Scalable Video Coding

Zafar Shahid, Marc Chaumont, William Puech

1

Scalable Video Coding

Z. Shahid, M. Chaumont and W. Puech
LIRMM / UMR 5506 CNRS / Universite´ Montpellier II
France

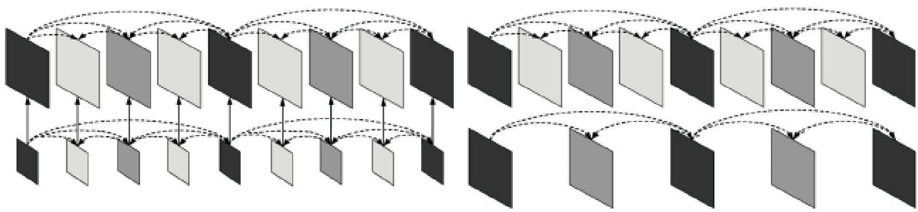
1. Introduction

With the evolution of Internet to heterogeneous networks both in terms of processing power and network bandwidth, different users demand the different versions of the same content. This has given birth to the scalable era of video content where a single bitstream contains multiple versions of the same video content which can be different in terms of resolutions, frame rates or quality. Several early standards, like MPEG2 video, H.263, and MPEG4 part II already include tools to provide different modalities of scalability. However, the scalable profiles of these standards are seldom used. This is because the scalability comes with significant loss in coding efficiency and the Internet was at its early stage. Scalable extension of H.264/AVC is named scalable video coding and is published in July 2007. It has several new coding techniques developed and it reduces the gap of coding efficiency with state-of-the-art non-scalable codec while keeping a reasonable complexity increase.
After an introduction to scalable video coding, we present a proposition regarding the scalable functionality of H.264/AVC, which is the improvement of the compression ratio in enhancement layers (ELs) of subband/wavelet based scalable bitstream. A new adaptive scanning methodology for *intra* frame scalable coding framework based on subband/wavelet coding approach is presented for H.264/AVC scalable video coding. It takes advantage of the prior knowledge of the frequencies which are present in different higher frequency subbands. Thus, by just modification of the scan order of the *intra* frame scalable coding framework of H.264/AVC, we can get better compression, without any compromise on PSNR.
This chapter is arranged as follows. We have presented introduction to scalable video in Section 2, while Section 3 contains a discussion on scalable extension of H.264/AVC. Comparison of scalable extension of different video codecs is presented in Section 4. It is followed by adaptive scan algorithm for enhancement layers (ELs) of subband/wavelet based scalable architecture in Section 5. At the end, concluding remarks regarding the whole chapter are presented in Section 6.

2. Basics of scalability

Historically simulcast coding has been used to achieve scalability. In simulcast coding, each layer of video is coded and transmitted independently. In recent times, it has been replaced by scalable video coding (SVC). In SVC, the video bitstream contains a base layer and number of enhancement layers. Enhancement layers 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.
In spatial scalability, the inter-layer prediction of the enhancement-layer is utilized to remove redundancy across video layers as shown in Fig. 1.a. The resolution of the enhancement layer is either equal or greater than the lower layer. Enhancement layer predicted (P) frames 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. 1.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 as shown in Fig. 2.
Individual scalabilities can be combined to form mixed scalability for a specific application. Video streaming over heterogeneous networks, which request same video content but with different resolutions, qualities and frame rates is one such example. The video content is encoded just once for the highest requested resolution, frame rate and bitrate, forming a scalable bitstream from which representations of lower resolution, lower frame rate and lower quality can be obtained by partial decoding. Combined scalability is a desirable feature for video transmission in networks with unpredictable throughput variations and can be used for bandwidth adaptation Wu et al. (2000). It is also useful for unequal error adaptation Wang et al. (2000), wherein the base layer can be sent over a more reliable channel, while the enhancement layers can be sent over comparatively less reliable channels. In this case, the connection will not be completely interrupted in the presence of transmission error and a base-layer quality can still be received.



(a) (b)

Fig. 1. 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) Temporal scalability in which, 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.

3. Scalable extension of H.264/AVC

Previous video standards such as MPEG2 MPEG2 (2000), MPEG4 MPEG4 (2004) and H.263+ H263 (1998) also contain the scalable profiles but they were not much appreciated because the quality and scalability came at the cost of coding efficiency. Scalable video coding (SVC) based on H.264/AVC ISO/IEC-JTC1 (2007) has achieved significant improvements both in terms of coding efficiency and scalability as compared to scalable extensions of prior video coding standards.
The call for proposals for efficient scalable video coding technology was made in October 2003. 12 of the 14 submitted proposals represented scalable video codecs based on a 3-D wavelet

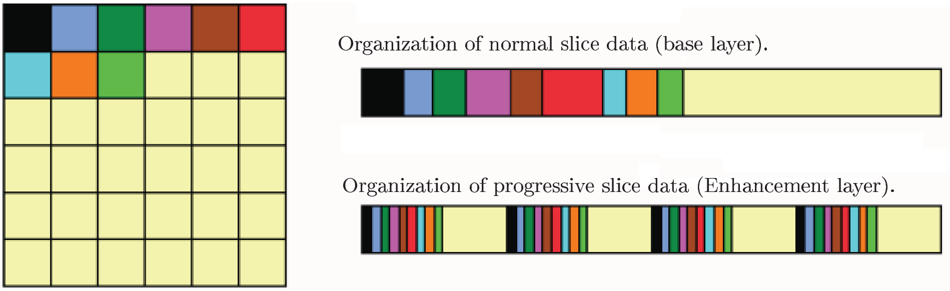


Fig. 2. SNR scalable architecture of SVC.

transform, while the remaining two proposals were extension of H.264/AVC. The scalable extension of H.264/AVC as proposed by Heinrich Hertz Institute (HHI) was chosen as the starting point of Scalable Video Coding (SVC) project in October 2004. In January 2005, ISO and ITU-T agreed to jointly finalize the SVC project as an Amendment of their H.264/AVC standard, named as scalable extension of H.264/AVC standard. The standardization activity of this scalable extension was completed and the standard was published in July 2007, which completed the milestone for scalable extension of H.264/AVC to become the state-of-the-art scalable video codec in the world. Similar to the previous scalable video coding propositions, Scalable extension of H.264/AVC is also built upon a predictive and layered approach to scalable video coding. It offers spatial, temporal and SNR scalabilities, which are presented in Section 3.1, Section 3.2 and Section 3.3 respectively.

