Selective And Scalable Encryption Of Enhancement Layers For Dyadic Scalable H.264/Avc By Scrambling Of Scan Patterns

Zafar Shahid, Marc Chaumont, William Puech

SELECTIVE AND SCALABLE ENCRYPTION OF ENHANCEMENT LAYERS FOR DYADIC SCALABLE H.264/AVC BY SCRAMBLING OF SCAN PATTERNS

Zafar Shahid, Marc Chaumont and William Puech

LIRMM,UMR CNRS 5506, University of Montpellier II,
161, rue Ada, 34392 Montpellier CEDEX 05, France
zafar.shahid@lirmm.fr, marc.chaumont@lirmm.fr, william.puech@lirmm.fr

ABSTRACT

This paper presents a new selective and scalable encryption (SSE) method for *intra* dyadic scalable coding framework based on wavelet/subband (DWTSB) for H.264/AVC. It has been achieved through the scrambling of quantized transform coefficients (QTCs) in all the subbands of DWTSB. To make the encryption scalable, it takes advantage of the prior knowledge of the frequencies which are dominant in different high frequency (HF) subbands, as traditional zigzag scan is not that efficient for them. Thus, by scrambling the scan order of QTCs in the *intra* scalable coding framework of H.264/AVC, we were able to get encryption and compression for enhancement layers (ELs) simultaneously. Watermarking has been integrated in the proposed architecture to avoid the requirement of separate *keys* for each spatial layer. The algorithm is better suited for multimedia streaming as the bitrate of the encrypted bitstream is lower than the original bitrate. Besides offering SSE, the proposed algorithm, when applied to different benchmark video sequences, outperformed the standard zigzag scan in terms of bitrate.

Index Terms— Scan patterns, selective and scalable video encryption, H.264/AVC

1. INTRODUCTION

Scalable video coding (SVC) is used to fulfill the demand of networks that are heterogeneous in terms of bandwidth and client resources. Different customers demand different quality levels of the same copyrighted multimedia content. In order to protect each quality layer independently, separate encryption of each spatial layer is suggested. In our approach, to avoid the use of multiple *keys* for higher layer decoding, watermarking has been incorporated in the framework.

Scalable architecture for H.264/AVC namely joint scalable video model (JSVM) [1] is based on pyramid coding architecture as shown in Fig. 1.a. In this kind of architecture, total spatial resolution of the processed video is the sum of resolutions of all the spatial layers and quality of each of the subsequent layer is dependent on the quality of the base layer. Hsiang [2] has presented a scalable dyadic intra coding method based on subband/wavelet coding (DWTSB). In this method, LL subband is encoded as the base layer while HF subbands are encoded as EL as shown in Fig. 1.b. With this method, if the LL residual is encoded, then higher layer can be encoded at a better quality than the base layer, as illustrated in Fig. 1.c. The results presented by Hsiang have proved to be better than those for H.264 scalable video coding method JSVM [3] for *intra* frame. In dyadic scalable intra coding, the image is transformed to wavelet subbands and then the subbands are encoded by traditional H.264/AVC. Since each wavelet subband possesses a certain range of frequencies, a single scan pattern is not equally efficient for scanning the QTCs in all the subbands. Jia *et al.* [4] have presented the idea of adaptive scan based on spatial prediction for intra frame of traditional H.264/AVC. In a previous work [5], we have presented an adaptive scan for each of the HF subbands in DWTSB scalable architecture as zigzag scan is not suitable for HF subbands. Scan patterns in the spatial domain have been discussed in literature by Maniccam and Bourbakis [6, 7] for compression, watermarking and encryption of multimedia content. Despite the fact that scrambling is a weak encryption technique and is vulnerable to chosen and known plain text attacks [8], it is interesting for copyright protection of ELs, owing to low computational cost, for having scalable access levels.

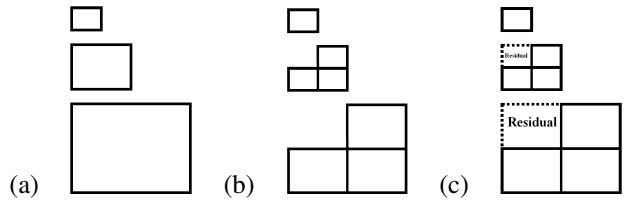


Fig. 1. Different SVC approaches: a) Pyramid coding used in JSVM, b) Wavelet subband coding used in JPEG2000, c) DWTSB for dyadic scalable intra frame of JSVM.

