SPECIAL ISSUE

Fast adaptive termination mode selection for H.264 scalable video coding

Jianfeng Ren · Nasser Kehtarnavaz

Abstract Although scalable video coding can achieve coding efficiencies comparable with single layer video coding, its computational complexity is higher due to its additional inter-layer prediction process. This paper presents a fast adaptive termination algorithm for mode selection to increase its computation speed while attempting to maintain its coding efficiency. The developed algorithm consists of the following three main steps which are applied not only to the enhancement layer but also to the base layer: a prediction step based on neighboring macroblocks, a first round check step, and a second round check step or refinement if failure occurs during the first round check. Comparison results with the existing algorithms are provided. The results obtained on various video sequences show that the introduced algorithm achieves about one-third reduction in the computation speed while generating more or less the same video quality.

Keywords Scalable video coding · Mode selection · Fast adaptive termination · H.264

1 Introduction

Scalable video coding (SVC) has been addressed in video coding standards [1, 2]. However, it is not widely used commercially due to its limited coding efficiency as compared to non-scalable video coding and also due to its high computational demand. Scalable video coding as an extension of the H.264/AVC standard allows the rate-distortion performance to come close to the state-of-the-art single layer coding. This improvement in coding efficiency is attributed to the use of hierarchical prediction structures, the inter-layer prediction mechanisms as well as features such as variable block size motion compensation with additional intra motion compensation. In order to gain the best rate-distortion performance in scalable video coding, an exhaustive search among inter-layer prediction, variable block inter prediction and intra prediction is done. However, such an approach results in high computational demand on the video encoding side, in particular for portable devices with limited CPU power and memory resources. This paper is concerned with the mode selection aspect of SVC, which is regarded as a computationally demanding part.

Fast mode selection has been extensively studied for single layer H.264/AVC video coding [3–6], utilizing coding information around neighboring blocks or texture information of current blocks. However, for SVC which includes additional inter-layer prediction, only a few techniques [7, 8] have been introduced for speeding up the mode selection in H.264/AVC SVC. In [7], a low complexity macroblock mode decision algorithm is introduced by combining CGS (coarse-gain-SNR) and temporal scalability. Depending on the macroblock coding modes and quantization parameters of the base layer, a look-up table is used recursively to determine the MB (macroblock) modes to be tested at the enhancement layer. The key point of this algorithm is the establishment of a robust look-up table between the base and the enhancement layers. In [8], it is shown that there is a relationship between the mode distribution of the base and the enhancement layers. Three different approaches are then used to speed up the mode selection. These approaches consist of a selective intra-

J. Ren (✉) · N. Kehtarnavaz
Department of Electrical Engineering, University of Texas at
Dallas, Richardson, TX, USA
e-mail: jxr041100@utdallas.edu

mode decision, a selective reduction of candidate modes and a temporal scalability.

Although the existing algorithms allow achieving time reduction in the SVC process, the gain in the computational complexity is limited to the enhancement layer with no consideration of the base layer. In this paper, we introduce a fast mode selection method which can be applied not only to the base layer but also to the enhancement layer. It is shown that the introduced algorithm is computationally simpler compared to the above existing algorithms.

The paper is organized as follows: Sect. 2 provide an overview of inter-layer video coding for SVC and covers the relationship between a current block and neighboring blocks. Our introduced fast mode selection algorithm is presented in Sect. 3. Sect. 4 includes the experimental results for spatial and quality scalable video coding. Finally, the conclusions are stated in Sect. 5.

2 Overview of scalable video coding

To set the stage for our introduced method, in this section, we provide an overview of the additional inter-layer prediction that is done in SVC to provide coding efficiency compared to single layer video coding. The statistical relationship between current and neighboring blocks is then shown for scalable video coding.

2.1 Inter-layer prediction in SVC

Scalable video coding supports different scalable resolutions including spatial, quality, and temporal. Temporal scalability is obtained using the concept of hierarchical B pictures. Figure 1 shows a representation of hierarchical B pictures with GOP (group of pictures) size of 16. Spatial scalability means each layer has a different spatial resolution. Quality scalability is regarded as a special case of spatial scalability with an identical resolution. For each spatial/quality layer, similar to the base H.264/AVC, a macroblock is reconstructed from spatial neighboring blocks using intra prediction or temporal blocks using
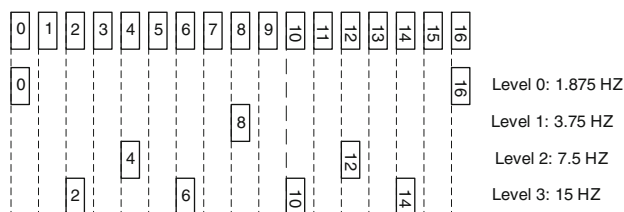
motion estimation. In order to improve coding efficiency in comparison to simulcasting different spatial/quality resolutions, an additional prediction step, referred to as inter-layer prediction, is applied to reduce the inter-layer redundancy. Figure 2 illustrates the H.264/AVC SVC and the inter-layer prediction step.

