

An Adaptive Scan of High Frequency Subbands for Dyadic Intra Frame in MPEG4-AVC/H.264 Scalable Video Coding

Z. Shahid, M. Chaumont and W. Puech

LIRMM Laboratory, UMR 5506 CNRS, University of Montpellier II
161, rue Ada, 34392 MONTPELLIER CEDEX 05, FRANCE

ABSTRACT

This paper develops a new adaptive scanning methodology for intra frame scalable coding framework based on a subband/wavelet(DWTSB) coding approach for MPEG-4 AVC/H.264 scalable video coding (SVC). It attempts to take advantage of the prior knowledge of the frequencies which are present in different higher frequency subbands. We propose dyadic intra frame coding method with adaptive scan (DWTSB-AS) for each subband as traditional zigzag scan is not suitable for high frequency subbands. Thus, by just modification of the scan order of the intra frame scalable coding framework of H.264, we can get better compression. The proposed algorithm has been theoretically justified and is thoroughly evaluated against the current SVC test model JSVM and DWTSB through extensive coding experiments for scalable coding of intra frame. The simulation results show the proposed scanning algorithm consistently outperforms JSVM and DWTSB in PSNR performance. This results in extra compression for intra frames, along with spatial scalability. Thus Image and video coding applications, traditionally serviced by separate coders, can be efficiently provided by an integrated coding system.

Keywords: scalable video coding, H.264/MPEG-4 AVC, adaptive scan

1. INTRODUCTION

Scalable video coding (SVC) standard¹ is based on pyramid coding architecture. In this kind of architecture, the total spatial resolution of the video processed is the sum of all the spatial layers. Consequently, quality of subsequent layers is dependent on quality of base layer as shown in Fig. 1.a. Thus, the process applied to the base layer must be the best possible in order to improve the quality.

Hsiang^{2,3} has presented a scalable dyadic intra frame coding method based on subband/wavelet coding (DWTSB). In this method, LL subband is encoded as the base layer while the high frequency subbands are encoded as subsequent layers 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 base layer, as illustrated in Fig. 1.c. The results presented by Hsiang has proved to be better than H.264 scalable video coding method JSVM⁴ for intra frame. In dyadic scalable intra frame coding, the image is transformed to wavelet subbands and then the subbands are encoded by traditional H.264/MPEG-4 AVC. Since each wavelet subband possesses a certain range of frequencies, zigzag scan is not equally efficient for scanning the transform coefficients in all the subbands. Jia *et al.*⁵ have presented the idea of adaptive scan based on spatial prediction for intra frame of traditional H.264/MPEG-4 AVC.

We organize our work as follows. In Section 2, overview of SVC including different types of scalability which it offers is presented. In Section 3, we introduce the basic scan methodology and the proposed scans for high frequency subbands. Section 4 contains its performance analysis and experimental results including its compression comparison for high frequency subbands and as a whole also with SVC test model JSVM and DWTSB for benchmark video sequences. The concluding remarks about the proposed algorithm are presented in Section 5.

