ARCHITECTURES OF INCORPORATING MPEG-4 AVC INTO THREE-DIMENSIONAL WAVELET VIDEO CODING

Xiangyang Ji[*1], *Jizheng Xu*[2], *Debin Zhao*[1], *Feng Wu*[2]

[1]Institute of Computing Technology,
Chinese Academy of Sciences, Beijing, China
{xyji,dbzhao}@jdl.ac.cn
[2] Microsoft Research Asia, No 49,
Zhichun Rd, Haidian District, Beijing, China100080
{i-jzhxu, fengwu}@microsoft.com

ABSTRACT

3D wavelet video coding provides an elegant full-dimensional solution for scalable video coding. However, due to the Motion-Compensated Temporal Filtering (MCTF) with open-loop structure and no good rate distortion optimization trade-off between the texture coding and overhead coding of the corresponding mode and motion vectors within narrow bandwidth, its performance is unsatisfactory at low bit-rates. On the other hand, video data below a certain spatial, temporal or SNR resolution are not required in most applications of scalable video coding. The introduction of a base layer in three-dimensional scalable video coding scheme can provide a lot of advantages such as close-loop temporal prediction, in-loop filter and rate-distortion optimization without any constraint occurred in pure 3D wavelet codec and also compatibility with existing video compression standards. This paper presents the architectures about how to efficiently incorporate base layer codec MPEG-4 AVC into 3D wavelet video coding.

1. INTRODUCTION

To reliably deliver video to varying clients over heterogeneous networks using available system resources, particularly in scenarios with unknown system resources and network conditions in advance, the coded bit-stream should provide the temporal, spatial, SNR and complexity etc. scalabilities to meet the requirements of the clients with diverse display resolutions, bandwidths, computational capability, memory and QoS capabilities[1].

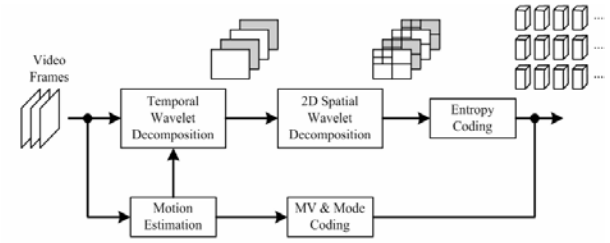

Figure 1. Three-dimensional wavelet video coding system

A generic three-dimensional wavelet video coding scheme [2] is illustrated in figure 1. The video signals are first decomposed by temporal subband transform and then the 2D spatial subband transform is applied to further decompose the signals. After the temporal and spatial transforms, the signals are organized into different spatio-temporal subbands, which correspond to different spatio-temporal resolution of the input video. With such nice property of subband transform, spatial scalability and temporal/frame-rate scalability can be supported very naturally. Moreover, with the bit-plane coding of the subband coefficients, quality/SNR scalability is also enabled. These scalabilities are very useful for the video transmission over error-prone networks, e.g, Internet or Wireless networks. Scalabilities also help different devices, which may have different memory, computation power, display size, accessible bandwidth, etc, to share the same compressed video. Recent results have showed that the three-dimensional subband video coding achieves compatible coding performance to the state-of-the-art non-scalable H.264 standard, at middle or high bit-rates.

Although wavelet-based codec can provide full-dimensional scalability with decent high coding efficiency within a wide range, at a certain bit-rate, it may not yield

