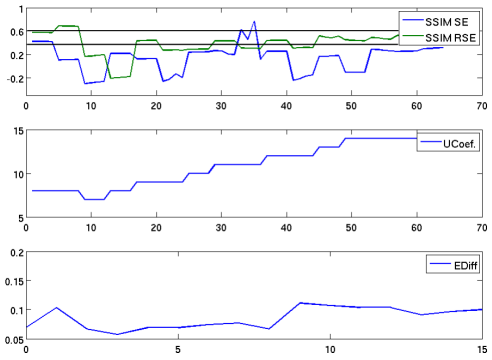

Fig. 6. SSIM measures of the *foreman* video sequence, with a GOP of 4 frames, when all the non-zeros coefficients of the Intra frames are encrypted (*blue*), compared with our proposed Reduced Selective Encryption algorithm (*green*). In middle are presented the amount of unencrypted coefficients with the RSE with respects to the frame number. On the bottom are displayed the differences of encrypted bits between the two encryption approaches, in percentage with respects to the GOP number.

5. CONCLUSION

In this paper, we have presented that similarity measures are an excellent way to improve SE by decreasing the number of encrypted bits. We have also introduced the blink phenomena created by the encryption of video sequences. Next, in the video field of the SE, encryption can be spread through the predicted frames. This allows to only encrypt *intra* frames and protected next predicted frames without encryption of these *inter* frames. Moreover, we developed a new method of RSE which decreases the encrypted non-zeros coefficients used in SE-CAVLC while keeping a good confidentiality. Further, we have implemented a measure of similarity that controls the amount of encrypted coefficients of the *intra* frames.

Foreman	0	4	8	12	16
PSNR (dB)	9.35	8.76	11.55	12.15	14.70
SSIM	0.18	0.33	0.56	0.64	0.75
Blink. M.	9752	6591	5512	4371	3340
Encryption	12.1%	8.6%	5.71%	3.64%	2.35%
Football	0	4	8	12	16
PSNR (dB)	12.6	14.0	16.0	17.8	19.6
SSIM	0.11	0.23	0.36	0.45	0.53
Blink. M.	5611	4270	3849	3030	2321
Encryption	13.7%	11.1%	8.9%	7.17%	5.78%
Mobile	0	4	8	12	16
PSNR (dB)	8.81	8.91	9.23	9.34	9.78
SSIM	0.08	0.09	0.14	0.17	0.23
Blink. M	10946	10494	9048	9113	8100
Encryption	17.6%	16.4%	15.2%	14.0%	12.9%

Table 1. Average of each similarity measure for three video sequences as a function of the number of clear non-zeros coefficients by macro-block. *Encryption* is the percentage of encrypted bits in the bitstream.

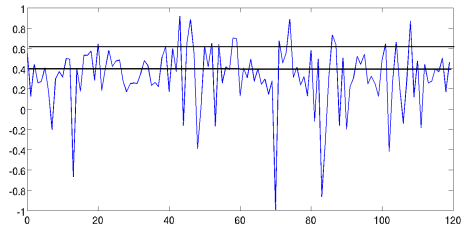

Fig. 7. Evolution of the SSIM of the encrypted video of *foreman* with respect to the frame number. In this experiment 8 non-zeros coefficients are unencrypted, the black lines stand for the SSIM thresholds set to 0.4 and 0.6.

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

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

[4] 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.

[5] 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.

[6] Siu-Kei Au Yeung, Shuyuan Zhu, and Bing Zeng, "Perceptual Video Encryption using multiple 8x8 transforms in H.264 and MPEG-4," *IEEE ICASSP*, pp. 2436–2439, May 2011.

[7] Z. Wang, A. C. Bovik, R. Hamid, Sheik, and E. P. Simoncelli, "Image Quality Assessment: From Evisibility to Structural Similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.