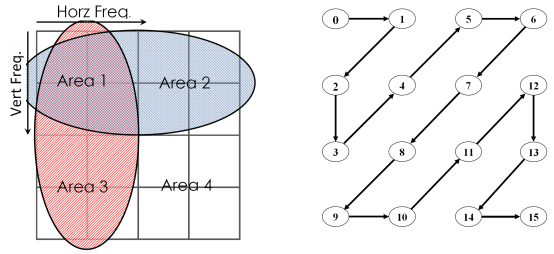
In Section 2, overview of scan methodology, including different types of scans for each subband, is explained. In Section 3, we present the proposed algorithm. Section 4 contains its performance analysis and experimental results, and concluding remarks are discussed in Section 5.

2. SCAN METHODOLOGY

In SVC, the video bitstream contains a base layer (BL) and number of ELs. ELs are added to the BL to further enhance the quality of coded video. In spatially scalable bitstream, the resolution of the EL is either equal to or greater than the lower layer.

Let the QTCs be a 2-dimensional array given as: $P_{m\times n} = \{p(i,j) : 1 \leq i \leq m, i \leq j \leq n\}$. After scanning the 2-dimensional array, we get a set: $Q_{mn} = \{1, ..., mn\}$. One can note that scanning is a bijective function from $P_{m\times n}$ to Q_{mn}. The output of scanning is fed to entropy coding module for source coding. The entropy coding module is designed to perform better when it gets most of the non-zero QTCs in the beginning of the block followed by long trail of zeros at the end of the block, with magnitude of non-zero QTCs higher at the start of the scanned array. Output of a good scanning process should fulfill all the above mentioned conditions, as much as possible.

To analyze the energy distribution in the transform domain,



(a) (b)

Fig. 2. Analysis of LL subband: a) Energy distribution in QTCs of LL subband, b) Scan used for this frequency distribution.

we can divide the 4×4 transform block in four areas. The LL subband consists of slow varying areas and contains lower frequencies, both horizontally and vertically; its energy distribution in the transform domain is shown in Fig. 2.a. Area 1 contains most of the energy as it contains the frequencies both in horizontal and vertical directions. Area 2 and 3 contain frequencies of only 1 direction, either horizontal or vertical. Area 4 contains the least amount of energy. The most appropriate scan for energy distribution of the LL subband is from top-left to bottom-right corner as shown in Fig. 2.b. This is the case for slowly changing video data wherein most of the energy exists in top-left corner of 4×4 transform block. In the case of HL subband, most of the energy is concentrated in top-right corner and QTCs should be scanned from top-right to bottom-left corner. Similarly, scan should be from bottom-left to top-right for the LH subband and the HH subband should be scanned from bottom-right to top-left corner [5]. We have used these scans to make scalable SE which is a good compromise between the bitrate and the security level of the encrypted bitstream.

SVC offers a new profile for *intra* video coding named Profile B Intra only. It can be used for fast random access to individual video frames and is useful for applications like digital cinema and satellite imaging requires. DWTSB can be used to enhance the efficiency of this profile and can be integrated to JSVM reference software without much modification. This framework is quite flexible in selecting the wavelet coefficients for generating low resolution video at the base layer.

3. PROPOSED METHOD

For our method, we have used wavelet critical sampling setting. Daubechies 9/7 wavelet filter set has been used to transform the image to the wavelet subbands. The work has been done on 'JVT-W097' [9] which is referenced H.264 JSVM 8.9 with wavelet framework integrated. The reference software is then modified to analyze the performances of the scrambling of the scan order for each subband.

H.264/AVC offers two types of entropy coding methods namely, context-based variable length coding (CAVLC) and context-based binary arithmetic coding (CABAC). CABAC updates its context model with encoding of every bit. If the synchronization is lost between encoder and decoder, the symbols are decoded differently. Owing to this fact, this SSE implementation for CABAC makes the bitstream undecodable by standard H.264/AVC decoder. Hence, we are using CAVLC entropy coding module with H.264/AVC in our work. In Section 3.1, we present the proposed algorithm for copyright protection of ELs while the scalability aspects are discussed in Section 3.2.