* This work has been done while the author is with Microsoft Research Asia

the best performance. One reason is that the codec has to trade-off between high bit-rate performance and low bit-rate performance. So there is a contradiction between the scalability and performance at a certain channel condition. The performance loss due to supporting scalability may be noticeable at low bit-rates because of the narrow bandwidth constrain. So the performance at low bit-rates still needs to be improved. On the other hand, in most applications of scalable video coding, it is not required to access video data below a certain spatial, temporal or SNR resolution. This minimum spatial, temporal and SNR resolution represents a base layer on which additional enhancement layers are built to provide scalability up to a maximum spatial, temporal and SNR resolution. Given the above two reasons, introduction a base layer at low bit-rate is worthy. The introduction of a base layer in wavelet-based scalable video coding scheme can provide a lot of advantages such as close-loop temporal prediction, in-loop filter, rate-distortion optimizing without any constraint occurred in pure 3D wavelet codec. The other advantage of embedding a codec into 3D wavelet video coding is that it provides compatibility to other codec. For example, if a MPEG-I codec is embedded, while it may not improve the coding performance, it can guarantee that any receiver with MPEG-I decoding capability can decode a rough video. Due to a number of aforementioned advantages, [3] proposed core experiment on introduction of a base layer.

In the following section, we are focused on describing how other video codec as base layer can be embedded into 3D wavelet video coding to improve the coding performance at low bit-rates, while the advantages of wavelet video coding still can be kept and then the simulated results and conclusions are presented.

2. INCORPORATING THE BASE LAYER CODEC INTO 3D WAVELET CODING

Suppose that 4-level temporal subband transform is applied on a video with frame rate f frame per second, each level of temporal transform divides the input signals into two subbands, i.e. low-pass subband and high-pass subband. Then each subband has the half frame rate of the input signals. Let t-Li and t-Hi denote the output low-pass subband and high-pass subband after the i-th level temporal transform. Then both t-Li and t-Hi have frame rate $f/2^i$. On the decoder side, we can reconstruct the half frame rate video from t-Li only and the full frame rate video from both t-Li and t-Hi. The signal t-Li can be further decomposed with temporal subband transform to enable lower frame rate video that can be decoded by the decoder.

The codec for the base layer is embedded among temporal subband transforms. To determine the place to embed the codec, we should first decide the setting for the base layer video (e.g. frame rate and resolution). For simplifying the whole description, suppose that the frame rate and resolution of the lowest bit-rate point handled by the base layer are the quarter and half of the original video, respectively. There are two different implementation schemes shown as follows.

2.1. The Scheme 1

Figure 2 gives the block diagram of the first scheme. In figure 2, the t-L2 subband after down-sampling is coded using the codec embedded. The down-sampling and later up-sampling processes are the optional modules in the scheme. They exist only when the video coded by the embedded codec has lower resolution than the t-L2 subbands. So for the base layer, the input video is of $f/4$ frame per second, and the resolution is also reduced with the down-sampling process. Basically, the base layer video provides a low frame rate and low resolution representation of the original video. The dashed rectangle in Figure 2 represents the operations related to the base layer codec after two-level temporal decomposition. The low-pass subbands ts_LL coded in the base layer are generated after down-sampling t-L2 by a given low-pass filter. For these ts_LL subbands as the input video, the base layer codec executes the encoding process on them and gets the corresponding reconstructed subbands ts_LL'. And then, the low-quality reconstructed subbands t-L2' with the same resolution as t_L2 can be generated by up-sampling the ts_LL' subbands from the base layer codec. The further two-level temporal decomposition will be conducted at the subbands ts_L2" obtained by subtracting the low-quality reconstructed subbands t-L2' from the original t_L2 subband. At last, the spatial transform and entropy coding will be executed on every subband to generate the scalable bit-stream. Embedding the codec after the t-L2 subband is only an example. We can embed the codec after the t-L1 subband or the t-L2 subband. It depends on which frame rate is required by applications at the base layer video.

In the exemplified decoder shown as in figure 3, the video with the lowest reconstructed frame rate and spatial resolution can be directly decoded from the base layer decoder. Otherwise, after the entropy decoding, inverse spatial decomposition and the third and fourth inverse temporal decomposition are conducted, the decoded subbands t_L2" are added with the corresponding subbands t-L2' upsampled from the decoded ts_LL' subbands from the base layer codec to generate the low-pass signal t-L2 for the rest inverse temporal decomposition.

