Research Paper

Fast Adaptive Hierarchical Prediction Algorithm for H.264/AVC Scalable Video Coding

ABDELRAHMAN ABDELAZIM^{1,a)} STEPHEN MEIN¹
MARTIN R. VARLEY¹ DJAMEL AIT-BOUDAUD²

Received: April 4, 2011, Accepted: December 11, 2011, Released: March 28, 2012

Abstract: The objective of scalable video coding (SVC) is to enable the generation of a unique bitstream that can adapt to various bit-rates, transmission channels and display capabilities. The scalability is categorised in terms of temporal, spatial, and quality. In order to improve encoding efficiency, the SVC scheme incorporates inter-layer prediction mechanisms to complement the H.264/AVC very refined Motion Estimation (ME) and mode decision processes. However, this further increases the overall encoding complexity of the scalable coding standard. In this paper several conditional probabilities are established relating motion estimation characteristics and the mode distribution at different layers of the H.264/SVC. An evaluation of these probabilities is used to structure a low-complexity prediction algorithm for Group of Pictures (GOP) in H.264/SVC, reducing computational complexity whilst maintaining similar RD performance. When compared to the JSVM software, the proposed algorithm achieves a significant reduction of encoding time, with a negligible average PSNR loss and bit-rate increase in temporal, spatial and SNR scalability. Experiments are conducted to provide a comparison between our method and recently developed fast mode selection algorithms. These demonstrate the proposed method achieves appreciable time savings for scalable spatial and scalable quality video coding, while maintaining similar PSNR and bit rate.

Keywords: video coding, Motion Estimation, interpolation, scalability, H.264

1. Introduction

In SVC the scalability is categorised into temporal, spatial, and quality. For temporal and spatial scalabilities, subsets of the bit stream (referred to as sub-streams) represent the source content with a reduced frame rate (temporal resolution) or picture size (spatial resolution), respectively. With quality scalability, the sub-stream provides the same spatial and temporal resolutions as the complete bit stream, but with a lower signal-to-noise ratio (SNR).

To enhance the coding efficiency when compared to prior video coding methods several new features have been added to the H.264/AVC standard [1]. Amongst those features are the exploitation of variable block-size, motion compensation with small block sizes and quarter-sample-accurate motion compensation. Since the H.264/SVC standard [2] was developed as an extension of the H.264/AVC therefore all of the AVC well-designed core coding tools were inherited. Additionally the SVC incorporates inter-layer prediction and Inter-Layer Residual Prediction mechanisms; those have been included to enable the usage of as much lower layer information as possible for improving rate-distortion efficiency of the enhancement layers. However, this further increases the overall encoding complexity of the SVC standard.

In the movement toward the Next Generation Video Coding (NGVC) H.265 the concept of adaptive interpolation filters that are specifically adapted to statistics of the current image are intro-

duced in Ref. [3], and motion compensated prediction with $\frac{1}{8}$ -pel displacement vector resolution is discussed in Ref. [4]. However, little research has been carried out to investigate and define situations when the interpolation process is unnecessary. In this paper the algorithm proposed in Ref. [5] is extended to a Group of Pictures (GOP) in H.264/SVC. Moreover, the algorithm has been utilised in the mode selection process of different scalability layers to reduce the computational complexity.

The rest of the paper is organised as follows. Section 2 gives a brief overview of the hierarchical prediction proposed in the H.264/SVC. Section 3 discusses some related work in the field. Section 4 describes the proposed mode decision algorithm, while Section 5 presents details and discussion of a comprehensive list of comparative experimental results. Section 6 concludes the paper.