3.1. Selective encryption method

In DWTSB, a given image is transformed to wavelet subbands and the LL subband is encoded as base layer by traditional H.264/AVC. In the EL, LL subband is predicted from the reconstructed base layer. Each HF subband is encoded as separate slice. They are transformed, quantized, scanned and then entropy coded. Our main contribution is to scramble the scan order of QTCs in ELs to encrypt them for copyright protection as shown in Fig. 3.

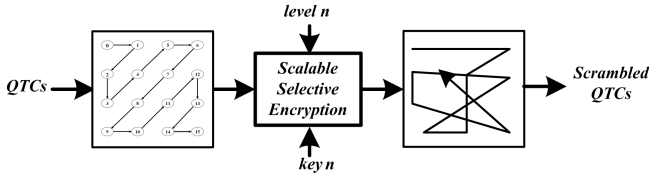
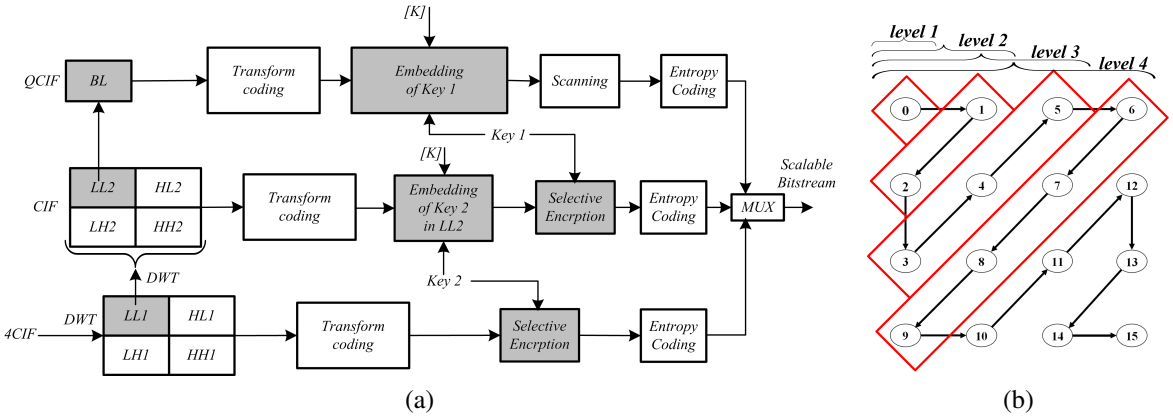


Fig. 3. Encryption of QTCs by scrambling of scan order.

We initialize a pseudo-random number generator (PRNG) with a secret *key*. BL is encoded without encryption. A secret *key* for the 1st EL is watermarked in base layer by embedding the bits of the *key* in the least significant bits (LSBs) of QTCs. In the 1st EL, prediction residual is transformed and quantized. In contrast to the HF subbands, the LL subband, which contains only prediction residual from lower layer, does not contain any frequency in dominance. This makes it the best candidate for embedding the *key* for the immediate higher



(a) (b)

Fig. 4. a) Block diagram of encryption process of ELs in scalable bitstream, b) SSE in which selected QTCs are scrambled in each encryption level.

EL. Embedding the *key* of immediate EL makes it possible to use only one *key* for decryption of certain quality level, thus avoiding the need of multiple *keys* for the decryption of higher quality layers. After watermarking of the LL subband, QTCs, in all the four subbands, are encrypted by scrambling. For a scalable bitstream containing n layers, LL subband in nth layer is not watermarked and all the subbands are only encrypted. The block diagram of the proposed method is presented in Fig. 4.a.

From the watermarking point of view, this scheme is quite flexible and we can use any suitable watermarking method which is best suited for the situation.