zafar.shahid@lirimm.fr, marc.chaumont@lirimm.fr, william.puech@lirimm.fr

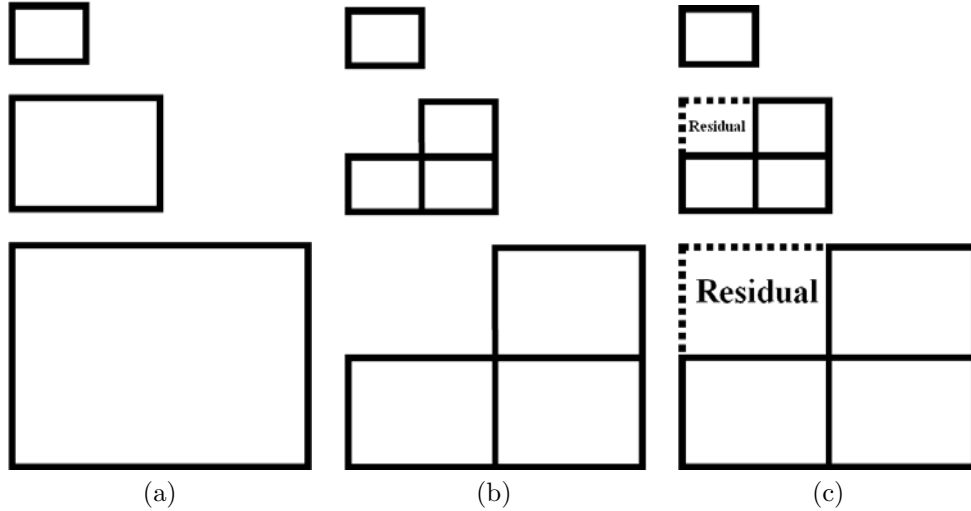


Figure 1. Different scalable video coding approaches: a) Pyramid coding used in JSVM, b) Wavelet subband coding used in JPEG2000, c) Wavelet subband coding for dyadic scalable intra frame of JSVM.

2. SCALABLE VIDEO CODING

In SVC, the video bit stream contains a base layer and number of enhancement layers. Enhancement layer are added to the base layer to further enhance the quality of coded video. The improvement can be made by increasing the spatial resolution, video frame-rate or video quality, corresponding to spatial, temporal and quality/SNR scalability. Previous video standards such as MPEG-2,⁶ MPEG-4⁷ and H.263+⁸ also contain the scalable profiles but they were not much appreciated because the quality and scalability came at the cost of coding efficiency. SVC based on H.264/AVC has achieved significant improvements both in terms of coding efficiency and scalability as compared to scalable extensions of prior video coding standards. Similar to the previous scalable video coding standards, SVC is also built upon a predictive and layered approach to scalable video coding.

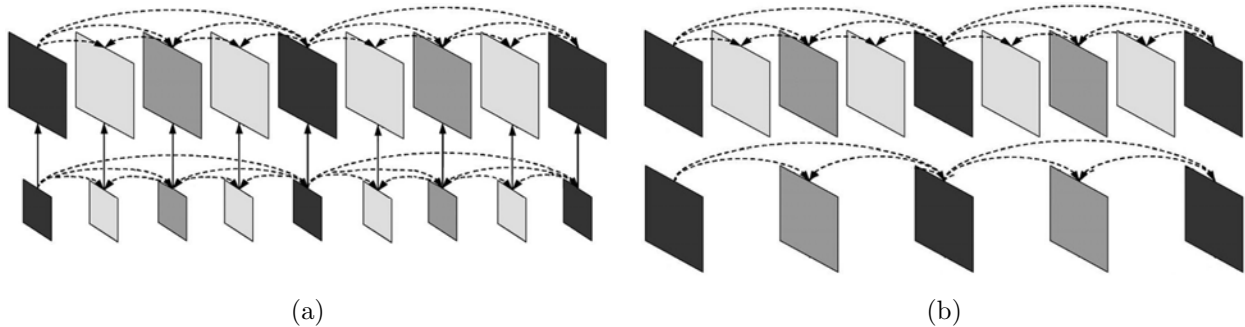


Figure 2. Spatial and temporal scalability offered by SVC: a) Spatial scalability in which resolution of enhancement layer can be either equal to or greater than resolution of base layer, b) First layer containing only I and P frames while second layer contains B frames also. Frame rate of second layer is twice the frame rate of first layer.

In spatial scalability, the inter-layer prediction of the enhancement-layer is utilized to remove redundancy across video layers as shown in Fig. 2.a. The resolution of the enhancement layer is either equal or greater than the lower layer. Enhancement layer P images can be predicted either from lower layer or from the previous frame in the same layer. In temporal scalability, the frame rate of enhancement layer is better as compared to the lower layer. This is implemented using I, P and B frame types. In Fig. 2.b, I and P frames constitute the base layer.

B frames are predicted from I and P frames and constitute the second layer. In quality/SNR scalability, the temporal and spatial resolution of the video remains same and only the quality of the coded video is enhanced.