2. Overview Of The Scalable Video Coding (SVC)

2.1 Rate Distortion Optimisation (RDO)

In the AVC, for the first stage of the ME, a picture is partitioned into fixed-size macroblocks that each cover a rectangular picture area of 16×16 pixels of the luma component, then an integer-pixel-motion-search is performed for each square block of the slice to be encoded in order to find one (or more) displacement vector(s) within a search range. The user defined search range is used for all Inter-modes and reference frames. The full-pel search positions are organised in a spiral structure around a

¹ University of Central Lancashire, Lancashire, England, PR1 2HE, UK

² University of Portsmouth, Hampshire, England, PO1 2UP, UK

^{a)} aabelazim@uclan.ac.uk

Fig. 1 Fractional pixel search positions.

prediction vector. This vector is predicted from vectors of previously encoded adjacent block. The exhaustive full-pel motion search is computed by minimising the following equation for all full-pel positions in the search region.

$$J(m, \lambda_{Motion}) = SAD[s, c(m) + \lambda_{Motion} \cdot R(m - p)] \quad (1)$$

with $m = (m_x, m_y)$ being the motion vector, $p = (p_x, p_y)$ being the prediction for the motion vector, λ_{Motion} being a Lagrange multiplier and SAD is the Sum of Absolute Differences. The rate term $R(m - p)$ represents the predicted motion vector error information only, and is computed by a table-lookup. The rate is estimated by using the universal variable length code (UVLC) table.

For an $M \times N$ block, SAD is computed as

$$SAD[s, c(m)] = \sum_{x=1}^M \sum_{y=1}^N |s(x, y) - c(x - m_x, y - m_y)| \quad (2)$$

with s being the original video block, and c the predicted video block at the position designated by m in the reference picture considered.

After the integer-pixel-motion-search finds the best match, the values at half-pixel positions around the best match are interpolated by applying a one-dimensional 6-tap FIR filter horizontally and vertically. Then the values of the quarter-pixel positions are generated by averaging pixels at integer and half-pixel positions.

Figure 1 illustrates the interpolated fractional pixel positions. Upper-case letters indicate pixels on the full-pixel grid, while numeric pixels indicate pixels at half-pixel positions and lower case pixels indicate pixels in between at quarter-pixel positions. Considering Fig. 1, if (for example) the integer best match is position E, the half-pixel positions 1, 2, 3, 4, 5, 6, 7, 8 are searched using Eq. (1). Suppose position 7 is the best match of the half pixel search. Then the quarter-pixel positions a, b, c, d, e, f, g, h are searched using again Eq. (1).

Similar to H.264/AVC the SVC has many different Inter macroblock size (mode) choices to code a macroblock as shown in **Fig. 2**. The modes are related to the partitioning possibilities of a 16×16 luma macroblock. When mode 4 is considered, then modes 5, 6, and 7 must be considered within each of the four individual 8×8 sub-macroblocks. Additionally a SKIP mode is introduced (mode 0) where the macroblock motion vector is identical to the prediction motion vector of the 16×16 motion compensated macroblock mode.

To choose the best mode, a Lagrangian multiplier is used to compute the cost for each mode and select the mode that gives the smallest cost. Therefore, the mode decision process minimises

Fig. 2 Different partition sizes in a macroblock.

Fig. 3 An example of a 2 layer GOP of size 16 for enabling temporal and spatial scalability.

the following Lagrangian

$$J = Distortion + \lambda_{Mode} \cdot Rate \quad (3)$$

The *Distortion* measurement quantifies the quality of the reconstructed pictures whilst *Rate* measures the bits needed to code the macroblock using the particular mode. The best mode will be selected as a compromise between the least distortion and the fewest number of bits to code the macroblock. Both factors are simultaneously quantified by the Lagrangian multiplier, λ_{Mode} .

Although the exploitation of different macroblock sizes significantly improves the RD performance, it increases the complexity of the ME process. This is due to the fact that for a given macroblock partition, motion estimation/compensation has to be performed before deciding on the best mode.