3.2. Scalable architecture

To make the SE scheme scalable, we take advantage of a prior knowledge of dominant frequencies in each subband. We have presented five levels of SE. The *level 0* scrambles all the 16 QTCs. It offers the highest security as scrambling space for this level is 16!. The bitrate in this level is approximately equal to that of standard zigzag scan. After that, we leave some QTCs unscrambled, in a diagonal fashion as illustrated in Fig. 4.b. For *level 1*, we leave the first line unscrambled which reduces the scrambling space to 15!. It assures that the bitrate is always lesser than the original bitstream. *level 2* reduces the scrambling space to 13! by leaving first two lines unscrambled and further reduces the bitrate. For *level n* encryption, we leave first n lines unscrambled. So we reduce the bitrate and the encryption strength for the sake of lesser computations.

It is pertinent to mention that for HF subbands, we leave the coefficients unscrambled according the frequency dominance [5] in that subband. For example, in HH subband, we do not scramble QTC at position 15 for *level 1* encryption.

4. EXPERIMENTAL RESULTS

For the experimental results, five benchmark video sequences, containing base layer of QCIF resolution and two ELs of CIF and 4CIF resolutions, have been used. Each of them represents different combinations of motion (fast/slow, pan/zoom/rotation), color (bright/dull), contrast (high/low) and objects (vehicle, buildings, people).

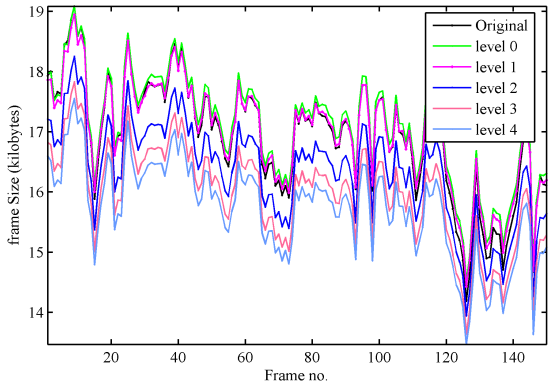
To demonstrate the efficiency of our proposed scheme,

(a) (b)

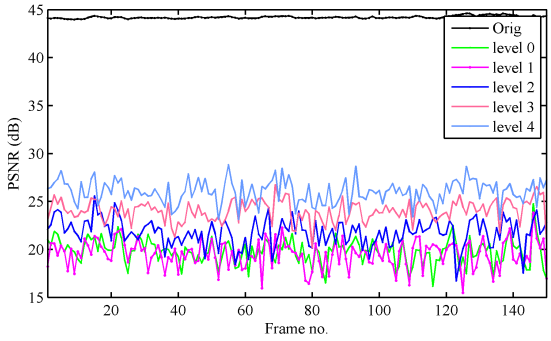
Fig. 5. Subframe of 280 × 240 pixels with offset of (400,200) in original frame of 2nd EL (4CIF) from 1st frame of *city* at QP 12: a) Without encryption, b) With SSE.

we have compressed 150 frames of each sequence as *intra*. Fig. 5.a and 5.b show a comparison of subframe of 1st frame of 2nd EL (4CIF) of *city* video sequence without encryption and with SSE. All refined detail of video frame including texture and edges, which constitute the EL, has been distorted in the encrypted frame. Frame-wise analysis of *city* is presented in the form of graph in Fig. 6.a and 6.b at QP value '18'. Performance analysis over the whole range of QP values of *city* is given in Table 1 for PSNR and bitrate comparison for all the encryption *levels*. Table 2 compares the PSNR and the bitrate of all benchmark video sequences at QP value '12' without encryption and with SSE at different encryption *levels*. Our results verifies that this scheme works well for various video contents and over the whole range of QP values.

For all the experimental simulations, 1st EL of CIF resolution has been decrypted properly and it is only the 2nd EL of 4CIF resolution which is left encrypted. If 1st EL is not decrypted, PSNR of encrypted 2nd EL will be very poor (about 10 dB on average).



(a)



(b)

Fig. 6. Frame-wise analysis of 2nd EL at 4CIF resolution with QP value '18' of *city*: a) Frame size, b) PSNR.

Table 1. Analysis of trade off among bitrate, PSNR and SE *level* for *city* over whole range of QP values.