For the inter-layer motion prediction, shown as arrow "A" in Fig. 2, the motion vector, mode partition and reference frame index from the base layer are used in the enhancement layer. If the lower layer contains a down-scaled version of the current layer, the up-scaling is applied for motion vector and mode partition. If a motion vector refinement is applied, additional information including motion vector refinement are transmitted for decoding purposes.

For the inter-layer residue prediction, indicated by arrow "B" in Fig. 2, the residues from the base layer are used for the prediction of residues of the current layer. Thus, only the difference is DCT transformed, quantized, and entropy coded. In case of spatial scalability, each corresponding 8×8 block residue is up-sampled by using a bilinear filter for the residue prediction of the enhancement layer.
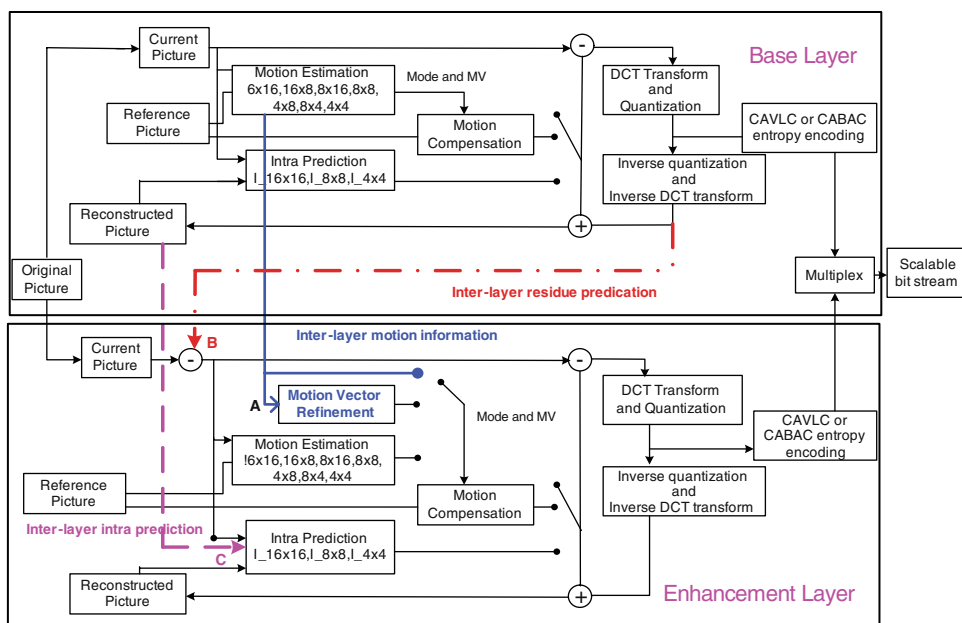
For the inter-layer intra prediction, indicated by arrow "C" in Fig. 2, the reconstructed blocks in the base layer are used for the prediction of a current macroblock in the enhancement layer. In case of spatial scalability, for the luma component, a one-dimensional 4-tap filter is applied horizontally and vertically. If the neighboring blocks are not intra coded, samples are generated through border extension.

It is time consuming to find the best prediction for a current macroblock in terms of rate-distortion efficiency due to the following reasons: (1) the measurement for the best prediction uses the rate-distortion cost, which requires the computations of SSD (sum of squared differences) and rate; (2) a macroblock partition for intra prediction can be 16×16, 8×8, and 4×4; (3) a macroblock partition for inter prediction can be 16×16, 16×8, 8×16, 8×8, 8×4, 4×8, 4×4; and (4) inter-layer prediction can be of inter-layer motion prediction, inter-layer residue prediction and inter-layer intra prediction from a lower layer.

2.2 Statistical relationship in SVC

In SVC, a current macroblock always has similar coding characteristics with neighboring blocks, such as mode information, motion vector, and reference frames. A number of fast motion estimation algorithms [9] have been previously introduced using motion vectors and reference frames of neighboring blocks. Here, the idea of the similarity between neighboring blocks and a current block is utilized for mode selection not only at the base layer but also at the enhancement layer of SVC. In H.264, the mode

Fig. 1 Temporal scalability with GOP size of 16 and output rate of 15 Hz

Level 0: 1.875 HZ

Level 1: 3.75 HZ

Level 2: 7.5 HZ

Level 3: 15 HZ

Fig. 2 Inter-layer prediction in H.264/AVC scalable video coding

is selected as the minimum rate-distortion cost by an exhaustive examination of all the modes.