Furthermore, for three-dimensional subband coding with embedded base layer codec, the motion information for high-pass subband at the further temporal decomposition can be generated by pure three-dimensional wavelet coding. However, due to still high temporal redundancy on the differences between the original low-pass subband and up-sampled subband from the reconstructed ts_LL subbands by base layer codec, the corresponding motion information also can be achieved by conducting the motion estimation on the differences.

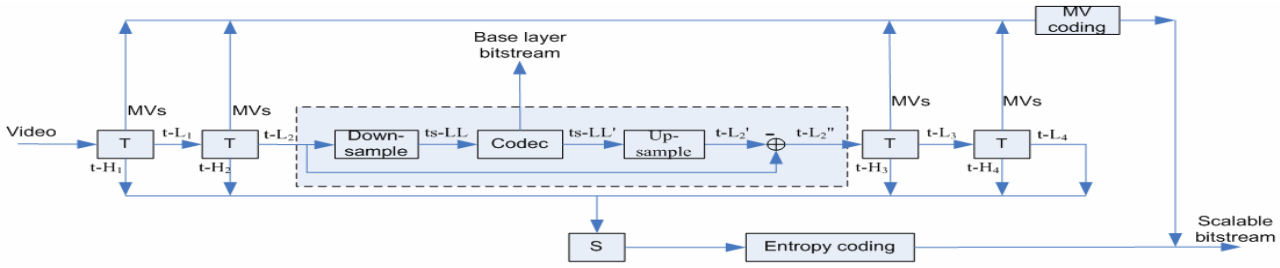

Figure 2: Three-dimensional subband video encoder w/ a codec embedded for scheme 1. T and S stand for temporal subband transform and spatial subband transform respectively.

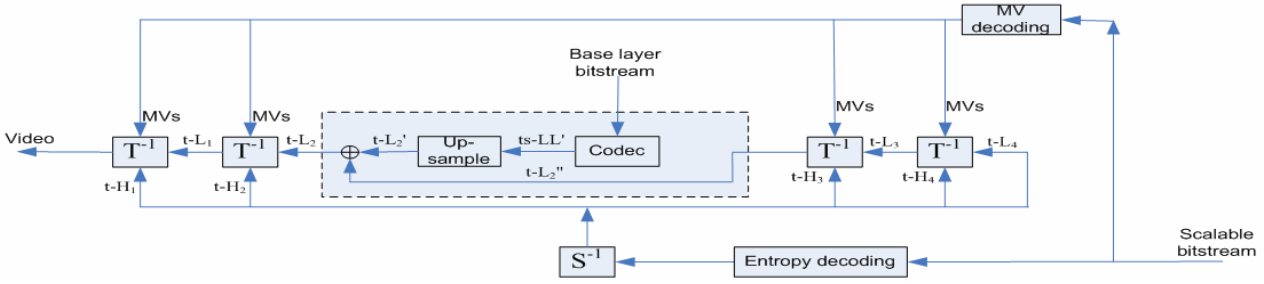

Figure 3: Three-dimensional subband video decoder w/ a codec embedded for scheme 1. T^{-1} and S^{-1} stand for inverse temporal subband transform and inverse spatial subband transform respectively.