3.1 Spatial scalability in scalable extension of H.264/AVC

Spatial scalability is achieved by pyramid approach. The pictures of different spatial layers are independently coded with layer specific motion parameters as illustrated in Fig. 3. In order to improve the coding efficiency of the enhancement layers in comparison to simulcast, additional inter-layer prediction mechanisms have been introduced to remove the redundancies among layers. These prediction mechanisms are switchable so that an encoder can freely choose a reference layer for an enhancement layer to remove the redundancy between them. Since the incorporated inter-layer prediction concepts include techniques for motion parameter and residual prediction, the temporal prediction structures of the spatial layers should be temporally aligned for an efficient use of the inter-layer prediction. Three inter-layer prediction techniques, included in the scalable extension of H.264/AVC, are:

- *Inter-layer motion prediction*: In order to remove the redundancy among layers, additional MB modes have been introduced in spatial enhancement layers. The MB partitioning is obtained by up-sampling the partitioning of the co-located 8x8 block in the lower resolution layer. The reference picture indices are copied from the co-located base layer blocks, and the associated motion vectors are scaled by a factor of 2. These scaled motion vectors are either directly used or refined by an additional quarter-sample motion vector refinement. Additionally, a scaled motion vector of the lower resolution can be used as motion vector predictor for the conventional MB modes.
- *Inter-layer residual prediction*:The usage of inter-layer residual prediction is signaled by a flag that is transmitted for all inter-coded MBs. When this flag is true, the base layer signal

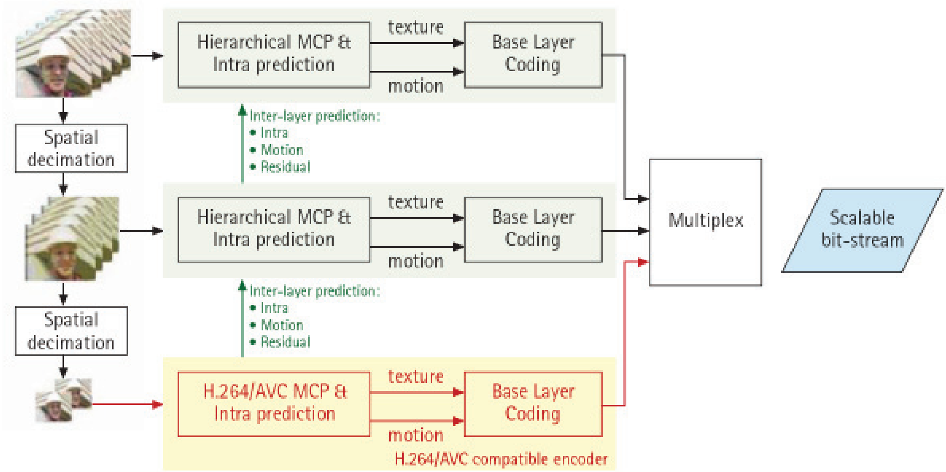


Fig. 3. Spatial scalable architecture of scalable extension of H.264/AVC.

of the co-located block is block-wise up-sampled and used as prediction for the residual signal of the current MB, so that only the corresponding difference signal is coded.

- *Inter-layer intra prediction*:Furthermore, an additional intra MB mode is introduced, in which the prediction signal is generated by up-sampling the co-located reconstruction signal of the lower layer. For this prediction it is generally required that the lower layer is completely decoded including the computationally complex operations of motion-compensated prediction and deblocking. However, this problem can be circumvented when the inter-layer intra prediction is restricted to those parts of the lower layer picture that are intra-coded. With this restriction, each supported target layer can be decoded with a single motion compensation loop.

3.2 Temporal scalability in scalable extension of H.264/AVC

Temporal scalable bitstream can be generated by using hierarchical prediction structure without any changes to H.264/AVC. A typical hierarchical prediction with four dyadic hierarchy stages is depicted in Fig. 4. Four temporal scalability levels are provided by this structure. The first picture of a video sequence is intra-coded as IDR picture that are coded in regular (or even irregular) intervals. A picture is called a key picture when all previously coded pictures precede this picture in display order. A key picture and all pictures that are temporally located between the key picture and the previous key picture consist of a group of pictures (GOP). The key pictures are either intra-coded or inter-coded using previous (key) pictures as reference for motion compensated prediction, while the remaining pictures of a GOP are hierarchically predicted. . For example, layer 0, 1, 2 and 3 contains 3, 5, 9 and 18 frames respectively in Fig. 4.

3.3 SNR scalability in scalable extension of H.264/AVC

For SNR scalability, scalable extension of H.264/AVC provides coarse-grain SNR scalability (CGS) and medium-grain SNR scalability (MGS). CGS scalable coding is achieved using the same inter-layer prediction mechanisms as in spatial scalability. MGS is aimed at increasing

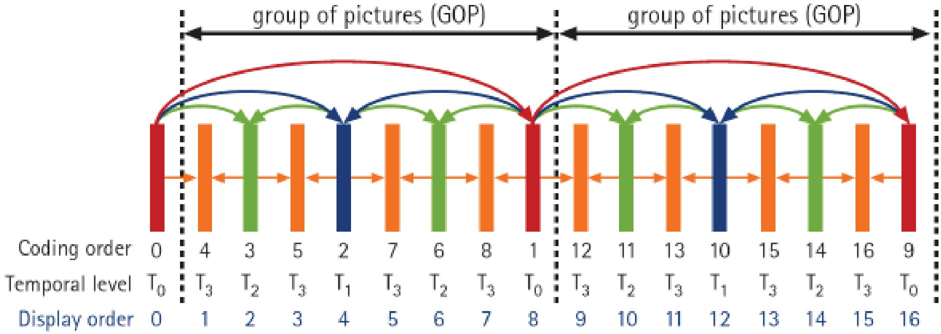


Fig. 4. Temporal scalable architecture of Scalable extension of H.264/AVC.

the granularity for SNR scalability and allows the adaptation of bitstream adaptation at network adaptation layer (NAL) unit basis. CGS and MGS are presented in details in Section 3.3.1 and Section 3.3.2 respectively.