If the mode of a current macroblock belongs to one of the modes in the neighboring macroblocks and its rate-distortion cost is less than an adaptive threshold (to be discussed in Sect. 3), which is based on the rate distortion of the neighboring blocks, the current macroblock is considered to have a similar statistical relationship with the neighboring macroblocks. This similarity is used here to design our fast mode selection algorithm discussed in Sect. 3.

At this point, it is useful to provide the outcome of some initial experiments that we carried out on several video sequences in order to highlight the similarity of the statistical relationship between a current macroblock and its neighboring blocks. For these experiments, the video coding parameters used were as follows: (1) 50 frames encoded, and (2) three quantizations used, QP = 24, 29, 34. The mode information and rate-distortion cost for each macroblock were recorded to provide an analysis between a current macroblock and its neighboring blocks including the collocated macroblock at the base layer for macroblocks of the enhancement layer. Table 1 shows the percentage of current macroblock having similar statistical mode information with neighboring macroblocks at the base and enhancement layers for both quality and spatial SVC.

From Table 1, one can see that a current macroblock exhibits similar mode information with its neighboring macroblocks. To save space, only one table for one video sequence is shown here. For other video sequences, similar outcomes are obtained.

Table 1 Percentage of current macroblock having similar statistical mode information with neighboring macroblocks at the base and enhancement layers for both quality and spatial scalable video coding (the mode is equal to the mode of at least one of the three neighbors as defined in Fig. 4)

| Quatization parameters	Quality Scalable video coding (50 frames to be encoded, news video with format QCIF at base and enhancement layers)		Spatial scalable video coding 50 frames to be encoded, news video with format QCIF at base and CIF at enhancement layers)	
	Base layer (%)	Enhancement layer (%)	Base layer (%)	Enhancement layer (%)
QP_base = 24 QP_enhance = 27	69.90	74.12	76.23	64.78
QP_base = 29 QP_enhance = 32	69.13	71.56	77.33	62.94
QP_base = 34 QP_enhance = 37	67.34	69.90	80.51	66.03
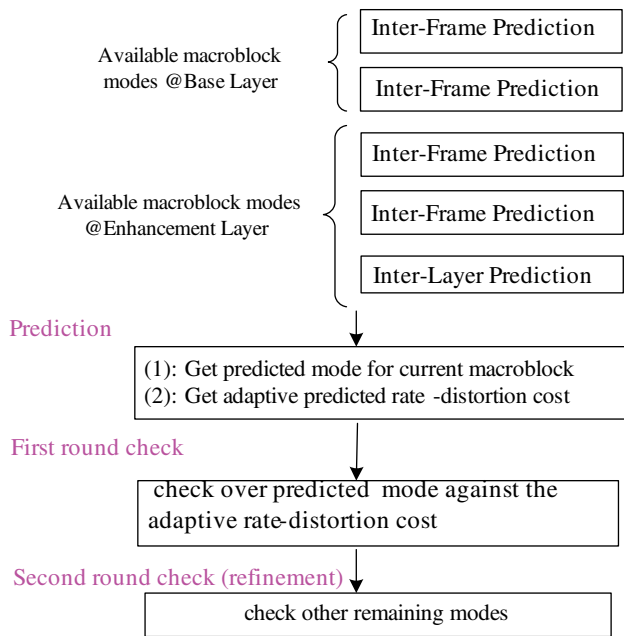
Fig. 3 Introduced fast adaptive mode selection algorithm for SVC

3 Fast adaptive termination mode selection

This section covers the details of our fast adaptive termination mode selection algorithm. Although the existing prediction (including inter-layer prediction) methods achieve coding efficiency, their amount of computation is high. Our introduced algorithm consists of three steps: (1) mode prediction based on spatial neighboring macroblocks or collocated macroblock in the base layer, (2) checking the rate-distortion cost of the predicated mode against an adaptive threshold for early termination, and (3) a refinement step by checking the other remaining modes. Figure 3 provides a flowchart of our algorithm applicable to both the base and enhancement layers in SVC. For each macroblock in the base layer, which is similar to single layer video coding, there are inter-frame prediction including skip, 16×16, 16×8, 8×16, 8×8, 8×4, 4×8, and 4×4 and intra-frame prediction including Intra 16×16, Intra 8×8, and Intra 4×4. However, for each macroblock in the enhancement layer, there is an additional inter-layer prediction available to improve coding efficiency. Therefore, compared with the macroblocks in the base layer, the macroblocks in the enhancement layer require more computation. Generally speaking, the introduced algorithm consists of three main steps: prediction of current macroblock, a first round check using predicted information and a second round check or refinement, see Fig. 3. More detailed explanation is covered in Sect. 3.3. In the following subsections, we address the prediction mode from
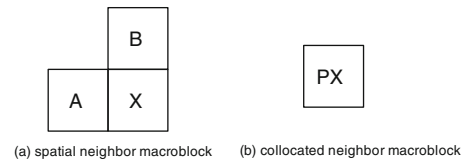