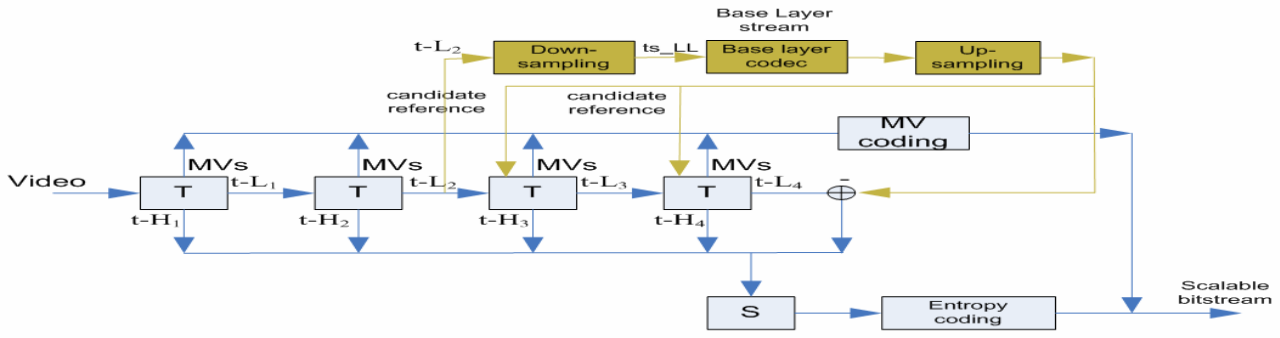

Figure 4. Three-dimensional subband video decoder with codec embedded for scheme 2.

2.2 The Scheme 2

For the above scheme, the basic idea is that the base layer codec is embedded in the temporal transforms. The signal further decomposed in the late temporal transforms is the residual one. But for this scheme, the further temporal transforms are still performed at the low-frequency frames of pure three-dimensional wavelet coding. Figure 4 depicts the block diagram of the second scheme. The input video for the base layer codec comes from the 3D subband coding scheme after the second temporal decomposition. The upper part of figure 4 is the base layer process. The input video is converted to the desired resolution with the down-sampling module. And then, the base layer codec will execute the encoding process on them to generate the base layer bit-stream of the lowest bit-rate.

For the further temporal transform in 3D subband coding, the high-pass subbands t_H3, t_H4 can use the temporally corresponding pictures, which are achieved by up-sampling the reconstructed picture from the base layer codec, as an additional candidate reference. The different reference combination cases of the temporal low-pass sub-

band in three-dimensional subband coding and/or the up-sampled version of the temporally corresponding reconstructed picture from the base layer can be used for the prediction and update step in 3D wavelet scalable coding. If the reference from base layer is used for t_H3, t_H4 in prediction step, the corresponding update step is not needed. Moreover, after the last temporal decomposition, the temporally corresponding up-sampled pictures from the base layer are always subtracted from the final low-pass subbands.

3. EXPERIMENTAL RESULTS

The proposed method is implemented based on core experiment reference software from MSRA and the MPEG-4 AVC reference software JM76 [4]. The test sequences in the CIF format include foreman, bus, mobile, and football with 30fps. Table 1 lists the detailed test conditions. The unit of bit-rate is kbits/s.

Table 1: Spatio-SNR-temporal scalability test conditions

Test sequences	QCIF@7.5Hz	QCIF@15Hz	CIF@15Hz	CIF@30Hz
BUS	64	96	192	512
			384	
FOOTBALL	128	192	384	1024
			512	
FOREMAN	32	48	96	256
			192	
MOBILE	48	64	128	384
			256	

In figure 2, the mpeg filter [5] and Daubechies 9/7 DWT are used for down-sampling module. During the up-sampling process, we adopt the three different schemes: AVC 6-tap filter (-1, 5, 20, 20, 5, -1), sinc 8-tap filter {-0.0105F, 0.0465F, -0.1525F, 0.6165F, 0.6165F, -0.1525F, 0.0465F, -0.0105F} [6] and Daubechies 9/7 IDWT (Inverse Discrete Wavelet Transform).

Table 2 shows the PSNR comparisons for the scheme 1 when mpeg filter is used for down-sampling and AVC filter, sinc filter and IDWT are used for up-sampling, respectively. MPEG_SINC represents the mpeg filter for down-sampling and sinc filter for up-sampling and the rest may be deduced by analogy. Wherein, MPEG_IDWT scheme is the worse and MPEG_SINC slightly outperform MPEG_AVC. Table 3 shows the PSNR comparisons for the scheme 1 when DWT is used for down-sampling and AVC filter, sinc filter and IDWT are used for up-sampling, respectively. Wherein, DWT_IDWT scheme is the worse and DWT_SINC slightly outperform DWT_AVC. Here, all the lowest bit-rate points are handled by base layer codec MPEG-4 AVC.