3.3.1 Coarse-grain SNR scalability

Coarse-grain SNR scalable coding is achieved using the concepts for spatial scalability. The same inter-layer prediction mechanisms are employed. The only difference is that base and enhancement layers have the same resolution. The CGS only allows a few selected bitrates to be supported in a scalable bitstream. In general, the number of supported rate points is identical to the number of layers. Switching between different CGS layers can only be done at defined points in the bitstream. Furthermore, the CGS concept becomes less efficient when the relative rate difference between successive CGS layers gets smaller.

3.3.2 Medium-grain SNR scalability

In order to increase the granularity for SNR scalability, scalable extension of H.264/AVC provides a variation of CGS approach, which uses the quality identifier Q for quality refinements. This method is referred to as MGS and allows the adaptation of bitstream adaptation at a NAL unit basis. With the concept of MGS, any enhancement layer NAL unit can be discarded from a quality scalable bitstream and thus packet based SNR scalable coding is obtained. However, it requires a good controlling of the associated drift. MGS in scalable extension of H.264/AVC has evolved from SNR scalable extensions of MPEG2/4. So it is pertinent to start our discussion from there and extend it to MGS of H.264/AVV.
The prediction structure of FGS in MPEG4 Visual was chosen in a way that drift is completely omitted. Motion compensation prediction in MPEG4 FGS is usually performed using the base layer reconstruction for reference as illustrated in Fig. 5.a. Hence loss of any enhancement packet does not result in any drift on the motion compensated prediction loops between encoder and decoder. The drawback of this approach, however, is the significant decrease of enhancement layer coding efficiency in comparison to single layer coding, because the temporal redundancies in enhancement layer cannot be properly removed.
For SNR scalability coding in MPEG2, the other extreme case was specified. The highest enhancement layer reconstruction is used in motion compensated prediction as shown in

Fig. 5.b. This ensures a high coding efficiency as well as low complexity for the enhancement layer. However, any loss or modification of a refinement packet results in a drift that can only be stopped by intra frames.

For the MGS in scalable extension of H.264/AVC, an alternative approach, which allows certain amount of drift by adjusting the trade off between drift and enhancement layer coding efficiency is used. The approach is designed for SNR scalable coding in connection with hierarchical prediction structures. For each picture, a flag is transmitted to signal whether the base representations or the enhancement representations are employed for motion compensated prediction. Picture that only uses the base representations (Q=0) for prediction is also referred as key pictures. Fig. 6 illustrates how the key picture can be combined with hierarchical prediction structures.

All pictures of the coarsest temporal level are transmitted as key pictures, and thus no drift is introduced in the motion compensated loop of temporal level 0. In contrast to that, all temporal refinement pictures are using the highest available quality pictures as reference in motion compensated prediction, which results in high coding efficiency for these pictures. Since key pictures serve as the resynchronization point between encoder and decoder reconstruction, drift propagation can be efficiently contained inside a group of pictures. The trade off between drift and enhancement layer coding efficiency can be adjusted by the choice of GOP size or the number of hierarchy stages.

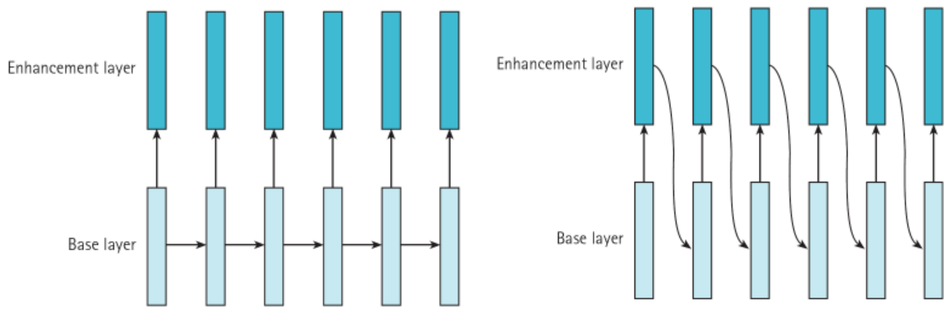


Fig. 5. SNR scalable architecture for (a) MPRG4, (b) MPRG2.

4. Performance comparison of different scalable architectures

In comparison to early scalable standards, scalable extension of H.264/AVC provides various tools for improving efficiency relative to single-layer coding. The key features that make the scalable extension of H.264/AVC superior than all scalable profiles are:

- The employed hierarchical prediction structure that provides temporal scalability with several levels improves the coding efficiency and effectiveness of SNR and spatial scalable coding.
- The concept of key pictures controls the trade off between drift and enhancement layer coding efficiency. It provides a basis for efficient SNR scalability, which could not be achieved in all previous standards.
- New modes for inter-layer prediction of motion and residual information improves coding efficiency of spatial and SNR scalability. In all previous standards, only residual information can be refined at enhancement layers.

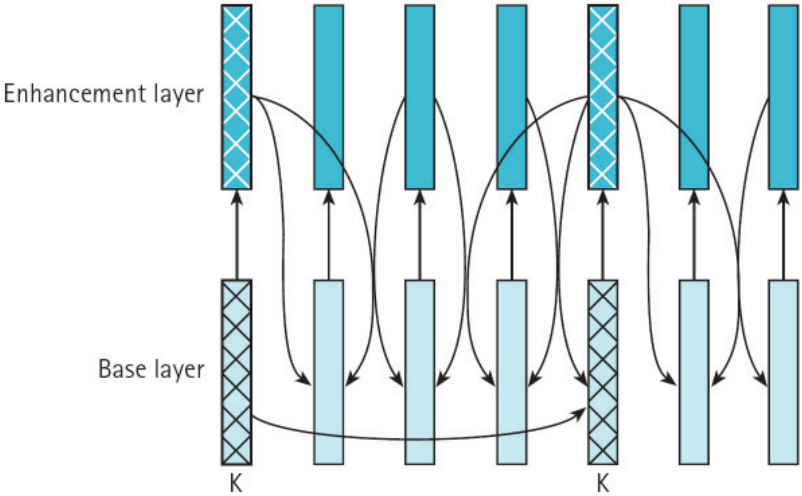


Fig. 6. SNR scalable architecture of Scalable extension of H.264/AVC.