Applications like digital cinema, motion picture production and satellite imaging requires fast random access to individual video frames. For that purpose, SVC offers a new profile for intra frame video coding named 'Scalable High Intra Profile'. DWTSB spatially scalable coding framework, presented by Hsiang,^{2,3} for intra frames can be used to enhance the efficiency of this profile. This framework is quite flexible in selecting the wavelet coefficients for generating low resolution video at the base layer. In contrast to visual texture coding (VTC)⁷ of MPEG-4 which is based upon separate zero-tree based system for coding wavelet coefficients, DWTSB framework can be integrated to H.264 JSVM reference software without much modification. DWTSB based H.264 coding system takes advantage of the benefits of wavelet coding without much increase in implementation complexity.

3. PROPOSED METHOD

In this section we present the basic scan methodology in Section 3.1 and the proposed adaptive scans for high frequency subbands in Section 3.2.

3.1. Scan methodology

Let the transform coefficients be 2-dimensional array given as:

$$P_{m \times n} = \{p(i, j) : 1 \leq i \leq m, i \leq j \leq n\}. \tag{1}$$

After scanning the 2-dimensional array, we get a set:

$$Q_{mn} = \{1, \dots, mn\}. \tag{2}$$

One can note that scanning is a bijective function from $P_{m \times n}$ to Q_{mn} . Indeed, scanning of a 2D array is a permutation in which each element of the array is accessed exactly once.

Natural images generally consist of slow varying areas and contain lower frequencies both horizontally and vertically. After a transformation in the frequency domain, there are lot of non-zero transform coefficients (NZ) in the top left corner. Consequently, zigzag scan is more appropriate than other scans to convert a 2D array of transform coefficients to one directional array.

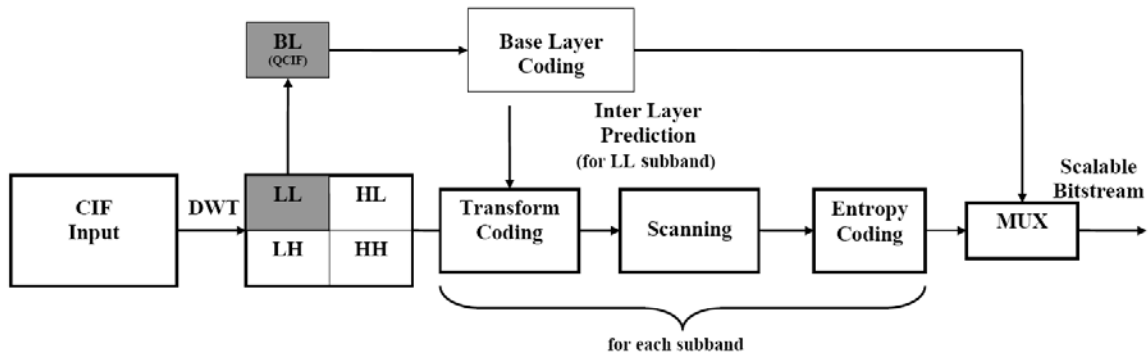


Figure 3. Architecture for dyadic scalable intra frame of H.264.

Entropy coding engine is designed to perform better when it gets most of the non-zero coefficients in the beginning of the block and long trail of zeros at the end of block, and magnitude of non-zero coefficients is higher at the start of the scanned array. This is the case for slowly changing video data when quantized coefficients

are scanned by traditional zigzag scan. Substituting the image by its wavelet subbands, each subband contains a certain range of frequencies. Traditional zigzag scan is not efficient for all the subbands as the energy is not concentrated in top left corner of 4x4 transform block and each subband should be scanned in a manner which generates most of the non-zero coefficients in the beginning and long trail of zeros at the end. We need only a single syntax element to indicate the number of the subband decomposition levels. No syntax is required for the scans since a specific scan is being used for each subband.

First of all, image is transformed to wavelet subbands and the LL subband is encoded as base layer by traditional H.264/MPEG4-AVC. In the enhancement layer, LL subband is predicted from the reconstructed base layer. Each high-frequency subband is encoded independently. They are transformed, quantized, scanned and then entropy coded as shown in Fig. 3. Our main contribution is to modify this scanning process and make it adaptive for every subband.