For calculating the PSNR of lowest bit-rate point, the references with QCIF resolution for table 2 is given by down-sampling with mpeg filter and the references for table 3 is achieved by down-sampling with 9/7 DWT. Although the PSNR in table 2 is far higher than the one in table 3 for each test sequence, the scheme with DWT_SINC whose down-sampling is implemented by DWT operation can keep the more high-frequency information compared to the scheme with MPEG_SINC. As shown in figure 5, for the lowest bit-rate point at 7.5 HZ frame rate and 48 kbps for mobile sequence in QCIF format, frame (a) reconstructed by base layer codec with DWT_SINC scheme can provide the better visual quality compared to frame (b) reconstructed frame by base layer codec with the MPEG_SINC scheme.

Table 4 provides average PSNR comparisons between the first scheme with DWT_SINC and 3D wavelet video coding at the test conditions shown in table 1. At low bit-rate, compared to pure 3D wavelet video coding scheme, 3D wavelet scheme with base layer embedded can gain 1.3-2.1 db. We also can see that at high bit-rate, the 3D wavelet scheme approximate to or outperform our scheme due to the overhead coding of modes and motion vectors, bit-plane segmentation entropy coding for enchantment layer, and motion align mismatch between base layer and further temporal decomposition.

To get rid of such disadvantage introduced by base layer codec, we can take the lowest bit-rate point as the independent part which will be handled by base layer codec. In table 5, the PSNR comparisons among pure 3D wavelet scheme, scheme 1 and scheme 2 show the scheme 2 can obviously outperform scheme 1 at high bit rate except for football sequence. For football, the scheme 1 can get better performance compared to scheme 2 because of small and irregular motion. The results reveal that scheme 2 outperforms the scheme 1 when one point with low bit-rate and the other points with high bit-rate are coded.

All above results are achieved by executing motion estimation at further temporal decomposition in pure 3D wavelet video coding (ME_On_Org). To find a better solution to efficiently exploit the temporal redundancy, the motion estimation based on difference (ME_On_Diff) is also tested. Table 6 shows PSNR comparisons between motion estimation on ME_On_Diff and ME_On_Org based on DWT_DWT scheme. The motion estimation with ME_On_Diff at further temporal decomposition always can get better performance compared to motion estimation with ME_On_Org. Table 7 shows PSNR comparisons between motion estimation on ME_On_Diff and ME_On_Org based on DWT_SINC scheme. The motion estimation with ME_On_Diff at further temporal decomposition always can get better performance compared to motion estimation with ME_On_Org except for bus sequence.

Table 2: PSNR comparisons for the first scheme when mpeg filter is used for down-sampling and AVC filter, sinc filter and IDWT are used for up-sampling, respectively.

	PSNR_Y	PSNR_U	PSNR_V	PSNR_Y	PSNR_U	PSNR_V	PSNR_Y	PSNR_U	PSNR_V		PSNR_Y	PSNR_U	PSNR_V	PSNR_Y	PSNR_U	PSNR_V	PSNR_Y	PSNR_U	PSNR_V
Rate	Bus									Rate	Foreman								
	MPEG_SINC			MPEG_IDWT			MPEG_AVC				MPEG_SINC			MPEG_IDWT			MPEG_AVC		
64	29.83	38.37	39.27	29.83	38.37	39.27	29.83	38.37	39.27	32	32.88	38.72	39.32	32.88	38.72	39.32	32.88	38.72	39.32
192	26.761	36.501	37.698	26.587	36.43	37.642	26.698	36.503	37.69	96	30.728	37.325	37.982	30.641	37.326	38.023	30.712	37.312	37.986
512	30.486	38.299	40.148	30.324	38.224	40.035	30.456	38.259	40.112	256	33.943	39.385	40.67	33.836	39.329	40.582	33.921	39.374	40.659
Rate	Football									Rate	Mobile								
	MPEG_SINC			MPEG_IDWT			MPEG_AVC				MPEG_SINC			MPEG_IDWT			MPEG_AVC		
128	33.35	37.55	39.56	33.35	37.55	39.56	33.35	37.55	39.56	48	27.67	31.33	30.96	27.67	31.33	30.96	27.67	31.33	30.96
384	30.837	35.413	37.993	30.741	35.31	37.944	30.823	35.405	37.989	128	23.361	28.018	27.442	23.287	28.046	27.291	23.343	28.026	27.42
1024	33.901	37.721	39.809	33.823	37.658	39.765	33.884	37.724	39.794	384	28.326	32.336	31.713	28.183	32.118	31.554	28.299	32.191	31.759