- The coder structure is designed in a more flexible way such that any layer can be configured to be the optimization point in SNR scalability. MPEG2 is designed in the sense that enhancement layer is always optimized but the base layer may suffer from a serious drift problem that causes significant quality drop. MPEG4 FGS, on the other way round, usually coded in a way to optimize base layer and the coding efficiency of enhancement layer is much lower than single layer coding. In scalable extension of H.264/AVC, the optimum layer can be set to any layer with a proper configuration Li et al. (2006).
- Single motion compensated loop decoding provides a decoder complexity close to single layer decoding.

To conclude, with the advances mentioned above, scalable extension of H.264/AVC, has enabled profound performance improvements for both scalable and single layer coding. Results of the rate-distortion comparison show that scalable extension of H.264/AVC clearly outperforms early video coding standards, such as MPEG4 ASP Wien et al. (2007). Although scalable extension of H.264/AVC still comes at some costs in terms of bitrate or quality, the gap between the state-of-the-art single layer coding and scalable extension of H.264/AVC can be remarkably small.

5. Adaptive scan for high frequency (HF) subbands in SVC

Scalable video coding (SVC) standard Schwarz & Wiegand (2007) 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. 7.a. Thus, the process applied to the base layer must be the best possible in order to improve the quality.

Hsiang Hsiang (2008) 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. 7.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. 7.c. The results presented by Hsiang have proved to be better than H.264 scalable video coding Wiegand et al. (April 2007) for intra frame. In dyadic scalable intra frame coding, the image is transformed to wavelet subbands and then the subbands are encoded by base-layer H.264/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.

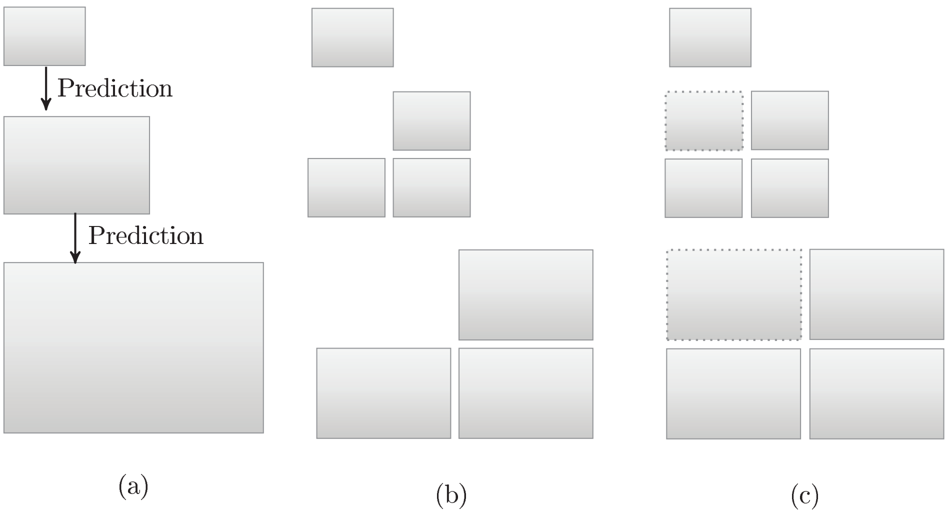


Fig. 7. Different scalable video coding approaches: (a) Pyramid coding used in JSVM, (b) Wavelet subband coding used in JPEG2000, (c) Wavelet subband coding for dyadic scalable intra frame.

This section presents a new scanning methodology for intra frame scalable coding framework based on a subband/wavelet (DWTSB) scalable architecture. It takes advantage of the prior knowledge of the frequencies which are present in different higher frequency (HF) subbands. An adaptive scan (DWTSB-AS) is proposed for HF subbands as traditional zigzag scan is designed for video content containing most of its energy in low frequencies. Hence, we can get better compression by just modification of the scan order of DCT coefficients. The proposed algorithm has been theoretically justified and is thoroughly evaluated against the current SVC test model JSVM and DWTSB through extensive coding experiments. The simulation results show the proposed scanning algorithm consistently outperforms JSVM and DWTSB in terms of PSNR.

5.1 Scan methodology

Let the QTCs be 2-dimensional array given as:

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

After scanning the 2-dimensional array, we get a 1-dimensional array $Q_{mn} = \{1, ..., mn\}$, using 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 to put QTCs with higher magnitude at the start of the array.
Entropy coding engine is designed to perform better when:

1. It gets most of the non-zero QTCs in the beginning of scanned and long trail of zeros at its end.
2. 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. Zigzag scan is thus no more efficient for all the subbands as the energy is not concentrated in top left corner of 4x4 transform block. Each subband should be scanned in a manner that entropy coding module do maximum possible compression. In other words, most of the non-zero QTCs should be in the beginning and a long trail of zeros at the end of the scanned array.

5.2 Analysis of each subband in transform domain

In DWTSB scalable video architecture, an image is transformed to wavelet subbands and the LL subband is encoded as base layer by traditional H.264/AVC. In the enhancement layer, LL subband is predicted from the reconstructed base layer. Each high-frequency subband is encoded independently using base-layer H.264/AVC as shown in Fig. 8.