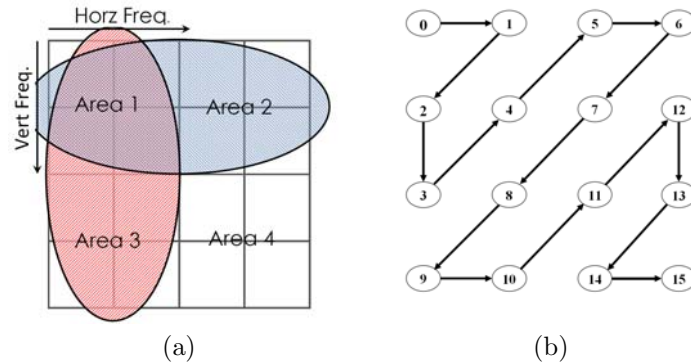


Figure 4. Analysis of LL subband: a) Dominant frequencies in transformed coefficients of LL subband, b) Zigzag scan used for such type of frequency distribution.

In order to analyze each subband in frequency domain, we propose to divide the 2D transform space into 4 areas, as shown in Fig. 4.a for LL subband. The area-1 contains most of the energy and has most of NZs. The area-2 and area-3 contain comparatively less number of NZs and only one frequency is dominant in these areas; either horizontal or vertical. The area-4 contains the least number of NZs. Fig. 4.a shows the frequency distribution in LL subband. It contains the lower frequencies in both horizontal and vertical directions and transform coefficients in this subband are scanned by traditional zigzag scan as illustrated in Fig. 4.b.

3.2. Proposed adaptive scan for high frequency subbands

In this section we propose to analyze the areas of the subbands HL, LH and HH in order to adapt the scanning processes. We have used wavelet critical sampling setting. Daubechies 9/7 wavelet filter set has been used to transform the image to the wavelet subbands at an enhancement layer. 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 several scan patterns for each subband. Nine standard video sequences containing various combinations of motion, color, contrast and objects have been used for the analysis.

HL and LH subbands do not contain horizontal and vertical frequencies in equal proportion. HL subband contains most of the high frequencies in horizontal direction while LH contains most of high frequencies in vertical direction. Because of non-symmetric nature of frequencies the scan pattern is not symmetric for HL and LH subbands except in the area-1 which contains both of the frequencies.

In HL subband, there are high horizontal frequencies and low frequencies in vertical direction. Area which contains many NZs should be then in top right corner, as illustrated in Fig. 5.a. Based on this, it should be scanned from top right corner to bottom left corner in a natural zigzag, as shown in Fig. 5.b. But separation of frequencies in subbands is not ideal and depends on the type of wavelet/subband filter used. It is also affected by rounding errors. So this simple zigzag scan is modified to get better results. Experimental results show that

DC coefficient still contains higher energy than other coefficients and should be scanned first. After that, we scan from the top left corner in a horizontal fashion till element 11, as illustrated in Fig. 5.c. At this position, we have two candidates to be scanned next: element 5 and element 15. We have already scanned the area-1 and zigzag scan is no more feasible. So, element 15 is then selected to be scanned first as it contains higher horizontal frequencies which are dominant in this subband. The same principle is true for the rest of scan lines and unidirectional scan from bottom to top gives better results, thus giving priority to the coefficients which contain higher horizontal frequencies. The final recommended scan for HL subband is illustrated in Fig. 5.c.

