Scalable Video Coding Using Hybrid DCT/Wavelets Architectures

Tamer Shanableh

American University of Sharjah, College of Engineering, Sharjah, UAE
tshanableh@aus.edu

Abstract. This paper proposes the use of wavelet image transformation and polyphase downsampling in scalable video coding. A wavelet-based inter-frame coding solution using the syntax and framework of both MPEG-4 H.264/AVC and it is scalable extension, SVC. In the former codec, redundant slices are employed for coding the high frequency subbands of wavelet transformed imaged. While in the latter codec, the wavelet subbands are arranged into separate Coarse Grain Scalability (CGS) layers. Additionally, the paper proposes the use of a modified polyphase downsampling in applications of scalability and error resiliency. It is shown that the coding efficiency of the proposed solutions is comparable to single layer coding.

Keywords: Digital Video Coding, Scalable Video Coding, MPEG.

1 Introduction

It is reported in [1] that the DCT block-based approach is suitable for coding wavelet subbands. It was proposed to code the wavelet subbands in the base and enhancement layer of MPEG-4 AVC/H.264 scalable video coding (SVC) [2]. The low frequency band is coded in the base layer, the resultant quantization error and the high frequency bands are arranged into one image and coded in the enhancement layer. Such an approach allowed for both SNR and dyadic spatial scalabilities. Both the base and enhancement layers are coded using the AVC intra-frame syntax. This paper extends the reviewed work by proposing an inter-frame wavelet coding scheme in two different coding arrangements using the framework of both AVC [3] and SVC.

For inter-frame wavelet coding, the high frequency subbands are time-variant because of the decimation process involved in the image wavelet decomposition. Thus, translation motion in the pixel-domain image cannot be accurately estimated from the wavelet coefficients. Complete-to-overcomplete Discrete Wavelet Transformation (DWT) can be used to solve this problem. For instance, in [4] and [5] complete-to-overcomplete DWT is applied to the locally decoded reference subbands. As a result, each frequency subband ends up with 4 representations with different directions of unit shifts. Motion estimation is then used to find a best match location in one of the four reference representations. An extra syntactic field is needed to indicate the reference subband representation to which the MV belongs. Clearly the complete-to-overcomplete

C.K. Loo et al. (Eds.): ICONIP 2014, Part III, LNCS 8836, pp. 266–275, 2014.

DWT of the reference subbands and the extra syntactical field violates the operations of the standardized codecs. Moreover, the results presented in [5] applies the above complete-to-overcomplete DWT in conjunction with pixel-accurate ME only. A traditional method for complete-to-overcomplete DWT was introduced in [6]. A time domain one dimensional signal is passed through a high pass and low pass filter followed by decimation to produce low and high frequency subbands. The original signal is also shifted by one unit and the decomposition procedure is repeated. In general at each decomposition level, the low frequencies are decomposed twice, with and without unit shifting. More advanced complete-to-overcomplete DWT methods are reported in [7] and [8].

In this paper, two solutions are proposed for interframe coding of wavelet coefficients. The first solution employs the redundant pictures of the AVC framework for the coding of wavelet subbands, while in the second solution, the wavelet subbands are coded in the enhancement layers of a SVC codec.

The paper is organized as follows. Section 2 introduces the proposed solution of using redundant pictures for video scalability. Section 3 introduces the proposed solution of using wavelet subbands with scalable video coding. Section 4 introduces the proposed polyphase downsampling approach to scalability. The experimental results are introduced in Section 5 and Section 6 concludes the paper.

2 Proposed Redundant Pictures Approach to Scalability

The AVC standard introduced the use of redundant pictures (or redundant slices) as an error resiliency tool. The idea is to allow the encoder to repeat the coding of a primary picture (or part of it) in a redundant picture syntax element. In case of transmission errors the decoder can choose to decode the redundant picture to conceal the error and alleviate picture drift. This paper proposes the use of the redundant pictures for the coding the high frequency wavelet subbands. The low frequency subbands on the other hand are coded using the primary picture syntax element. Note that the AVC standard indicates that a compliant decoder does not have to decode redundant pictures. Therefore, the proposed coding arrangement does not violate the standard in this regard.