2.2 Group of Pictures Hierarchical Prediction

In H.264/SVC a Group of Pictures (GOP) is an encoding of a contiguous subset of frames from a video sequence; each GOP consists of all frames between two successive frames at the lowest temporal resolution, plus the second of the temporal base layer frames (the first is considered to belong to the previous GOP). All information required to decode any one frame from the GOP is contained within it. Using the GOP concept, SNR gains of more than 1 dB can be obtained for medium bit rates when compared to the widely used IBBP coding structure [2].

Figure 3 shows an example of a 2 layer GOP of size 16. This example shows a dyadic coding structure as it is the most used structure for its demonstrated coding efficiency. Furthermore, Fig. 3 shows the base and the enhancement layers have the same frame rate, however the H.264/SVC supports non-dyadic hierar-

chical prediction structures and different frame rates for the base and the enhancement layer. Frame 0 belongs to the previous GOP but is used in the prediction of the current GOP.

2.3 Temporal Scalability

The enhancement layer pictures are typically coded as B-pictures, where the reference picture lists 0 and 1 are restricted to the temporally preceding and succeeding picture, respectively, with a temporal layer identifier less than the temporal layer identifier of the predicted picture. In the figure, frames 0 and 16 are coded first as intra-frames, then frame 8 is coded as a B-picture using frame 0 and 16 as references. Then the rest of the frames are coded in the order shown in the figure. As an example when frame 10 is coded, up to 3 references can be used in list 0 (left references), and 2 references in list 1 (right references). Additionally bi-prediction signal, formed by a weighted sum of list 0 and list 1, can be used. In the JSVM software [6] the exact number of references is a user-configurable parameter.

2.4 Spatial Scalability

In each spatial layer, motion-compensated prediction and intra-prediction are employed similar to the processes used for the base layer. Additionally an inter-layer prediction mechanism is incorporated as shown in Fig. 3. For instance, when frame 10 is coded, in addition to the 5 possible references described in the temporal scalability, information from the corresponding frame in the base layer can be used in motion-compensated prediction.

2.5 Quality Scalability

Quality scalability is considered as a special case of spatial scalability with identical picture sizes for base and enhancement layers, therefore the same motion-compensated predictions including the inter-layer prediction are employed.

3. Related Work

To reduce the implementation complexity, several fast techniques have been presented recently [7], [8], [9], [10], [11]; most of them share the same concept of using the correlation between a macroblock (MB) and its neighbours in different layers and also that between a MB in the base layer and its corresponding position in the enhancement layer.

In Ref. [7], a fast mode decision algorithm for inter-frame coding for temporal, spatial and quality scalability is presented. It makes use of the mode-distribution correlation between the base layer and enhancement layers. Specifically, after the exhaustive search technique is performed at the base layer, the number of candidate modes for enhancement layers is reduced.

In Refs. [8] and [9] fast mode decision algorithms exploiting correlations between MBs and their neighbours are proposed. These algorithms have an obvious limitation when applied to fast video sequences or sequences with complex backgrounds, as the irregularity between the MB and its neighbours increases, which results in limited benefits and significant performance degradation. Implementation from Ref. [9] is shown in Fig. 4 as a flowchart.

In Ref. [10] a scheme that generates reduced candidate list for

Fig. 4 Algorithm proposed in Ref. [9].

Fig. 5 Algorithm proposed in Ref. [10].

the current macroblock according to co-located reference macroblocks is presented. It works by calculating the average PSNR of the frames in same temporal levels for a specific mode, and then this PSNR is subtracted from the PSNR of the same mode for the current MB and compared to a threshold. Based on the result an early termination decision is made. A simplified implementation of the algorithm is shown in Fig. 5 as a flowchart.

This scheme shares the same limitation with Refs. [8] and [9] when applied to fast video sequences as the MB and the co-located reference MBs are likely to have different modes.