Fig. 4 a Spatial neighboring macroblocks around a current macroblock X and **b** a collocated macroblock PX in the base layer

neighboring macroblocks and the adaptive threshold which is obtained from neighboring coded blocks.

3.1 Fast mode prediction

In H.264/AVC, video coding is performed on macroblocks from the up-left to the right-bottom direction. Adjacent macroblocks being encoded in the same frame generally have the same characteristics such as the same motion or similar detailed regions. For example, as illustrated in Fig. 4, a current macroblock X in the enhancement layer may have similar characteristics with its neighboring macroblocks A and B, or a collocated macroblock PX at the base layer. For example, if A and B have both 16×16 mode, and a collocated macroblock has 16×8 mode, then the mode prediction for a current macroblock is limited to 16×16 and 16×8 mode. This means that during the first round check, our algorithm only checks the 16×16 and 16×8 modes. After obtaining the mode prediction, an adaptive threshold is set up for comparing the rate distortion of a current macroblock. Note that in order to provide a reliable adaptive threshold for comparison, a modulator term named β based on the predicted cost is used to establish a trade-off between complexity and accuracy as indicated in Sect. 2.

3.2 Adaptive rate-distortion threshold

In [10], we introduced an adaptive threshold for mode selection in H.264. In this subsection, the same approach is used for SVC. Assuming the minimum rate-distortion cost for a current macroblock is RD_{best}, and the predicted rate-distortion cost according to spatial and temporal correlations is RD_{pred}, a threshold RD_{thresh} for early termination is defined according to the formula $RD_{thresh} = (1 + \beta) \times RD_{pred}$, where the modulator$_{dsds}$ provides a trade-off between computational efficiency and accuracy. If the rate-distortion cost of a current macroblock for a specific mode is less than RD_{thresh}, it stops checking the other modes and exits the current macroblock. It is important to assign an appropriate value for the modulator β. In what follows, we present how to assign an appropriate modulator value.

Assuming the minimum SAD for a current macroblock is SAD_{best} and the SAD for exiting is SAD_{exit}, the threshold SAD_{thresh} is defined to meet the requirement $SAD_{best} \leq SAD_{exit} \leq SAD_{thresh}$. After motion estimation, the residue information is used for the DCT transformation using the following formula:

$$F(u,v) = \left(\frac{2}{N}\right)^{\frac{1}{2}}\left(\frac{2}{M}\right)^{\frac{1}{2}}\sum_{i=0}^{N-1}\sum_{j=0}^{M-1} \text{diff}(i,j)C(i)C(j)$$
$$\cos\left(\frac{\pi u}{2N}(2i+1)\right)\cos\left(\frac{\pi v}{2M}(2j+1)\right) \tag{1}$$

where $C(i) = \begin{cases} 1/\sqrt{2}; i = 0 \\ 1; i \neq 0 \end{cases}$ and diff (i, j) denotes the pixel difference between the pixel value in the current macroblock at (i, j) position and the corresponding pixel value in the best match block in the reference frame.

Next, the quantization step Q_{step} is used to quantize the DCT coefficients. If the following condition $\left|\frac{F(u,v)}{Q_{step}}\right| < 1$ is satisfied, no information is coded. As a result, we have

$$|F_1(u,v) - F_2(u,v)| \leq Q_{step}$$

$$\Rightarrow |F_1(u,v) - F_2(u,v)| = \left(\frac{2}{N}\right)^{\frac{1}{2}}\left(\frac{2}{M}\right)^{\frac{1}{2}}\sum_{i=0}^{N-1}$$

$$\sum_{j=0}^{M-1}\left|\text{Diff}_1^B(i,j) - \text{Diff}_2^B(i,j)\right|$$

$$C(i)C(j)\cos\left(\frac{\pi u}{2N}(2i+1)\right)\cos\left(\frac{\pi v}{2M}(2j+1)\right) \leq Q_{step}$$

$$\Rightarrow \left(\frac{2}{N}\right)^{\frac{1}{2}}\left(\frac{2}{M}\right)^{\frac{1}{2}}\sum_{i=0}^{N-1}\sum_{j=0}^{M-1}\left|\text{Diff}_1^B(i,j) - \text{Diff}_2^B(i,j)\right| \leq Q_{step}$$