In this arrangement, if a video server streams the primary pictures only then a low spatial resolution of the original video is received. This is fully compliant with any AVC decoder. On the other hand, if the server streams both the primary and redundant pictures then a scalable decoder will be able to reconstruct the video at a high spatial resolution.

The first stage of this solution is a pre-process in which the input images are transformed into the wavelet domain. High frequency subbands are then rearranged and coded as redundant pictures. This is illustrated in Figure 1 below. The rearrangement of high frequency subbands is necessary to guarantee that similar subbands are predicted from each other thus increasing the efficiency of motion estimation and compensation.

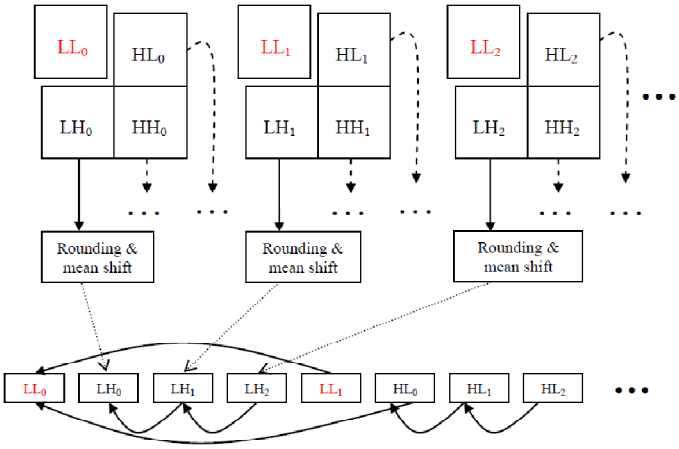

Fig. 1. Arrangement of wavelet subbands into primary and redundant pictures in the AVC framework

In the pre-processing it is also important to round and shift the mean of wavelet coefficients such that they can be represented with unsigned short data types. In this implementation and with one level of wavelet decomposition, the coefficients are represented by 10 bits only. The AVC implementation can be configured accordingly. Note that the rounding causes an imperfect reconstruction of the wavelet coefficients. Nevertheless it was noticed that loss in image quality is negligible. Empirically, the reconstructed rounded images have a PSNR of around 50 dB.

In the AVC coding stage, the standard specifies that primary pictures cannot be predicted from redundant ones. And a redundant picture cannot be predicted from its primary picture as well. Referring to Figure 1, clearly the prediction of say HL_0 (the subscript refers to the time index of the input image) from LL_0 is useless and the AVC coder will decide to perform an intra-frame coding instead. The rest of the high frequency subbands in this case i.e. HL_1 and HL_2 will be efficiently predicted from each other. Upon decoding, an extra post-process is required in which the decoded high frequency subbands and the decoded primary pictures are regrouped and inverse transformed into the higher spatial resolution.

3 Proposed Scalability Solution Based on Wavelet Subbands

In this proposed solution, the SVC scalable framework is used to encode both the low and high frequency subbands. The input images are DWT, rounded and mean shifted as in the aforementioned redundant pictures solution. The low frequency subband is coded as a base layer in this case. The high frequency subbands on the other hand are coded in separate SVC enhancement layers as illustrated in Figure 2 below.

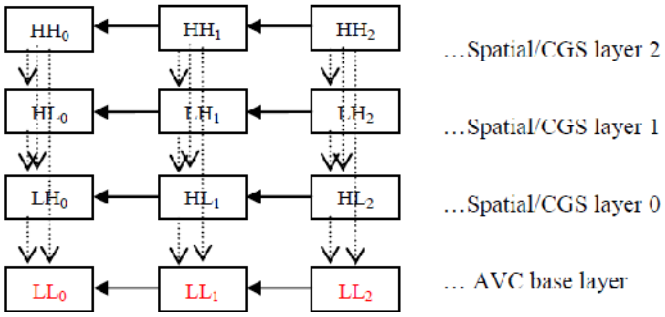

Fig. 2. Arrangements of wavelet subbands into 4 SVC layers

The SVC standard specifies that the spatial resolution of the enhancement layers can be greater than or equal to the spatial resolution of the base layer. In this case the upsamling filter of interlayer prediction is disabled and the deblocking of the base layer is omitted because the block boundaries between the layers are already aligned. Thus the arrangement of Figure 2 above is syntax friendly.