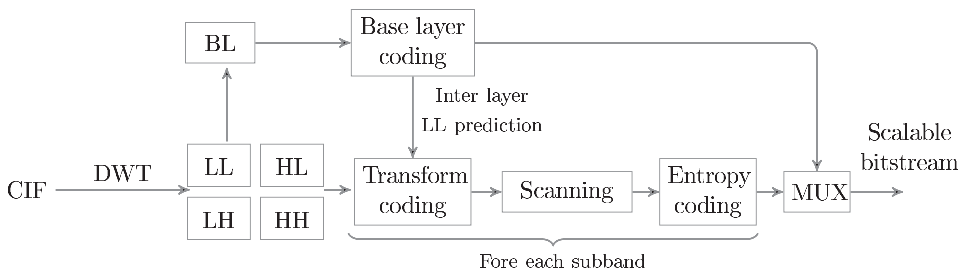
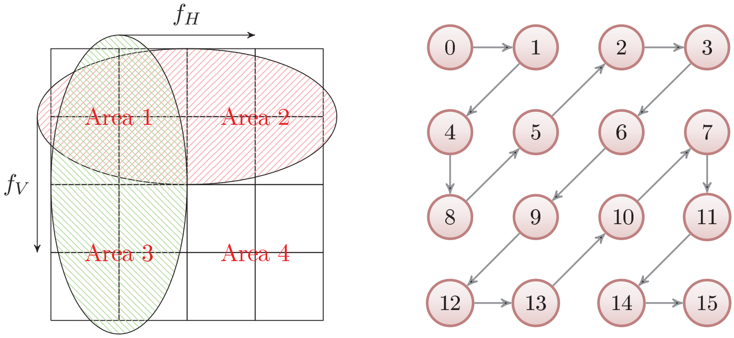


Fig. 8. DWTSB scalable architecture based on H.264/AVC.

For this work, we have used wavelet critical sampling setting. Daubechies 9/7 wavelet filter set has been used to transform the video frame to four wavelet subbands. The work has been performed on 'JVT-W097' Hsiang (2007) which is referenced H.264 JSVM 8.9 with wavelet framework integrated.
In order to analyze each subband in transform domain, we propose to divide the 2D transform space into 4 areas, *e.g.* as shown in Fig. 9.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. 9.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. 9.b.



(a) (b)

Fig. 9. Analysis of LL subband: (a) Dominant frequencies in transformed coefficients of LL subband, (b) Zigzag scan is suitable for such type of frequency distribution.

5.3 Adaptive scan for HF subbands

In this section we present our proposition which is to use DWTSB scalable architecture along-with adaptive scan (DWTSB-AS) for HF subbands. We analyze the frequencies present in HL, LH and HH subbands in order to adapt the scanning processes.

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 in 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. 10.a. Based on this, it should be scanned from top right corner to bottom left corner in a natural zigzag, as shown in Fig. 10.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. It is followed by a scan from the top left corner in a horizontal fashion till element 11, as illustrated in Fig. 10.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.

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. 11.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. 11.b. But due to reasons similar to HL subband,

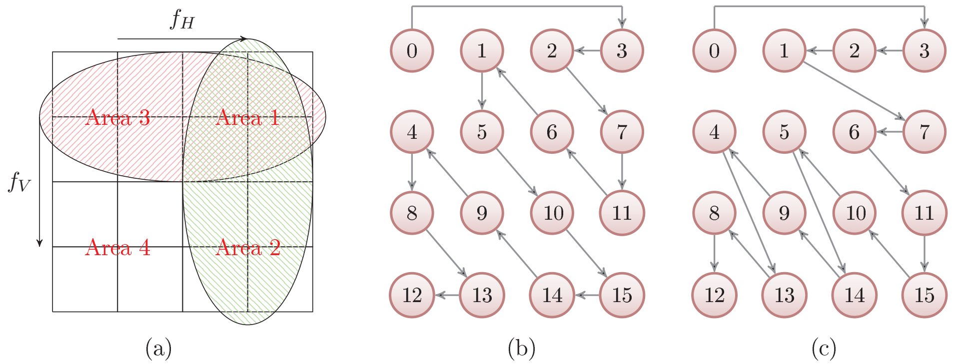


Fig. 10. Analysis of HL subband: (a) Dominant frequencies in QTCs of this subband, (b) Simple zigzag scan proposed for such type of frequency distribution, (c) Proposed scan for HL subband.

this simple scan is modified to a more efficient scan, illustrated Fig. 11.c. DC coefficient is scanned first. It is followed by a 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.

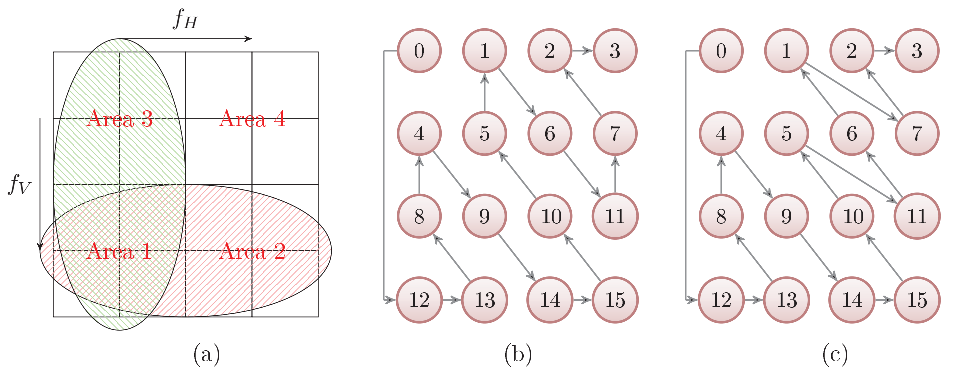


Fig. 11. Analysis of LH subband: (a) Dominant frequencies in QTCs of this subband, (b) Simple zigzag scan proposed for such type of frequency distribution, (c) Proposed scan for LH subband.

HH subband contains higher frequencies both in horizontal and vertical directions as shown in 12.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 as shown in Fig. 12.b.

5.4 Experimental results

For the experimental results, nine standard video sequences have been used for the analysis in CIF and QCIF format. To apply our approach we have compressed 150 frames of each sequence at 30 fps.