Table 3: PSNR comparisons for the first scheme when DWT is used for down-sampling and AVC filter, sinc filter and IDWT are used for up-sampling, respectively.

	PSNR_Y	PSNR_U	PSNR_V	PSNR_Y	PSNR_U	PSNR_V	PSNR_Y	PSNR_U	PSNR_V		PSNR_Y	PSNR_U	PSNR_V	PSNR_Y	PSNR_U	PSNR_V	PSNR_Y	PSNR_U	PSNR_V
Rate	Bus									Rate	Foreman								
	DWT_SINC			DWT_IDWT			DWT_AVC				DWT_SINC			DWT_IDWT			DWT_AVC		
64	27.539	37.081	38.102	27.539	37.081	38.102	27.539	37.081	38.102	32	31.481	37.795	38.313	31.481	37.795	38.313	31.481	37.795	38.313
192	26.691	36.206	37.527	26.538	36.157	37.495	26.673	36.234	37.521	96	30.944	37.357	38.092	30.864	37.369	38.058	30.92	37.378	38.053
512	30.304	38.356	40.067	30.087	38.312	39.819	30.285	38.303	40.054	256	33.952	39.439	40.613	33.795	39.402	40.55	33.919	39.495	40.665
Rate	Football									Rate	Mobile								
	DWT_SINC			DWT_IDWT			DWT_AVC				DWT_SINC			DWT_IDWT			DWT_AVC		
128	32.226	36.647	38.457	32.226	36.647	38.457	32.226	36.647	38.457	48	24.369	29.341	28.892	24.369	29.341	28.892	24.369	29.341	28.892
384	31.314	35.75	38.056	31.263	35.76	38.008	31.316	35.762	38.036	128	23.598	28.416	27.837	23.486	28.398	27.837	23.583	28.391	27.839
1024	34.082	37.852	39.873	34.012	37.791	39.833	34.072	37.864	39.875	384	28.022	32.127	31.528	27.78	31.964	31.244	28.005	32.078	31.432

(a) (b)

Figure 5: Visual quality comparisons of frame 0 for the lowest bit-rate point at 7.5 HZ frame rate and 48 kbps for mobile sequence in QCIF format between (a) reconstructed frame by base DWT_SINC scheme base layer codec and (b) reconstructed frame by base MPEG_SINC scheme base layer codec

Table 4: PSNR comparisons between the scheme 1 with DWT_SINC and 3D wavelet coding