The prediction of high frequency subbands will naturally be intra-layer as opposed to inter-layer prediction. Nevertheless the vertical prediction lines in the figure indicate that other prediction modes can be applied. The SVC standard specifies a number of inter-layer predictions such as prediction of motion fields, prediction of MB partitioning, MB coding modes and so forth.

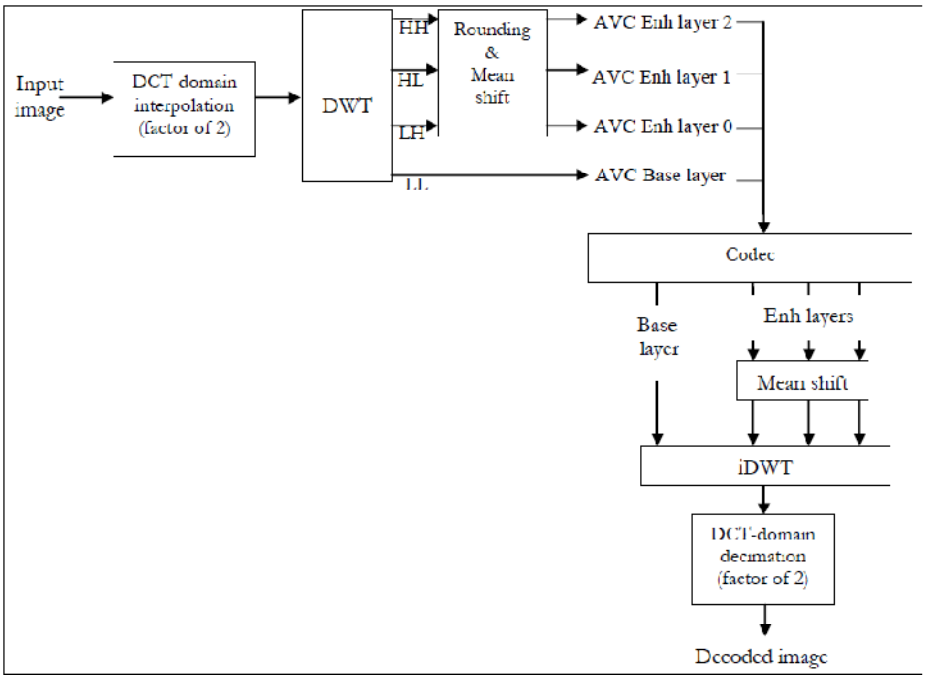

Fig. 3. Interpolation of input images prior to DWT and AVC coding

In comparison to the previous redundant picture solution, the perdition of high frequency subbands in the enhancement layers is continuous and does not suffer from the aforementioned problems where third of the redundant pictures have to be either intra coded or predicted from a primary picture which is the LL subband in this case. Hence more efficient coding is expected as illustrated in the experimental results section.

In this solution we also experiment with spatially interpolating the input images prior to DWT in an attempt to increase the correlation between the same frequency subbands across different images. The pre-processing, coding and post processing of such a system is illustrated in Figure 3 below.

On the other hand in an attempt to reduce the bitrate generated by the high frequency band, an opposite solution can be thought of. Such bands can be spatially decimated prior to coding. However this arrangement will in some cases affect the efficiency of motion estimation and compensation. The pre-processing, coding and post processing of such a system is illustrated in Figure 4 below.