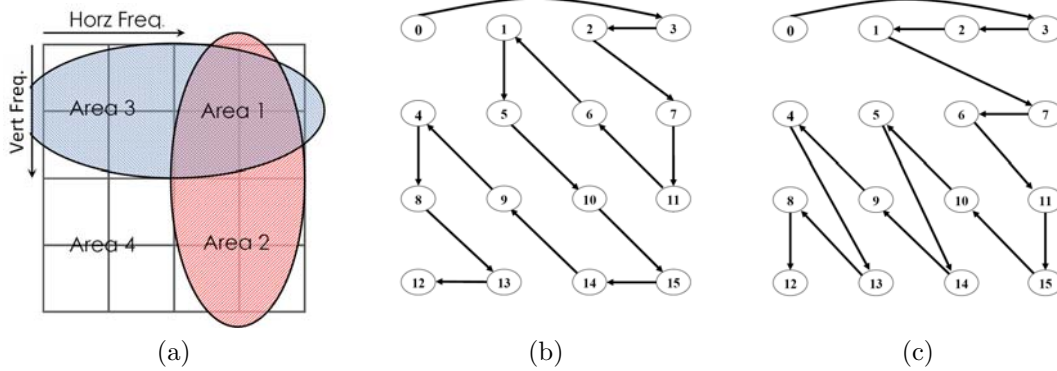


Figure 5. Analysis of HL subband: a) Dominant frequencies in transformed coefficients of this subband, b) Simple zigzag scan proposed for such type of frequency distribution, c) Proposed scan for HL subband.

Similarly for LH subband, there are low horizontal frequencies and high frequencies in vertical direction. This subband contains most of the NZs in bottom left corner, as illustrated in Fig. 6.a. Based on this, LH subband should be scanned in a zigzag fashion from bottom left corner to top right corner as shown in Fig. 6.b. But due to reasons similar to HL subband, this simple scan is modified to the final scan, illustrated Fig. 6.c, which is more appropriate for this subband. DC coefficient is scanned first and after, we scan from the bottom left corner in a zigzag fashion till element 5. Thereafter, unidirectional scan which gives priority to the coefficients containing higher vertical frequencies, is performed. The final recommended scan for LH subband is illustrated in Fig. 6.c.

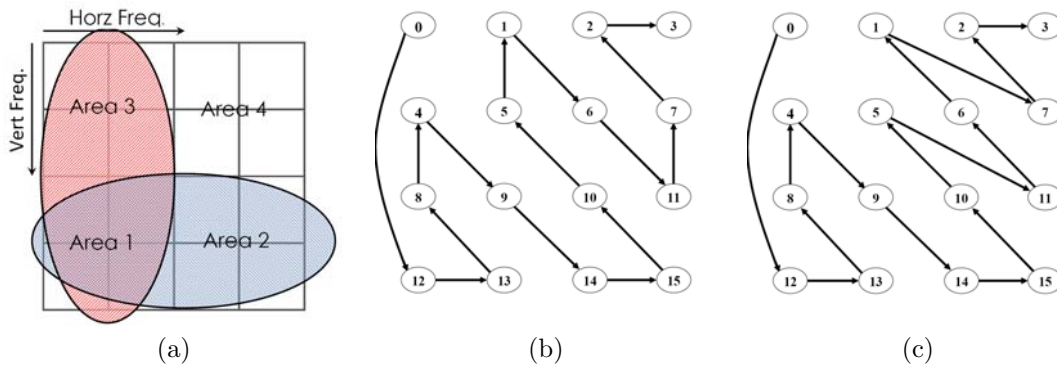


Figure 6. Analysis of LH subband: a) Dominant frequencies in transformed coefficients of this subband, b) Simple zigzag scan proposed for such type of frequency distribution, c) Proposed scan for LH subband.

HH subband contains higher frequencies both in horizontal and vertical directions as shown in 7.a. Frequencies which contain NZs should then be in bottom right. In this subband, DC coefficient contains the least energy and is scanned at the end. So it should be scanned from bottom right corner to top left corner in a zigzag fashion

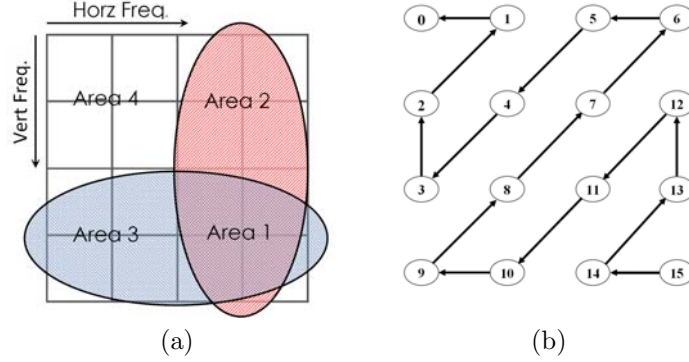


Figure 7. Analysis of HH subband: a) Dominant frequencies in transformed coefficients of this subband, b) Inverse zigzag scan proposed for such type of frequency distribution.

as shown in Fig. 7.b.