$$\Rightarrow \sum_{i=0}^{N-1}\sum_{j=0}^{M-1}\text{Diff}_1^B(i,j) - \sum_{i=0}^{N-1}$$

$$\sum_{j=0}^{M-1}\left|\text{Diff}_2^B(i,j)\right| \leq Q_{step} * \left(\frac{N}{2}\right)^{\frac{1}{2}}\left(\frac{M}{2}\right)^{\frac{1}{2}} \tag{2}$$

where F_1 and F_2 are predicted and best outcomes, respectively. Now, let us consider

$$SAD_{thresh} = \sum_{i=0}^{N-1}\sum_{j=0}^{M-1}\left|\text{Diff}_1^B(i,j)\right|$$
$$SAD_{best} = \sum_{i=0}^{N-1}\sum_{j=0}^{M-1}\left|\text{Diff}_2^B(i,j)\right| \tag{3}$$

by assuming $SAD_{pred} \approx SAD_{best}$, the modulator β can be written as follows:

$$SAD_{thresh} - SAD_{best} \leq Q_{step} * \left(\frac{N}{2}\right)^{\frac{1}{2}}\left(\frac{M}{2}\right)^{\frac{1}{2}}$$

$$\Rightarrow SAD_{pred}(1 + \beta) - SAD_{best} \leq Q_{step} * \left(\frac{N}{2}\right)^{\frac{1}{2}}\left(\frac{M}{2}\right)^{\frac{1}{2}}$$

$$\Rightarrow \beta \leq Q_{step} * \left(\frac{N}{2}\right)^{\frac{1}{2}}\left(\frac{M}{2}\right)^{\frac{1}{2}}/SAD_{pred} \tag{4}$$

The modulator β is seen to depend on two factors, namely quantization step and block size. Furthermore, the SAD threshold for a current macroblock can be obtained via $SAD_{thresh} = SAD_{pred} \times (1 + \beta)$. Similarly, if one replaces SAD with RD cost, the following formula is obtained $RD_{thresh} = RD_{pred} \times (1 + \beta)$. The process is adaptive as it depends on the predicted rate-distortion cost derived from spatial and temporal correlations.

It is worth stating two points here. First, the idea behind the above equations is whether the residue after quantization is almost zero as this is considered to be good enough and thus it is not necessary to search additional blocks. Second, the derivation of beta value is based on SAD. Since in our algorithm we do not consider the low complexity mode for video encoding based on SAD, the beta value is not applied in the low complexity mode. However, the derived beta can be extended into RD, which is regarded as the high complexity mode during video encoding.

3.3 Fast adaptive mode decision algorithm

As mentioned previously, fast adaptive mode decision is based on the statistical similarity between a current macroblock and its neighboring macroblocks. For macroblocks at the enhancement layer, collocated macroblocks at the base layer is also considered. For the base layer, the available modes for one macroblock include the inter-frame prediction from 16×16 to 4×4 and the intra-frame prediction. And, at the enhancement layer, in addition to the inter-frame and the intra-frame prediction, the available modes for one macroblock includes the additional inter-layer prediction shown in Fig. 2. Here the best mode is selected by checking all the rate-distortion costs from the available modes. However, in order to maintain accuracy, the rate-distortion cost of the predicted mode is first checked against the adaptive threshold. If failed, all the other modes are checked and then the mode corresponding to the minimum rate-distortion cost is selected. Basically, our algorithm consists of the following three steps:

(1) Prediction from neighboring blocks and collocated macroblocks. During prediction, we obtain the prediction mode and rate-distortion threshold cost for a current macroblock. For example, in Fig. 4, if the modes from A, B and PX are 16×16, 16×16, 16×8 with the rate-distortion cost being {373,492, 415}, among all the modes at the enhancement layer, 16×16 and 16×8 are flagged as true to be checked in the first round check and all the other modes are flagged as false, and RD_{pred} is taken as the average rate distortion cost, that is $(373 + 492 + 415)/3 = 427$.

(2) First round check. During the first round check, the predicted mode is checked. For example, the 16×16 and 16×8 modes get chosen in the above example. In order to provide a balance between complexity and accuracy, the adaptive threshold $RD_{thresh} = RD_{pred} \times (1 + \beta)$ is made dependent on the average prediction cost from neighboring macroblocks and modulator β. First, the rate-distortion of the current macroblock with the mode 16×16 is checked against RD_{thresh}. If it is less than RD_{thresh}, the algorithm will terminate and go to the next macroblock. Otherwise, it will go to check the rate-distortion of the current macroblock with the mode 16×8 against RD_{thresh}. If it is less than RD_{thresh}, it will terminate and go to the next macroblock. Otherwise, it checks all the other modes.