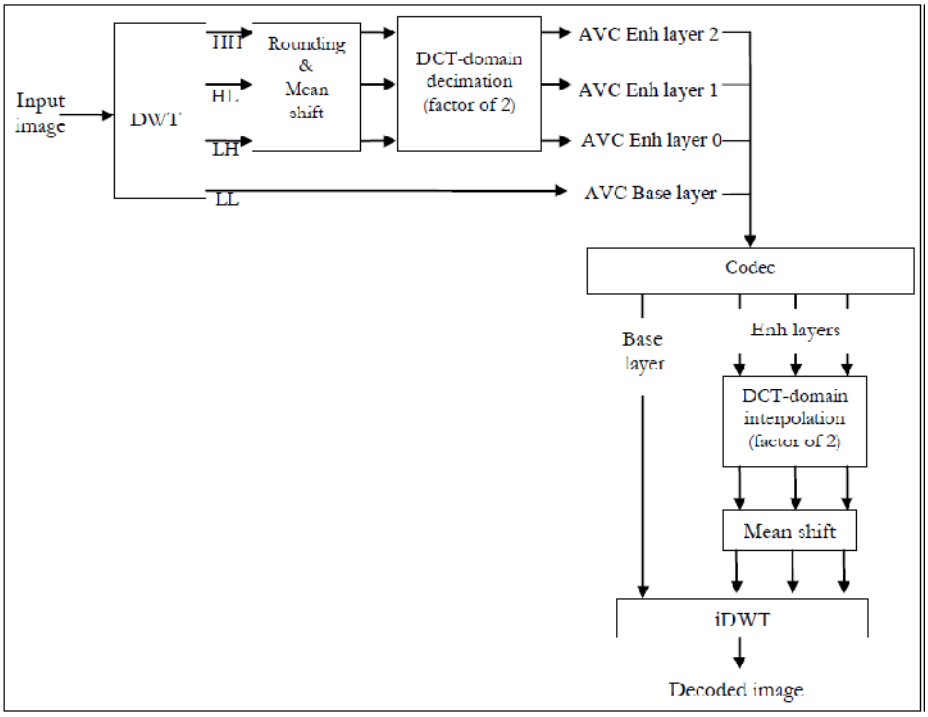

Fig. 4. Decimation of wavelet subbands prior to SVC coding

4 Proposed Polyphase Scalability Solution

One potential drawback of the proposed inter-frame wavelet solution suing the AVC redundant pictures is that it defeats the purpose of error resiliency. Thus one can think

of an alternative solution in which redundant pictures can be used for both error resiliency and spatial scalability. The solution is based on polyphase downsampling which is usually used in Multiple Description (MD) coding [9,10,11].

In [12] a source video is polyphase down sampled and fed into separate AVC coders. The paper then focuses on transmission errors and proposes different concealment solutions and post processing to attenuate visual effects related to MD coding and transmission errors.

In this work we propose the use of polyphase downsampling as a scalability and error resiliency tool. One of the polyphase down sampled images (or descriptors) is used as a primary image within the AVC framework and the rest of the descriptors are used as redundant pictures. The redundant pictures can serve as an error resiliency tool because their visual content is very similar to the primary pictures. Likewise the redundant pictures can be used for enhancing the spatial scalability of the primary pictures.

For completeness, the concept of polyphase down sampling is illustrated in the Figure 5.

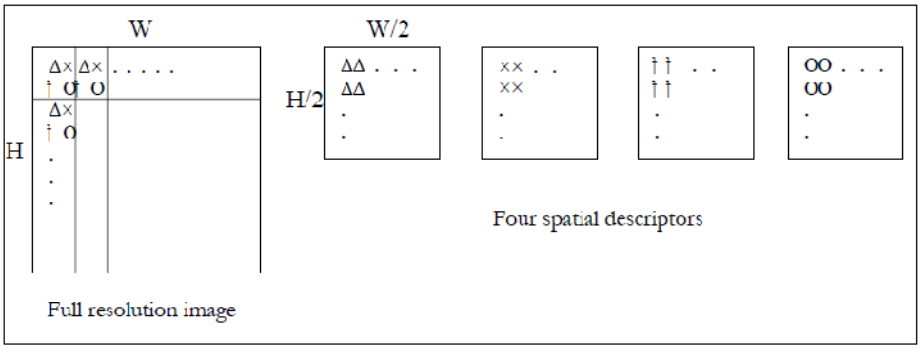

Fig. 5. Illustrating the concept of image polyphase downsampling.