	PSNR_Y	PSNR_U	PSNR_V	PSNR_Y	PSNR_U	PSNR_V		PSNR_Y	PSNR_U	PSNR_V	PSNR_Y	PSNR_U	PSNR_V
Rate	Bus						Rate	Football					
	pure 3D wavelet			scheme1				pure 3D wavelet			scheme1		
64	26.118	36.326	36.714	27.539	37.081	38.102	128	30.581	34.577	36.893	32.226	36.647	38.457
96	26.383	36.575	37.44	26.965	37.041	38.106	192	29.68	33.893	36.688	30.781	35.507	37.783
192	27.599	37.051	37.774	26.691	36.206	37.527	384	31.271	35.177	37.432	31.314	35.75	38.056
384	30.596	38.715	40.144	29.747	38.128	39.431	512	32.666	36.274	38.432	32.563	36.755	38.769
512	30.991	39.161	40.52	30.304	38.356	40.067	1024	34.163	37.681	39.67	34.082	37.852	39.873
Rate	Foreman						Rate	Mobile					
	pure 3D wavelet			scheme1				pure 3D wavelet			scheme1		
32	29.366	36.569	35.53	31.481	37.795	38.313	48	22.889	26.812	25.957	24.369	29.341	28.892
48	29.773	36.63	37.196	31.28	37.765	38.273	64	23.33	27.414	26.839	24.085	29.351	28.919
96	31.027	37.448	38.092	30.944	37.357	38.092	128	23.847	28.656	28.039	23.598	28.416	27.837
192	33.755	39.278	40.366	33.385	39.093	40.211	256	27.029	31.4	30.789	26.535	30.6	29.841
256	34.329	39.694	40.916	33.952	39.439	40.613	384	28.642	32.791	32.399	28.022	32.127	31.528

Table 5: PSNR comparisons of pure 3D wavelet scheme, scheme 1 and scheme 2

	PSNR Y	PSNR U	PSNR V	PSNR Y	PSNR U	PSNR V	PSNR Y	PSNR U	PSNR V		PSNR Y	PSNR U	PSNR V	PSNR Y	PSNR U	PSNR V	PSNR Y	PSNR U	PSNR V
Rate	Bus									Rate	Foreman								
	pure 3D wavelet			scheme 2			scheme 1				pure 3D wavelet			scheme 2			scheme 1		
64	26.118	36.33	36.714	27.539	37.081	38.102	27.539	37.081	38.102	32	29.366	36.569	35.53	31.481	37.795	38.313	31.481	37.795	38.313
192	27.599	37.05	37.774	26.721	36.037	37.355	26.691	36.206	37.527	96	31.027	37.448	38.092	30.724	37.042	37.448	30.944	37.357	38.092
384	30.596	38.72	40.144	30.082	38.257	39.494	29.747	38.128	39.431	192	33.755	39.278	40.366	33.523	38.915	40.049	33.385	39.093	40.211
512	30.991	39.16	40.52	30.627	38.41	40.197	30.304	38.356	40.067	256	34.329	39.694	40.916	34.111	39.5	40.748	33.952	39.439	40.613
Rate	Football									Rate	Mobile								
	pure 3D wavelet			scheme 2			scheme 1				pure 3D wavelet			scheme 2			scheme 1		
128	30.581	34.58	36.893	32.226	36.647	38.457	32.226	36.647	38.457	48	22.889	26.812	25.957	24.369	29.341	28.892	24.369	29.341	28.892
384	31.271	35.18	37.432	30.07	34.149	36.757	31.314	35.75	38.056	128	23.847	28.656	28.039	23.794	28.481	27.819	23.598	28.416	27.837
512	32.666	36.27	38.432	31.624	35.418	37.591	32.563	36.755	38.769	256	27.029	31.4	30.789	26.941	30.687	30.179	26.535	30.6	29.841
1024	34.163	37.68	39.67	33.619	37.255	39.385	34.082	37.852	39.873	384	28.642	32.791	32.399	28.537	32.594	32.122	28.022	32.127	31.528

Table 6: PSNR comparisons between motion estimation on ME_On_Diff and ME_On_Org based on DWT_DWT scheme

	PSNR_Y	PSNR_U	PSNR_V	PSNR_Y	PSNR_U	PSNR_V		PSNR_Y	PSNR_U	PSNR_V	PSNR_Y	PSNR_U	PSNR_V
Rate	Bus						Rate	Foreman					
	ME_On_Diff			ME_On_Org				ME_On_Diff			ME_On_Org		
64	27.539	37.081	38.102	27.539	37.081	38.102	32	31.481	37.795	38.313	31.481	37.795	38.313
192	26.768	36.21	37.485	26.538	36.157	37.495	96	31.05	37.352	38.142	30.864	37.369	38.058
512	30.186	38.187	39.865	30.087	38.312	39.819	256	33.806	39.337	40.528	33.795	39.402	40.55
Rate	Football						Rate	Mobile					
	ME_On_Diff			ME_On_Org				ME_On_Diff			ME_On_Org		
128	32.226	36.647	38.457	32.226	36.647	38.457	48	24.369	29.341	28.892	24.369	29.341	28.892
384	31.401	35.754	38.021	31.263	35.76	38.008	128	23.611	28.448	27.856	23.486	28.398	27.837
1024	34.062	37.829	39.859	34.012	37.791	39.833	384	27.814	31.988	31.275	27.78	31.964	31.244