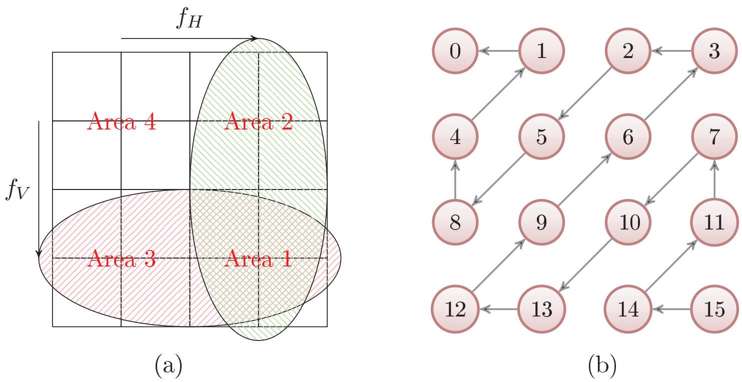
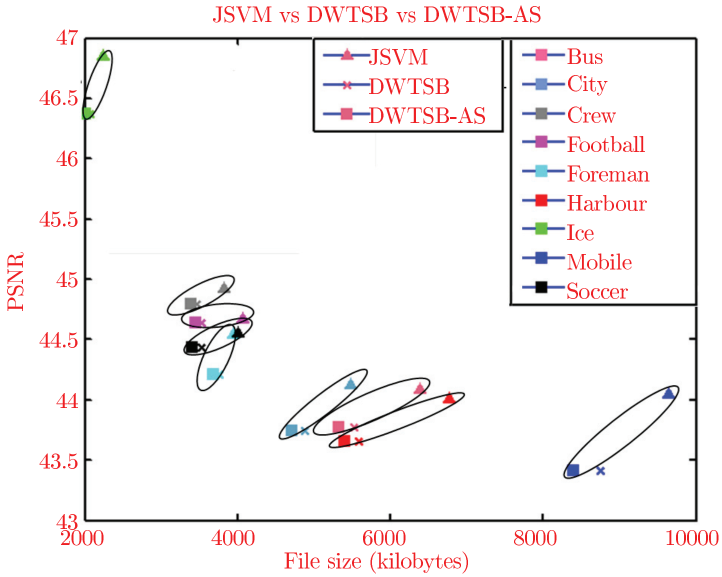
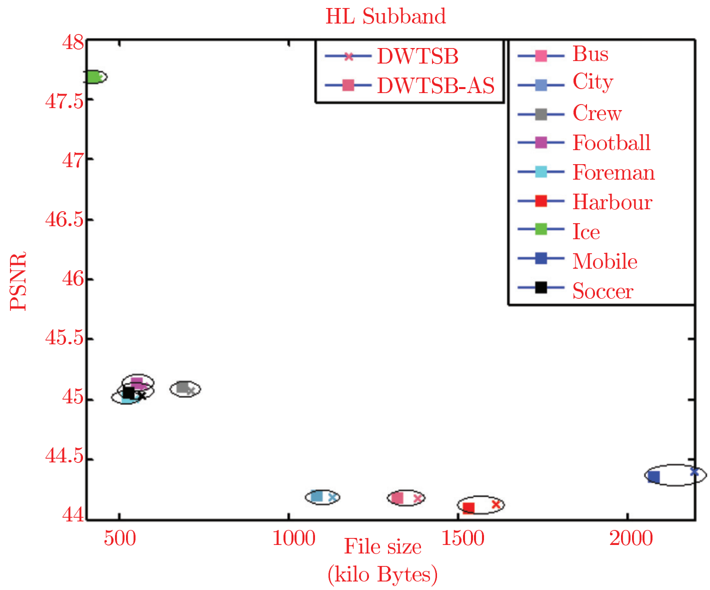


Fig. 12. Analysis of HH subband: (a) Dominant frequencies in QTCs of this subband, (b) Inverse zigzag scan proposed for such type of frequency distribution.

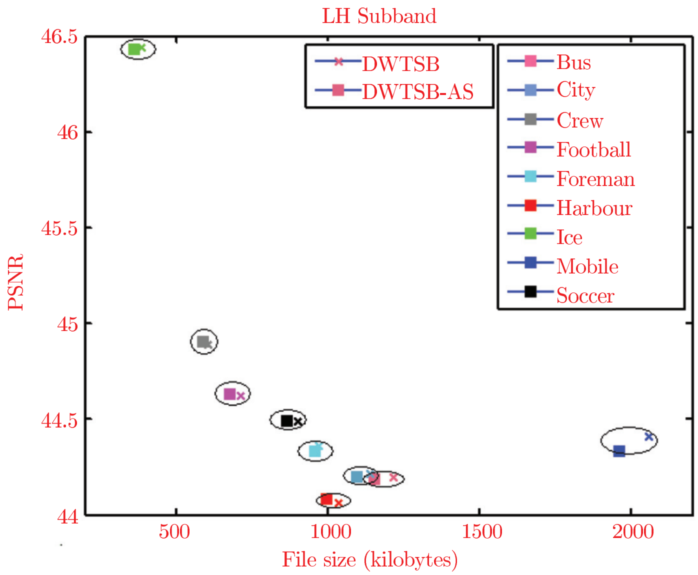
For the experimental results, nine benchmark video sequences have been used for the analysis in QCIF format. Each of them represents different combinations of motion (fast/slow, pan/zoom/rotation), color (bright/dull), contrast (high/low) and objects (vehicle, buildings, people). The video sequences '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 complex still background and foreground motion.
DWTSB dyadic intra frame coding has already been demonstrated to perform better results than JSVM. Results illustrated in Fig. 13 for QP value 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 overall bitstream size for the three high frequency subbands (HL, LH and HH) from **2049 kB to 1863 kB** as shown in Fig. 13.a. 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. 13.a for all the video sequences. Fig. 13.b-d show the file size reduction for HL, LH and HH subbands respectively.
To see the performance as a function of the QP value over the whole rate distortion (R-D) curve, we have tested the proposed scans over 150 frames of the same benchmark video sequences with QP values of 18, 24, 30 and 36. The results show that the performance of adaptive scan is consistent over the whole curve for all the benchmark sequences. Rather adaptive scans perform at high QP values times. Hence our scan performs better for all high frequency subbands over the whole R-D curve. Fig. 14.a gives the performance analysis overall 2-layer video *mobile* at different QP values since Fig. 14.b-d give the performance analysis for the video *mobile* at different QP values for the three subbands HL, LH and HH respectively.
To summarize, we have presented a new adaptive scanning methodology for DWTSB scalable architecture of dyadic intra frames in Section 5.4. We have described in detail the DWTSB-AS scheme. DWTSB-AS has done a significant file size reduction without any computation load