A scalable solution based on such descriptors suffers from aliasing artifacts in the primary pictures (or base layer in this case) due to the lack of image filtering prior to down sampling. Hence this work proposes to replace the first descriptor (indicated by the 'Δ' samples) by the average of the four descriptors Δ, ×, 1 , O. This will provide a filtered and downsampled base layer which can be coded using the AVC primary pictures. Again, the rest of the descriptors are coded using redundant pictures. If all the descriptors are decoded then the original samples of the base layer descriptor can be recovered from the decoded average (in the primary pictures) and the '×', 'l' and 'O' samples decoded from the redundant pictures.

For an alternative approach for filtering, an adaptive average can be used based on localized edge detection. In this case the 'Δ' samples are averaged with a predictor '*y*' defined as:

$$\begin{aligned} y &= \max(\times, l) \quad \text{if } O \geq \max(\times, l) \text{ or} \\ y &= \min(\times, l) \quad \text{if } O \leq \min(\times, l) \text{ or} \\ y &= \times + l - O \text{ (otherwise)} \end{aligned} \tag{1}$$

Both methods of averaging and adaptive averaging generates similar results. However, it was noticed that the upsamling quality of the latter approach generated a higher PSNR (around 2 dB).

Figure 6 below visually shows the results of the proposed polyphase downsampling in comparison to the traditional approach. The aliasing artifacts on the background edges are evident in the descriptors generated from the '×', 'l' and 'O' samples. However such aliasing affects are greatly attenuated in the averaged descriptor.

Figure 7 illustrates that similar to the inter-frame wavelet solution, the descriptors can be arranged into primary and redundant pictures in the AVC framework following the arrangement illustrated in Figure 1above. Notice that similar descriptors are grouped into one redundant picture group thus rendering the motion compensation process more efficient.

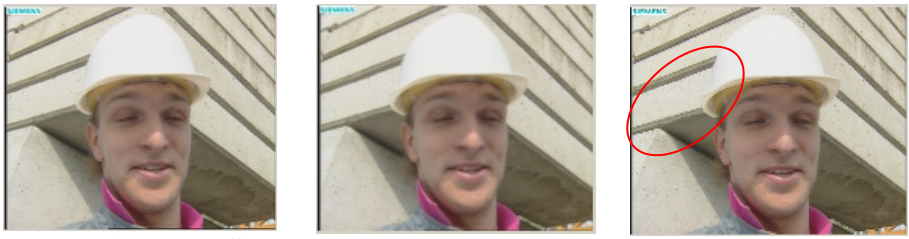

Fig. 6. Reducing aliasing artifacts in one of the polyphase image descriptors

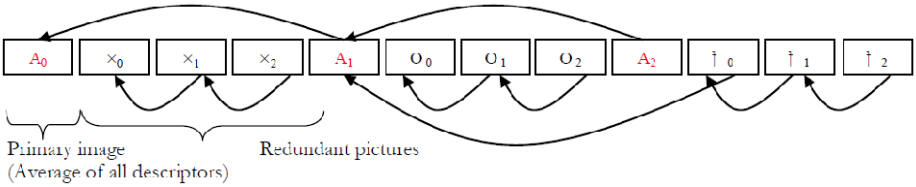

Fig. 7. Arrangement of image descriptors into primary and redundant pictures

5 Experimental Setup and Results

The experimental results used the following software; the JM reference software for (AVC) [13] and the JSVM reference software for SVC [14]. Both reference software are available on HHI institute, image and video coding website.

Figure 8 compares between the rate distortion curves of the proposed solutions against AVC single layer coding. Three test sequences are used; Crew and Harbour with a spatial resolution of 704x576 and IntoTree with a spatial resolutions of 1920x1080.

It is shown in the figure that in some cases the proposed inter-frame wavelet SVC solution outperforms single layer coding. In other cases, the proposed solution was slightly inferior to single layer coding. The figure also shows that the proposed SVC

solutions slightly outperform the inter-frame wavelet coding based on redundant pictures (In the figure this is referred to as 'Proposed RP. AVC'). As mentioned previously, this is due to the fact that a redundant picture preceded by a primary picture will not be inter-frame coded. Again such pictures count for third of the redundant pictures.