(3) Second round check or refinement. If the above rate-distortion cost of the predicted modes is still larger than the adaptive threshold, the algorithm enters into a refinement stage. By doing so, it is made sure the coding efficiency does not suffer much

Generally speaking, our algorithm includes the following attributes:

1. Mode prediction is done based on neighboring and collocated macroblocks,
2. An adaptive rate distortion cost based on neighboring macroblocks is used through a modulator term β reflecting the quantization parameter,
3. Both the base and enhancement layers are considered.

4 Experimental results

The introduced algorithm was implemented and incorporated into the JSVM9.8 reference software. Its outcome was compared with the mode decision algorithm in JSVM9.8 to verify its effectiveness. In addition, a comparison with the existing fast mode selection algorithms in SVC [7, 8] was also performed. A two-layer spatial layer situation was considered here, where the base layer had a resolution of 176×144 and the enhancement layer had a resolution of 352×288. In addition, a two-layer quality was performed, where the base layer and the enhancement layer had an identical resolution of 176×144. Additional parameters used in our experiments are listed below:

1. motion search range $= 16$;
2. GOP size $= 16$;
3. reference frame $= 1$;
4. MV resolution $= ¼$ pel;
5. QP $= 22$, 27,32,37 for base layer; and QP $= 19$, 24, 29,34 for enhancement layer;
6. frame out rate $= 15$ HZ;
7. total number of encoded frames $= 50$;

The comparison measures considered were DBR (delta bitrate change) and DSNR (delta PSNR). These measures are normally used to measure video quality for two different algorithms, which can be derived by using the software *avsnr4* [11]. This software normalizes the parameters involved before comparing two different algorithms—DSNR is the delta PSNR between Algorithm A and Algorithm B assuming both are operating at the same bitrates at all measured points. DBR, also called Percent Mode, gives the percent bitrate increase assuming Algorithm A and Algorithm B have the same PSNR at all measured points. Computational efficiency was measured by the amount of time reduction (TR), which was computed as follows:

$$\text{Time\_reduction} = \frac{\text{Time}_{standard} - \text{Time}_{introduced}}{\text{Time}_{standard}} \times 100\%$$

(5)

Table 2 shows a summary of the comparison results between our mode decision algorithm and the one in JSVM9.8 for scalable spatial and quality video coding. As can be observed from this table, our algorithm provides a time reduction in the range of 20–50% depending on the video content, while generating the same video quality for scalable spatial video coding. And for scalable quality video coding, the time reduction is in the range of 30–55%. It is seen that the video quality is maintained within 5% of the original. Tables 3 and 4 provide the comparison results between the existing fast mode selection algorithms and JSVM9.8. It is seen that the method in [7] achieves time reduction with a quality loss of more than 5%. And the method in [8] achieves good results in quality SVC, but poor results in spatial SVC.

We also applied our algorithm to HD video sequences. The same video coding parameters mentioned above were used except the frame size was higher at $1,280 \times 720$. Table 5 shows the comparison outcome for the quality scalable video coding over four different HD video sequences [12].

For better visualization, for one of the HD sequences, the comparison between the rate-distortion of our algorithm and the JSVM9.8 implementation is shown in Fig. 5.

Table 2 Comparison outcome between our introduced algorithm and the one in JSVM9.8 for spatial and quality scalable video coding

	Sequence	Introduced versus mode decision algorithm in JSVM9.8 @scalable spatial video coding			Introduced versus mode decision algorithm in JSVM9.8 @scalable quality video coding		
		DSNR (dB)	DBR (%)	TR (%)	DSNR (dB)	DBR (%)	TR (%)
Base layer@15HZ	Hall	−0.071	1.23	30.36	−0.071	1.24	49.40
	Foreman	−0.247	4.67	31.50	−0.259	4.89	37.72
	Mobile	−0.171	2.58	21.69	−0.159	2.40	28.62
	News	−0.207	2.47	52.29	−0.208	2.48	55.59
	Silent	−0.172	2.41	37.86	−0.182	2.53	48.90
Enhancement layer @15HZ	Hall	−0.049	1.45	30.36	−0.159	2.99	49.40
	Foreman	−0.130	2.66	31.50	−0.228	4.13	37.72
	Mobile	−0.094	1.32	21.69	−0.178	2.55	28.62
	News	−0.261	4.32	52.29	−0.338	3.87	55.59
	Silent	−0.137	2.26	37.86	−0.233	3.00	48.90