4. EXPERIMENTAL RESULTS

For the experimental results, nine standard standard video sequences have been used for the analysis in CIF and QCIF format. Each of them represents different combinations of motion (fast/slow, pan/zoom/rotation), color(bright/dull), contrast (high/low) and objects (vehicle, buildings, people). The video sequences 'bus', 'city' and 'foreman' contain camera motion while 'football' and 'soccer' contain camera panning and zooming along with object motion and texture in background. The video sequences 'harbour' and 'ice' contain high luminance images with smooth motion. 'Mobile' sequence contains a still complex background and motion in foreground. To demonstrate the effectiveness of our proposed scan, we have compressed 150 frames of each sequence at 30 fps.

DWTSB dyadic intra frame coding has already been demonstrated to perform better results than JSVM. Results illustrated in Fig. 8 for QP = 18 show that DWTSB-AS coding improves results comparing to DWTSB coding. In particular, adaptive scanning helps the entropy coder to perform a better coding and then gives a better compression without any compromise on quality. HH subband offers the best results since the appropriate scan for this subband is exactly opposite to simple zigzag scan. For example, for 'bus' video sequence, DWTSB-AS has reduced the over all size for the three high frequency subbands (HL, LH and HH) from 2983 kB to 2783 kB. File size of base layer and its residual remains the same since no modification has been made in their scan pattern. The improvements for the overall 2-layer video have been shown in Fig. 8.a for all the video sequences. Fig. 8.b-d show the file size reduction for HL, LH and HH subbands respectively.

To see the performance as a function of the QP over the whole rate distortion (R-D) curve, we have tested the proposed scans with QP values of 18, 24, 30, 36 over 150 frames of the same benchmark video sequences at 30 fps. The results show that the performance of adaptive scan is consistently better over the whole curve for all the benchmark sequences. Rather adaptive scans performs better at high QP values many times. Hence our scan performs better for all high frequency subbands over the whole R-D curve. Fig. 9.a gives the performance analysis overall 2-layer video 'mobile' at different QP values while Fig. 9.b-d give the performance analysis for the video 'mobile' at different QP values for the three subbands HL, LH and HH respectively.