Figure 9 on the other hand presents the results using the interpolation and decimation ideas of Figures 3 and 4 above. As for decimating the high frequency subbands prior to coding, the figure shows that a gain in PSNR was achieved for the Crew but not the Harbour sequence. This can be justified as follows. The Crew sequence is less spatially active than Harbour thus, the coarse representation of high frequencies by means of decimation means that more bits can be allocated to the low frequency band and therefore enhancing the overall image quality. In contrast, coarsely representing the high frequencies of the spatially active Harbour sequence has a counter effect on image quality.

Moreover, the figure shows that implementing the interpolation solution of Figure 3 above the opposite effect is observed. The Harbour sequence benefited from such a solution and the overall PSNR was higher than the proposed SVC solution. In conclusion it seems that the use of the interpolation and decimation techniques should be adaptive according to the spatial activity of the image content.

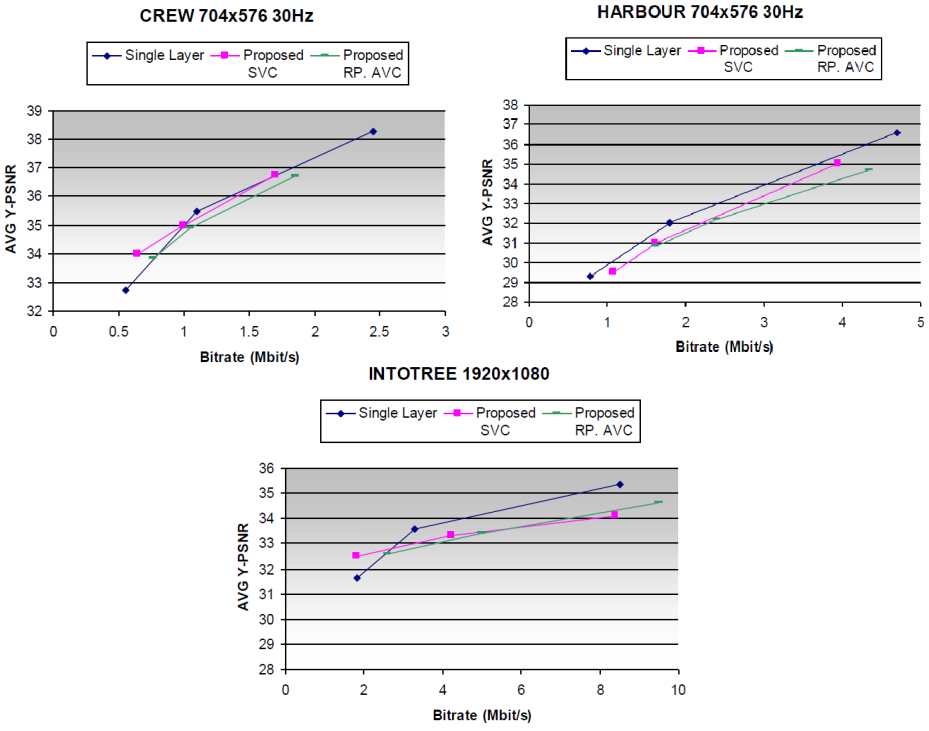

Fig. 8. Rate distortion curves for the proposed solutions in comparison to single layer coding

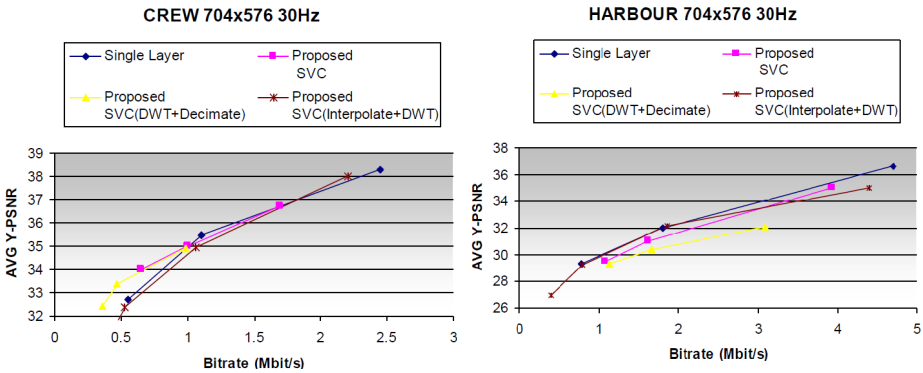

Fig. 9. Rate-distortion curves for the proposed interpolation and decimation solutions with inter-frame wavelet coding