Table 3 Comparison outcome between W029 [7] and mode decision algorithm in JSVM9.8 for spatial and quality scalable video coding

	Sequence	W029 versus mode decision algorithm in JSVM9.8 @scalable spatial video coding			W029 versus mode decision algorithm in JSVM9.8 @scalable quality video coding		
		DSNR (dB)	DBR (%)	TR (%)	DSNR (dB)	DBR (%)	TR (%)
Base layer@15HZ	Hall	0.00	0.00	0.00	0.00	0.00	0.00
	Foreman	0.00	0.00	0.00	0.00	0.00	0.00
	Mobile	0.00	0.00	0.00	0.00	0.00	0.00
	News	0.00	0.00	0.00	0.00	0.00	0.00
	Silent	0.00	0.00	0.00	0.00	0.00	0.00
Enhancement layer @15HZ	Hall	−0.192	5.92	67.91	−0.144	2.86	13.45
	Foreman	−0.615	13.42	66.52	−0.298	5.62	24.34
	Mobile	−0.348	4.96	67.94	−0.263	3.83	21.07
	News	−0.395	6.68	66.40	−0.269	3.04	14.34
	Silent	−0.390	6.54	66.13	−0.268	3.43	14.03

Table 4 Comparison outcome between He [8] and mode decision algorithm in JSVM9.8 for spatial and quality scalable video coding

	Sequence	He versus mode decision algorithm in JSVM9.8 @scalable spatial video coding			He versus mode decision algorithm in JSVM9.8 @scalable quality video coding		
		DSNR	DBR	TR (%)	DSNR (dB)	DBR (%)	TR (%)
Base layer@15HZ	Hall	0.00 dB	0.00%	0.00	0.00	0.00	0.00
	Foreman	0.00 dB	0.00%	0.00	0.00	0.00	0.00
	Mobile	0.00 dB	0.00%	0.00	0.00	0.00	0.00
	News	0.00 dB	0.00%	0.00	0.00	0.00	0.00
	Silent	0.00 dB	0.00%	0.00	0.00	0.00	0.00
Enhancement layer @15HZ	Hall	NA	NA	11.84	−0.175	3.34	19.43
	Foreman	NA	NA	11.93	−0.282	5.08	11.65
	Mobile	NA	NA	23.99	−0.268	3.89	17.08
	News	NA	NA	1.11	−0.292	3.32	18.39
	Silent	NA	NA	6.33	−0.248	3.18	17.97

NA means the result was unacceptably poor compared with JSVM9.8

Table 5 Comparison outcome between our introduced and the mode decision algorithm in JSVM9.8 for quality scalable video coding over high definition video sequences

|	Sequence	Introduced versus mode decision algorithm in JSVM9.8 @scalable spatial video coding		
		DSNR (dB)	DBR (%)	TR (%)
Base layer@15HZ	720p50_mobcal_ter	−0.036	1.10	27.19
	720p50_parkrun_ter	−0.089	1.96	20.47
	720p50_shields_ter	−0.045	1.55	29.40
	720p5994_stockholm_ter	−0.045	2.11	25.58
Enhancement layer @15HZ	720p50_mobcal_ter	−0.032	0.88	27.19
	720p50_parkrun_ter	−0.134	2.54	20.47
	720p50_shields_ter	−0.052	1.78	29.40
	720p5994_stockholm_ter	−0.035	1.59	25.58

Fig. 5 Rate-distortion curve comparison for 720p50_mobcal_ter HD sequence between our algorithm and the JSVM implementation at different temporal levels
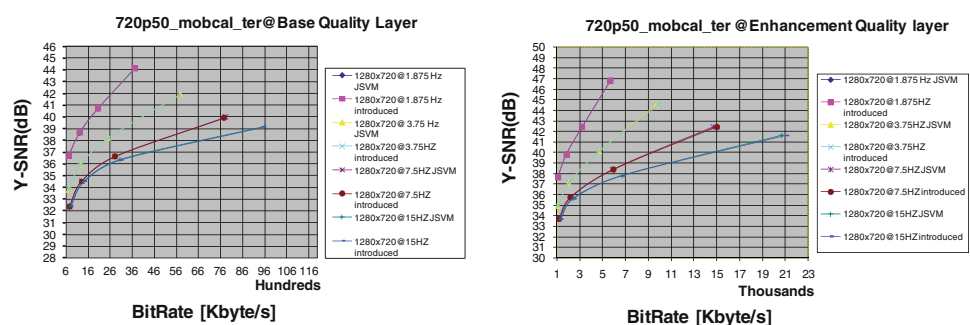
Figure 6 exhibits the time spent for video coding in JSVM9.8 as compared to our algorithm on a 2.4 GHz Pentium-4 PC.