JSVM vs DWTSB vs DWTSB-AS
PSNR
File size (kilobytes)
47
46.5
46
45.5
45
44.5
44
43.5
43
2000
4000
6000
8000
10000
JSVM
DWTSB
DWTSB-AS
Bus
City
Crew
Football
Foreman
Harbour
Ice
Mobile
Soccer

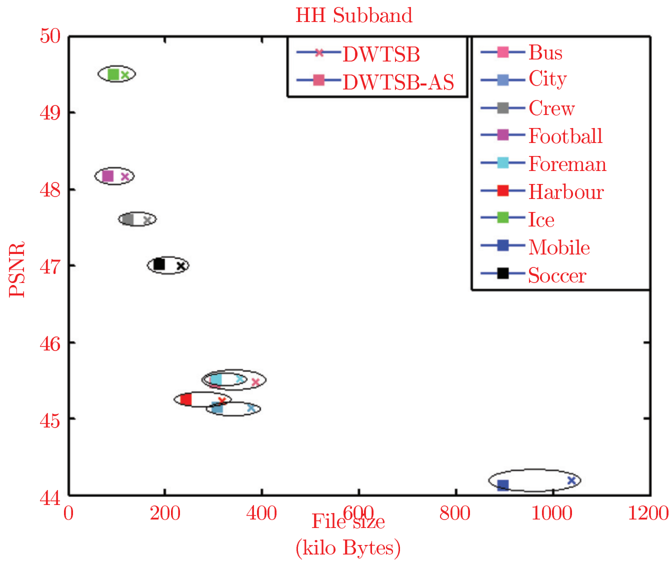
(a)


HL Subband
PSNR
File size
(kilo Bytes)
48
47.5
47
46.5
46
45.5
45
44.5
44
500
1000
1500
2000
DWTSB
DWTSB-AS
Bus
City
Crew
Football
Foreman
Harbour
Ice
Mobile
Soccer

(b)

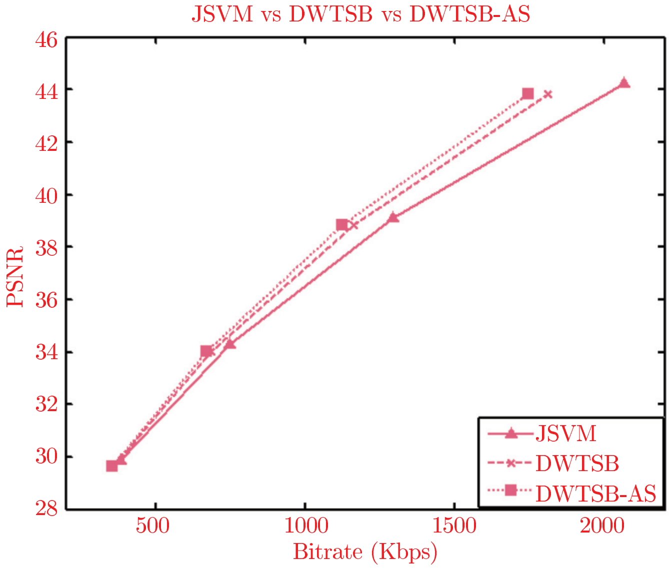


(c)

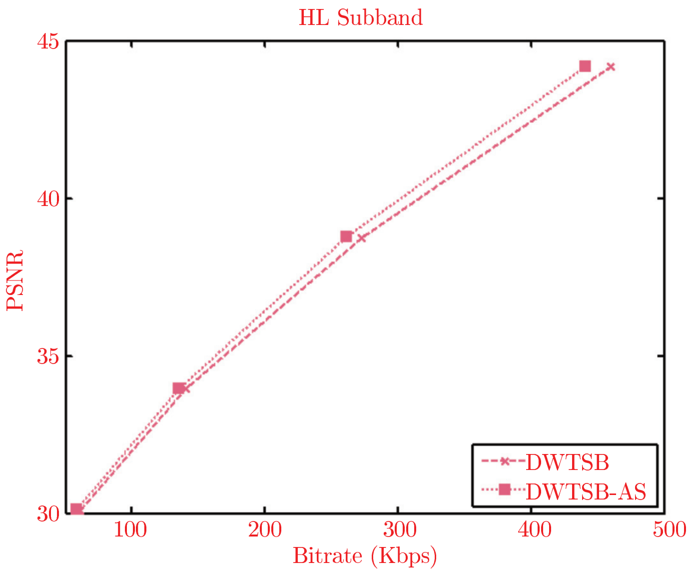


(d)

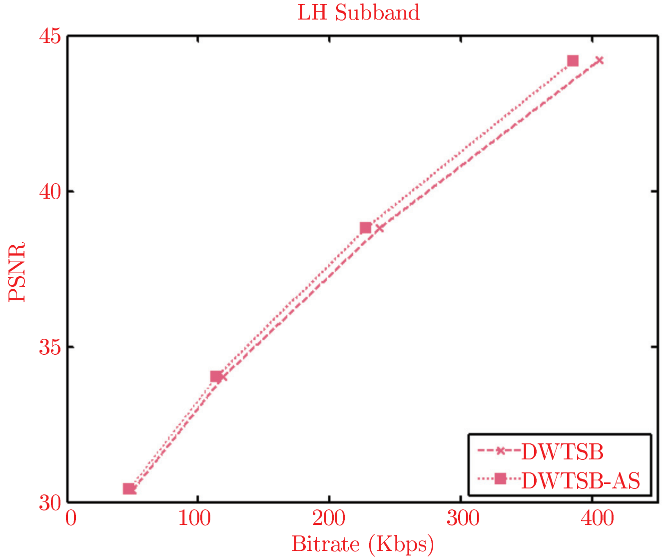
Fig. 13. Comparison of JSVM, DWTSB and DWTSB-AS: (a) Global comparison for two layer scalable bitstreams, (b) HL subband comparison, (c) LH subband comparison, (d) HH subband comparison.