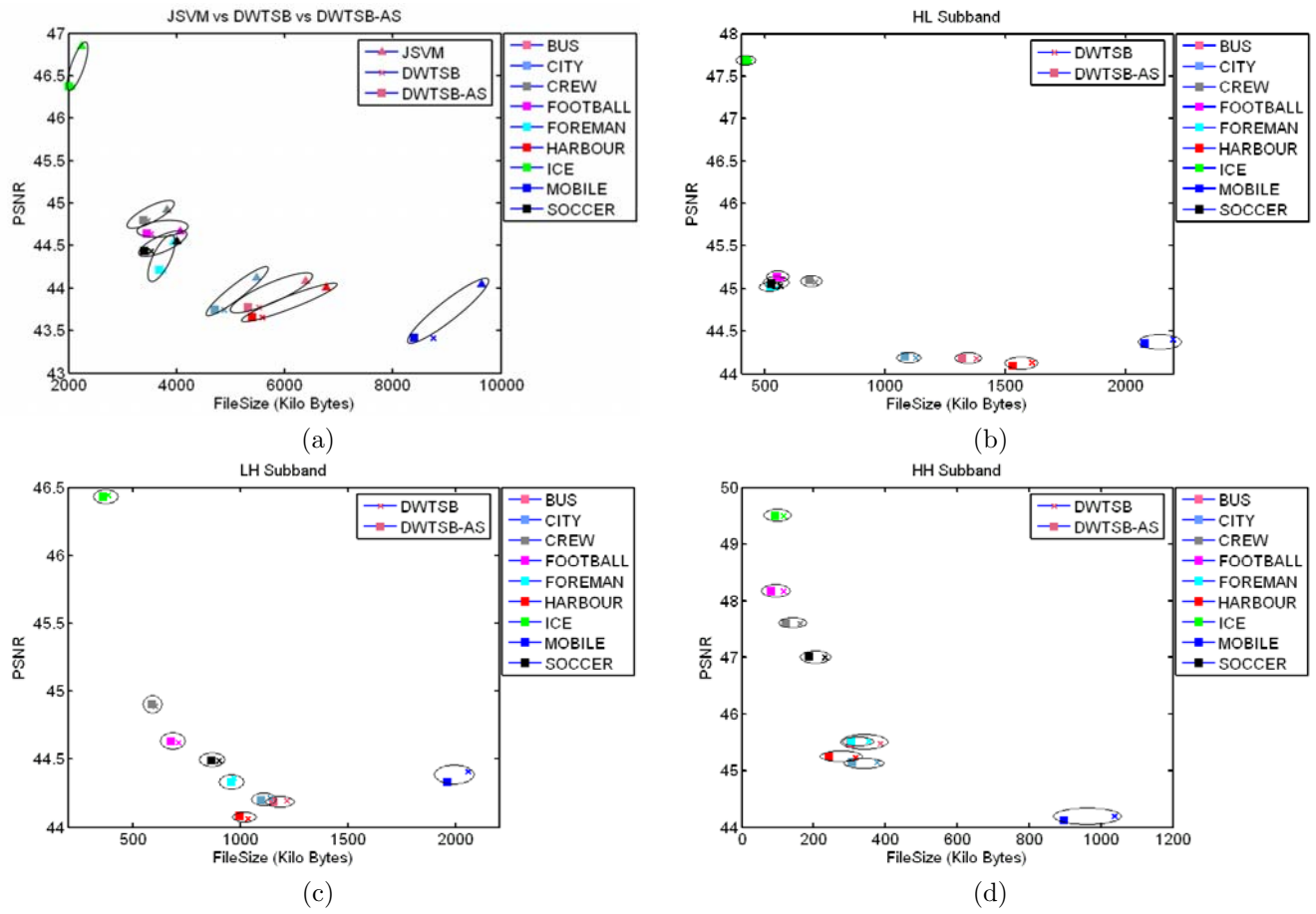


Figure 8. Comparison of JSVM, DWTSB and DWTSB-AS: a) Global comparison for two layer scalable bit streams, b) Only HL subband, c) Only LH subband, d) Only HH subband.