From Figs. 5 and 6, it can be seen that our algorithm achieves about 30% time reduction compared with the JSVM9.8 implementation. It is important to note that this time reduction is done at little or no expense in video coding efficiency, as indicated by the graphs in Fig. 5.
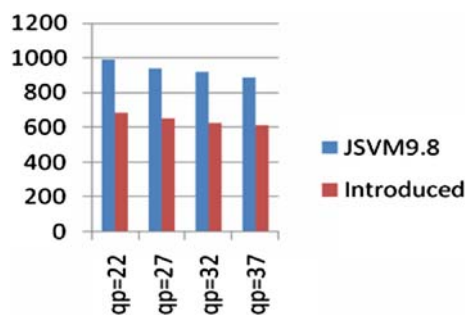
Fig. 6 Comparison of video coding time between our algorithm and the JSVM9.8 implementation at different quantization levels (vertical axis unit is seconds)

5 Conclusion

In order to obtain coding efficiency comparable to single layer video coding, in addition to the traditional inter prediction and intra prediction, an additional inter-layer prediction is performed in SVC. To lower the computational complexity associated with this additional prediction, this paper has introduced an algorithm for fast mode selection in SVC. The reduction in the computational complexity is achieved by utilizing a fast early termination approach for mode selection. It has been shown that at least a one-third reduction in computation time is obtained at the expense of a tolerable loss in video quality.

References

1. Video Coding for Low Bit Rate Communications, ITU-T, ITU-T Rec. H.263, Version 1: November 1995, Version 2: January 1998, Version 3: November 2000
2. Coding of Audio-Visual Objects Part 2-Visual, ISO/IEC JTC1, ISO/IEC 14492-2, Version 1: April 1999, Version 2: February 2000, Version 3: May 2004
3. Yin, P., Tourapis, H.-Y.C., Tourapis, A.M., Boyce, J.: "Fast mode decision and motion estimation for JVT/H.264,"

Proceedings of International Conference on Image Processing, pp. 853–856 (2003)

4. Jeon, B.: "Fast Mode Decision for H.264," Document JVT-J033, December 2003

5. Pan, F., Lin, X., Susanto, R., Lim, K., Li, Z., Feng, G., Wu D., Wu, S.: "Fast mode decision for intra prediction," Doc. JVT-G013, March 2003

6. Wu, D., Pan, F., Lim, K., Wu, S., Li, Z., Lin, X., et al.: Fast intermode decision in H.264/AVC video coding. IEEE Trans Circ Syst Video Tech **15**(7), 953–958 (2005). doi:10.1109/TCSVT.2005.848304

7. Lin, H., Peng W., Hang, H., "Low-complexity macroblock mode decision algorithm for combined CGS and Temporal Scalability", JVT-W029, April 2007

8. Li, H., Li, Z.G., Wen, C.: Fast Mode Decision algorithm for inter-frame coding in fully scalable video coding. IEEE Trans Circ Syst Video Tech **16**(7), 889–895 (2006). doi:10.1109/TCSVT.2006.877404

9. Tourapis, A.M., Au, O.C., Liou, M.L.: Highly efficient predictive zonal algorithm for fast block-matching motion estimation. IEEE Trans Circ Syst Video Tech **12**, 934–947 (2002). doi:10.1109/TCSVT.2002.804894

10. Ren, J., Kehtarnavaz, N.: Fast Adaptive Early Termination for Mode Selection in H.264 Scalable Video Coding, to appear in Proceedings of IEEE ICIP, San Diego (2008)

11. Bjontegaard, G.: Calculation of Average PSNR Difference between RD-Curves, Document VCEG-M33, April 2001

12. High Definition video sequences, (online) ftp://ftp.ldv.e-technik.tu-uenchen.de/pub/test_sequences/

Author Biographies

Jianfeng Ren received his BS degree from Xi'an University of Technology, P. R. China, in 2000 and MS degree from NorthWestern Polytechnic University, P. R. China, in 2003. Currently, he is pursuing his PhD in Electrical Engineering at the University of Texas at Dallas. His research interests include video coding and real-time image processing, particularly interested in low-power video coding.

Nasser Kehtarnavaz received the PhD degree from Rice University, Houston, TX, in 1987. He is currently a professor with the Department of Electrical Engineering and Director of the Signal and Image Processing Lab at the University of Texas at Dallas. He was previously a professor of Electrical Engineering at Texas A&M University. His research interests include digital signal and image processing, real-time image processing, and pattern recognition. He has authored or co-authored five books and numerous articles as related to these areas. Dr. Kehtarnavaz is currently serving as the Chair of the Dallas Chapter of the IEEE Signal Processing Society.