In Ref. [11] it has been observed that there is a high correlation between the MB and its enclosed partitions when using Forward (FW), Backward (BW) and BI-directional (BI) predictions. Based on this correlation a 3 steps algorithm has been proposed; first, an exhaustive search is carried out for the 16×16 MB including FW, BW and BI search. Secondly based on the result from the first step the BI prediction is selectively eliminated for the 16×8 , 8×16 and 8×8 partitions. Finally, similar step to the second one is implemented to selectively reduce the prediction modes for the 8×4 , 4×8 and the 4×4 partitions based on the 8×8 results. The algorithm is shown in Fig. 6 as a flowchart.

Although the algorithm can provide some time saving whilst maintaining a good RD performance. This time saving is limited

Fig. 6 Algorithm proposed in Ref. [11].

to one prediction type (the BI prediction type) especially for the large block partitions. Furthermore, this prediction type is the least used type.

4. Proposed Algorithm

In our previous work [5] it has been observed that there is a high correlation between the MB and its enclosed partitions when estimating the motion at different resolution. Therefore a two step fast sub-pixel motion estimation scheme based on this observation has been developed.

- (1) In the first step, if the 16×16 MB finds a best match in the full-pixel motion search that does not change after performing the sub-pixel motion search ($cond_1$), then the sub-pixel motion search for all the enclosed 16×8 and 8×16 blocks is disabled.
- (2) Similarly, if in the second step the 8×8 block partitions of the 16×16 MB find the same best match in the full and sub-pixel motion searches ($cond_2$), the sub-pixel motion search for all the enclosed 8×4 , 4×8 and 4×4 blocks is disabled.

In the following sub-sections, several conditional probabilities that relate $cond_1$ and $cond_2$ to the motion estimation and the mode distribution at different layers of the H264/SVC have been evaluated. Moreover, additional conditional probabilities of the mode distribution have been calculated.

4.1 Probability of Mode Selection in Temporal Scalability

The first conditional probabilities focus on measuring the probability of the mode distribution of the different temporal layers in relation to $cond_1$ and the mode distribution of lower temporal layers. Two probabilities were analysed;

- $P1$, the probability of the best mode belonging to the following set Skip = Mode 0, 16×16 = Mode 1, 16×8 = Mode 2 and 8×16 = Mode 3 when $cond_1$ is true.
- $P2$, the probability of the best mode belonging to the same set (Mode < 4) given that the best modes for all the corresponding MBs in list 0 and list 1 at lower temporal layer in the same GOP are < 4.

This analysis was conducted using a GOP of size 16 for the different temporal layers. It reflects the average of using 9 Quanti-

 Table 1 Conditional probabilities for $P1$ and $P2$.

Sequence	QP	$P1$ (%)	$P2$ (%)			
			T1	T2	T3	T4
City	10-42	94.39	95.29	96.02	95.75	97.67
Harbour		97.02	94.06	98.97	99.13	91.73
Hall		99.28	99.53	99.68	98.54	99.57
Foreman		95.43	96.71	94.37	94.66	92.57
Mobile		96.13	92.16	95.68	97.24	92.33
Silent		99.67	99.88	99.98	98.65	99.78
News		98.79	98.69	98.02	96.87	97.76

 Table 2 Conditional probabilities for $P3$, $P4$ and $P5$.

Sequence	Spatial			SNR		
	$P3$ (%)	$P4$ (%)	$P5$ (%)	$P3$ (%)	$P4$ (%)	$P5$ (%)
City	94.39	99.19	33.88	85.09	84.37	50.02
Harbour	97.02	98.82	26.81	80.14	80.55	42.25
Hall	99.98	99.28	13.53	98.23	99.8	19.19
Foreman	96.11	98.15	18.21	82.62	83.67	37.5
Mobile	95.76	97.23	23.62	81.19	81.22	22.13
Silent	99.69	99.93	36.18	95.46	98.35	51.68
News	99.99	99.65	33.36	96.78	98.71	53.17

sation Parameters (QP) (10-14-18-22-26-30-34-38-42), for seven standard test video sequences of size CIF and QCIF, as shown in Table 1.