Table 7: PSNR comparisons between motion estimation on ME_On_Diff and ME_On_Org based on DWT_SINC scheme

	PSNR_Y	PSNR_U	PSNR_V	PSNR_Y	PSNR_U	PSNR_V		PSNR_Y	PSNR_U	PSNR_V	PSNR_Y	PSNR_U	PSNR_V
Rate	Bus						Rate	Foreman					
	ME_On_Diff			ME_On_Org				ME_On_Diff			ME_On_Org		
64	27.539	37.081	38.102	27.539	37.081	38.102	32	31.481	37.795	38.313	31.481	37.795	38.313
192	25.645	36.07	37.151	26.691	36.206	37.527	96	31.089	37.424	38.17	30.944	37.357	38.092
512	30.249	38.307	39.785	30.304	38.356	40.067	256	33.92	39.461	40.581	33.952	39.439	40.613
Rate	Football						Rate	Mobile					
	ME_On_Diff			ME_On_Org				ME_On_Diff			ME_On_Org		
128	32.226	36.647	38.457	32.226	36.647	38.457	48	24.369	29.341	28.892	24.369	29.341	28.892
384	31.404	35.775	38.019	31.314	35.75	38.056	128	23.686	28.398	27.845	23.598	28.416	27.837
1024	34.11	37.868	39.852	34.082	37.852	39.873	384	28.045	32.07	31.524	28.022	32.127	31.528

4. CONCLUSIONS

This paper provides two architectures of incorporating base layer codec MPEG-4 AVC into three-dimensional wavelet video coding. The base layer codec in scheme 1 is embedded in the temporal transforms. The base layer codec in scheme 2 is an independent module and the further temporal transforms are still performed at the low-frequency frames of pure three-dimensional wavelet coding. Both of them can efficiently improve the coding performance of video signal at low bit-rate. The different down-sampling and up-sampling combination cases for scheme 1 are tested. Corresponding to scheme 1, in most case, scheme 2 can provide the better performance at high bit-rate when only one point with low bit-rate and the other points with high bit-rate are coded. Furthermore, for scheme1, the experimental results verify the motion estimation on difference at the further temporal decomposition is more effective than the motion estimation by pure three-dimensional wavelet coding.

5. REFERENCES

[1] ISO/IEC JTC1, "Requirements and Applications for Scalable Video Coding," ISO/IEC JTC1/WG11 Doc. N6025, Oct. 2003.

[2] Jizheng Xu, Ruiqin Xiong, Bo Feng, Gary Sullivan, Ming-Chieh Lee, Feng Wu, Shipeng Li, "3D Sub-band Video Coding using Barbell lifting," ISO/IEC JTC1/SC29/WG11 MPEG2004/M10569/S05, Munich, March 2004.

[3] "Description of Core Experiments in SVC," ISO/IEC JTC 1/SC 29/WG11 MPEG2004/N6373

[4] JVT Reference Software, Version JM76, http://bs.hhi.de/~suehring/tml/download/

[5] Gisle Bjøntegaard and Karl Olav Lillevo, "H.263 Anchors - Technical Description". MPEG95/0322, Dallas, November 1995.

[6] P. Chen, J. W. Woods, "Improved MC-EZBC with quarter-pixel motion vectors, ISO/IEC JTC1/SC29/WG11", MPEG2002/m8366, Fairfax, May 2002.