JSVM vs DWTSB vs DWTSB-AS
PSNR
46
44
42
40
38
36
34
32
30
28
500
1000
1500
2000
Bitrate (Kbps)
JSVM
DWTSB
DWTSB-AS

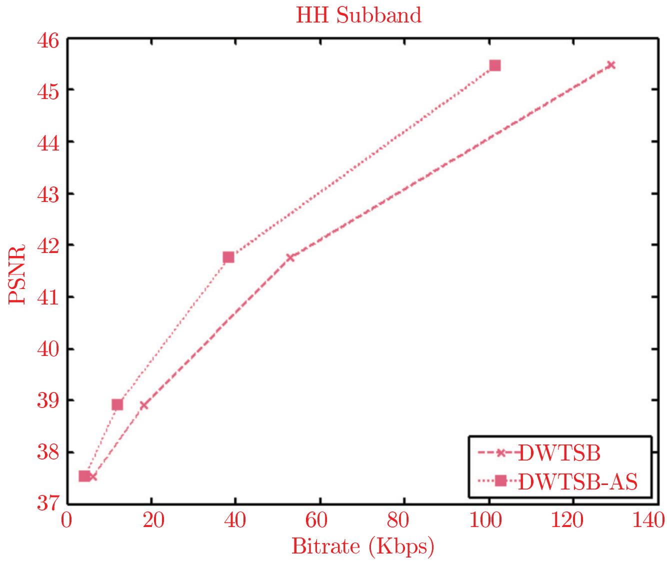
(a)


HL Subband
PSNR
45
40
35
30
100
200
300
400
500
Bitrate (Kbps)
DWTSB
DWTSB-AS

(b)



(c)



(d)

Fig. 14. Performance comparison of JSVM, DWTSB and DWTSB-AS for *mobile* video sequence over whole QP range: (a) Global comparison for two layer scalable bitstreams, (b) HL subband, (c) LH subband, (d) HH subband.

for the same quality as compared to DWTSB coding. Effectiveness of subband-specific scan for DWTSB scalable video has been elaborated by showing experimental results on several benchmark video sequences containing diverse content.

6. Summary

In this chapter, we have presented the scalable extension of H.264/AVC and its comparison with previous scalable video architectures. Extra prediction modes in spatial scalability and SNR scalability have resulted in extra performance of this architecture. It is followed by our contribution related to spatially scalable video. First of all, we have presented the DWT based spatial scalable architecture. It is followed by proposed adaptive scanning methodology for DWTSB 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 Shahid et al. (2009). We have then elaborated the effectiveness of subband-specific scan for two layers by showing experimental results applied on several standard video sequences.

7. References

H263 (1998). ITU Telecommunication Standardization Sector of ITU, Video Coding for Low Bitrate Communication, *ITU-T Recommendation H.263 Version 2*.

Hsiang, S. (2007). CE3: Intra-frame Dyadic Spatial Scalable Coding Based on a Subband/Wavelet Filter Banks Framework, Joint Video Team, Doc. JVT-W097.

Hsiang, S. (2008). A New Subband/Wavelet Framework for AVC/H.264 Intra-Frame Coding and Performance Comparison with Motion-JPEG2000, *SPIE, Visual Communications and Image Processing*, Vol. 6822, pp. 1–12.

ISO/IEC-JTC1 (2007). Advanced Video Coding for Generic Audio-Visual Services, *ITU-T Recommendation H.264 Amendment 3, ISO/IEC 14496-10/2005:Amd 3 - Scalable extension of H.264 (SVC)* .

Li, Z., Rahardja, S. & Sun, H. (2006). Implicit Bit Allocation for Combined Coarse Granular Scalability and Spatial Scalability, *IEEE Transactions on Circuits and Systems for Video Technology* 16(12): 1449 –1459.

MPEG2 (2000). ISO/IEC 13818-2:2000 Information Technology – Generic Coding of Moving Pictures and Associated Audio Information: Video, 2nd Edition.

MPEG4 (2004). ISO/IEC 14496-2:2004 Information Technology – Coding of Audio-Visual Objects: Visual, 3rd Edition.

Schwarz, H. & Wiegand, T. (2007). Overview of the Scalable Video Coding Extension of the H.264/AVC Standard, *IEEE Transactions on Circuits and Systems for Video Technology* 17(9): 1103–1120.

Shahid, Z., Chaumont, M. & Puech, W. (2009). An Adaptive Scan of High Frequency Subbands of Dyadic Intra Frame in MPEG4-AVC/H.264 Scalable Video Coding, *Proc. SPIE, Electronic Imaging, Visual Communications and Image Processing*, Vol. 7257, San Jose, CA, USA, p. 9.

Wang, Y., Wenger, S., Wen, J. & Katsaggelos, A. (2000). Error Resilient Video Coding Techniques, *IEEE Signal Processing Magazine* 17(4): 61 –82.

Wiegand, T., Sullivan, G., Richel, J., Schwartz, H., Wien, M. & eds. (April 2007). Joint Scalable Video Model (JSVM) 10, *JVT-W202*.

Wien, M., Schwarz, H. & Oelbaum, T. (2007). Performance Analysis of SVC, *IEEE Transactions on Circuits and Systems for Video Technology* 17(9): 1194 –1203.

Wu, M., Joyce, R. & Kung, S. (2000). Dynamic Resource Allocation via Video Content and Short-Term Traffic Statistics, *Proc. IEEE International Conference on Image Processing*, Vol. 3, pp. 58–61.