6 Conclusion

This work proposed a number of novel arrangements for scalable video compression. It was proposed to high wavelet frequency subbands as either redundant pictures using AVC or scalable layers using SVC. It was shown that the latter provided higher prediction efficiency for coding the high frequency subbands. It was also shown that depending on the spatial activity of a given image the high frequency subbands can be decimated for bitrate reduction. On the other hand interpolating the images prior to DWT increased the correlation between subsequent subbands leading to higher prediction efficiency in sequences with high spatial activities. Lastly a framework for a solution based on modified polyphase downsampling was proposed. It is anticipated that such an approach can achieve both spatial scalability and error resiliency.

References

1. Hsiang, S.-T.: Intra-frame spatial scalability coding based on a subband/wavelet framework for MPEG-4 AVC/H.264 scalable video coding. In: ICIP 2007, San Antonio, Texas (September 2007)
2. Schwarz, H., Marpe, D., Wiegand, T.: 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 (2007)
3. Wiegand, T., Sullivan, G., Bjontegaard, G., Luthra, A.: Overview of the H.264/AVC video coding standard. IEEE Transactions on Circuits and Systems for Video Technology 13(7), 560–576 (2003)
4. Ohm, J.-R., Schaarb, M., Woods, J.: Interframe wavelet coding—motion picture representation for universal scalability. Signal Processing: Image Communication 19(9), 877–908 (2004)
5. Andreopoulos, Y., Schaar, M., Munteanu, A., Barbarien, J., Schelkens, P., Cornelis, J.: Fully scalable wavelet video coding using in-band motion-compensated temporal filtering. In: Proc. ICASSP 2003, Hong kong (April 2003)

6. Park, H.-W., Kim, H.-S.: Motion Estimation Using Low-Band-Shift Method for Wavelet-Based Moving Picture Coding. IEEE Trans. Image Processing 9(4), 577–587 (2000)
7. Andreopoulos, Y., Munteanu, A., Auwera, G., Schelkens, P., Cornelis, J.: A new method for complete-to-overcomplete discrete wavelet transforms. In: Proc. of 14th International Conference on Digital Signal Processing, vol. 2, pp. 501–504 (2002)
8. Li, X.: New results of phase shifting in the wavelet space. IEEE Signal Processing Letters 10(7), 193–195 (2003)
9. Caramma, M., Fumagalli, M., Lancini, R.: Polyphase down sampling Multiple Description Coding for IP Transmission. In: Proc. SPIE, vol. 4310, pp. 545–552 (December 2000)
10. Franchi, N., Fumagalli, M., Lancini, R., Tubaro, S.: A space domain approach for multiple description video coding. In: Proc. ICIP 2003, Barcelona, Spain (September 2003)
11. Gallant, M., Shiranit, S., Kossentiniq, F.: Standard-compliant multiple description video coding. In: ICIP 2001, Thessaloniki, Greece, vol. 1, pp. 946–949 (2001)
12. Bernardini, R., Durigon, M., Rinaldo, R., Celetto, L., Vitali, A.: Polyphase spatial subsampling multiple description coding of video streams with H264. In: Proc. ICIP 2004, Singapore, October, vol. 2, pp. 3213–3216 (2004)
13. JM reference software for H.264/MPEG-4 AVC is a video coding standard, HHI institute, `http://iphome.hhi.de/suehring/tml/`
14. JSVM reference software for the Scalable Video Coding (SVC), HHI institute, `http://www.hhi.fraunhofer.de/fields-of-competence/image-processing/research-groups/image-video-coding/svc-extension-of-h264avc/jsvm-reference-software.html`