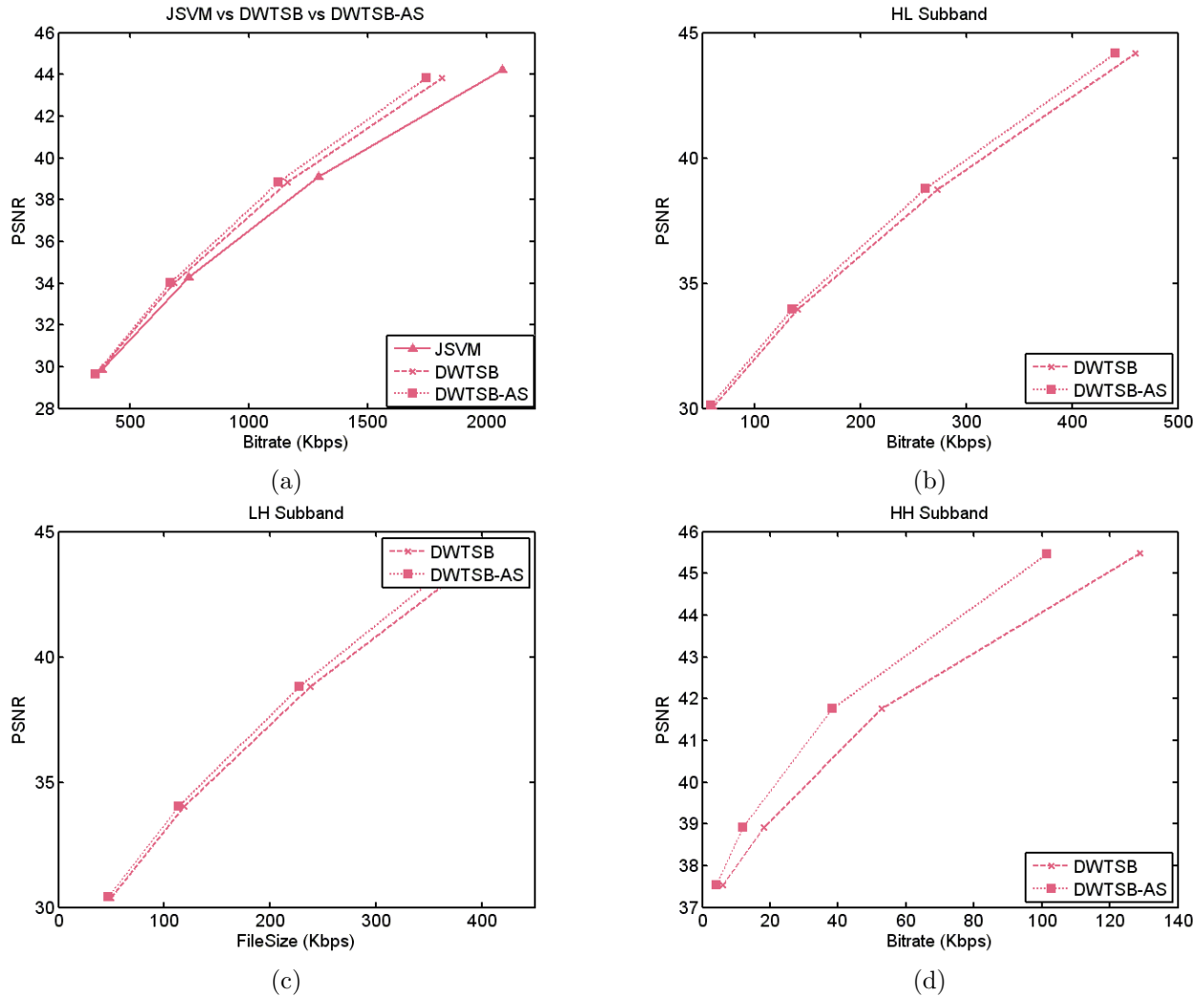


Figure 9. Performance analysis for the video 'mobile': a) Global comparison for two layer scalable bit streams, b) Only HL subband, c) Only LH subband, d) Only HH subband.

5. CONCLUSION

In this paper we have presented a new adaptive scanning methodology for intra frame scalable coding framework. We have described in detail the DWTSB-AS coding and we have shown that DWTSB-AS coding has done a significant file size reduction without any computation load for the same quality as compared to DWTSB coding. We have then elaborated the effectiveness of subband-specific scan for two layers by showing experimental results applied on several standard video sequences.

ACKNOWLEDGMENTS

This work is supported by the French National Project VOODOO of the ANR Content and Interaction. We also would like to thank the "Région Languedoc Roussillon", France.

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* **17(9)**, pp. 1103–1120, September 2007.
2. S.-T. Hsiang, "A New Subband/Wavelet Framework for AVC/H.264 Intra-Frame Coding and Performance Comparison with Motion-JPEG2000," in *VCIP08 Proceedings of the SPIE*, **6822**, 2007.
3. 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*, pp. 73–76, 2007.
4. T. Wiegand, G. Sullivan, J. Richel, H. Schwartz, and M. Wien, "Joint scalable video model (JSVM) 10," in *JVT-W202*, April 2007.
5. J. Jie, J. Eun-Ku, and K. Hae-Kwang, "Adaptive Transform Coefficient Scan for H.264 Intra Coding," *IEICE Trans Inf Syst* **90(10)**, pp. 1709–1711, 2007.
6. ISO/IEC, *13818-2:2000 Information technology, Generic coding of moving pictures and associated audio information: Video, 2nd Ed.*, 2000.
7. ISO/IEC, *14496-2:2004 Information technology, Coding of audio-visual objects: Visual, 3rd Ed.*, 2004.
8. ITU, *Telecommunication Standardization Sector of ITU, Video Coding for Low Bitrate Communication, ITU-T Recommendation H.263 Version 2*, 1998.
9. S.-T. Hsiang, "CE3: Intra-Frame Dyadic Spatial Scalable Coding Based on a Subband/Wavelet Filter Banks Framework," Joint Video Team, Doc. JVT-W097, April 2007.