SE level	12.0 (kbps) (dB)	18.0 (kbps) (dB)	24.0 (kbps) (dB)	30.0 (kbps) (dB)	36.0 (kbps) (dB)	42.0 (kbps) (dB)
Orig	5337 48.71	3386 44.20	1999 39.60	1088 35.19	543 31.17	250 27.60
level 0	5327 19.20	3418 19.75	2045 20.61	1144 20.12	595 18.79	288 18.20
level 1	5285 18.48	3394 19.47	2038 20.24	1147 19.88	598 18.65	286 18.12
level 2	5119 20.79	3270 21.75	1955 22.16	1100 21.04	576 19.28	281 18.87
level 3	5012 22.91	3193 23.99	1910 23.92	1082 22.22	573 20.21	283 19.43
level 4	4942 24.82	3146 25.99	1882 25.73	1071 23.40	571 20.82	283 20.06

5. CONCLUSION

In this paper, a novel framework for encryption of ELs of *intra* dyadic scalable framework for H.264/AVC has been presented. Real-time constraints have been handled successfully by making the encryption scalable and by reducing the bitrate.

Table 2. Analysis of change in bitrate and PSNR without encryption and with SSE of benchmark video sequences at QP value '12'.

Seq.	Orig (kbps) (dB)	level 0 (kbps) (dB)	level 1 (kbps) (dB)	level 2 (kbps) (dB)	level 3 (kbps) (dB)	level 4 (kbps) (dB)
city	5337 48.71	5327 19.20	5285 18.48	5119 20.79	5012 22.91	4942 24.82
crew	4124 49.08	4100 23.16	4066 22.54	3962 25.12	3883 28.04	3831 28.63
harbour	5260 48.71	5291 19.00	5263 18.57	5080 21.23	4966 23.39	4892 24.67
ice	2729 49.51	2790 24.48	2772 24.34	2715 26.46	2674 29.13	2649 30.23
soccer	2348 49.38	2404 21.66	2350 21.19	2282 23.12	2227 26.09	2198 29.04

The experiments have shown that we can achieve the desired level of encryption without any escalation in bitrate, if we do not scramble only the first coefficient in the scan pattern. Owing to embedding of *key* in LL subband of subsequent lower layer, a single *key* is required for the decoding of some specific quality level. In future, the proposed scheme can be extended for protection of P and B frames.

6. REFERENCES

[1] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17(9), pp. 1103–1120, September 2007.

[2] S.T. Hsiang, "Intra-Frame Dyadic Spatial Scalable Coding Based on a Subband/Wavelet Framework for MPEG-4 AVC/H.264 Scalable Video Coding," in *ICIP07, San Antonio, Texas, USA*, 2007, pp. I: 73–76.

[3] T. Wiegand, G. Sullivan, J. Richel, H. Schwartz, and M. Wien, "Joint Scalable Video Model (JSVM) 10," in *JVT-W202*, April 2007.

[4] J. Jia, E.K. Jung, and H.K. Kim, "Adaptive Transform Coefficient Scan for H.264 Intra Coding," *IEICE Trans Inf Syst*, vol. 90, no. 10, pp. 1709–1711, 2007.

[5] Z. Shahid, M. Chaumont, and W. Puech, "A New Subband/Wavelet Framework for AVC/H.264 Intra-Frame Coding and Performance Comparison with Motion-JPEG2000," in *Proceedings of SPIE Volume 7259 - Electronic imaging.*, 2009.

[6] S. S. Maniccam and N. G. Bourbakis, "Image and Video Encryption Using SCAN Patterns," *Pattern Recognition*, vol. 37, no. 4, pp. 725 – 737, 2004.

[7] S. S. Maniccam and N. Bourbakis, "Lossless Compression and Information Hiding in Images," *Pattern Recognition*, vol. 37, no. 3, pp. 475 – 486, 2004.

[8] S. Li, C. Li, G. Chen, N. G. Bourbakis, and K. Lo, "A General Quantitative Cryptanalysis of Permutation-Only Multimedia Ciphers Against Plaintext Attacks," *Signal Processing: Image Communication*, vol. 23, no. 3, pp. 212 – 223, 2008.

[9] S.T. Hsiang, "CE3: Intra-Frame Dyadic Spatial Scalable Coding Based on a Subband/Wavelet Filter Banks Framework," in *Joint Video Team, Doc. JVT-W097*. April 2007, Joint Video Team, Doc. JVT-W097.