T1, T2, T3 and T4 refer to the different temporal layers shown in Fig. 3. For example when encoding a MB from frame 10 in Fig. 3, $P2$ is calculated for T1, T2 and T3. If the MB to encode is from frame 9, $P2$ is calculated for T1, T2, T3 and T4.

From the table it can be seen that the conditional probabilities are always higher than 90%. Therefore, $P1$ and $P2$ can be used as reliable indicators for mode predication.

4.2 Correlation of Mode Selection in Spatial & SNR Scalability

A second set of conditional probabilities was evaluated to calculate the probabilities of the mode distribution in the enhancement layers in relation to $cond_1$ and the mode distribution in the base layer for spatial and SNR scalability. In the study three probabilities were analysed;

- $P3$, the probability of the best mode in the enhancement layer being less than 4 given that $cond_1$ is true for the corresponding MB in the base layer.
- $P4$, the probability of the best mode in the enhancement layer being < 4 given that the best mode of the corresponding MB in the base layer is < 4.
- $P5$, the probability of the best mode in the enhancement layer being equal to the best mode of the corresponding MB in the base layer.

Table 2 shows an average of $P3$, $P4$ and $P5$ when; GOP = 16, enhancement and base frame rate = 15, base QP = 14 – 42 and enhancement QP = 12 – 40.

The table shows high correlation between the MB motion characteristics in the base layer and the best mode in the enhancement layers; this is reflected by the high $P3$ value. Also, the mode distribution characteristics of the enhancement layer are highly correlated to mode distribution characteristics of the base layer and should be used to predict the mode in the enhancement layer, this is shown by the high $P4$ value. In Ref. [7] an observation related to $P4$ has been used to limit the mode selection process. The low

Table 3 Conditional probabilities for P_6 and P_7 .

Sequence	Spatial		SNR	
	P_6 (%)	P_7 (%)	P_6 (%)	P_7 (%)
City	87.60	89.62	85.09	83.09
Harbour	92.86	95.28	87.54	95.15
Hall	85.80	84.86	84.78	83.09
Foreman	82.15	86.26	83.31	81.72
Silent	97.77	93.57	89.14	93.26
News	95.23	97.48	91.38	97.66
Mobile	86.43	82.39	88.95	83.17

value of P_5 demonstrates that no assumptions can be made on the base layer and the enhancement layer having the same best mode.

4.3 Motion Characteristic Probability in Spatial & SNR Scalability

A final set of conditional probabilities was evaluated to calculate the probabilities of the motion characteristics in the enhancement layers in relation to $cond_1$ and the $cond_2$ in the base layer. Two probabilities were calculated;

- P_6 , the probability of the 16×16 in the enhancement layer $cond_1$ being true given that $cond_1$ for the corresponding MB in the base layer is true.
- P_7 , the probability of the 8×8 partition in the enhancement layer $cond_2$ being true when $cond_2$ for the corresponding block in the base layer is true.

Similar experimental conditions as in Section 4.2 were used in the evaluation. The probabilities are shown in **Table 3** below.

The table shows that, on average 88% of the time when integer motion occurs in the base layer similar motion occurs in the enhancement layers. Thus, P_6 and P_7 can be used to adaptively enable and disable the sub-pixel motion search.

4.4 Low Complexity Hierarchical Prediction Algorithm

Based on the above observations, a fast mode decision algorithm and fast sub-pixel motion estimation for inter-frame prediction were developed as follows:

4.4.1 Temporal Scalability:

- (1) If the 16×16 find a best match in the full pixel ME that does not change after performing the fractional pixel ME (Integer pixel Motion), then:
 - limit the mode to 16×16 , 16×8 and 8×16 ,
 - disable the sub pixel ME for the 16×8 and 8×16 .
- (2) If the collective mode in previously encoded macroblock in lower temporal layers is < 4 , then:
 - limit the mode selection to 16×16 , 16×8 and 8×16 .
- (3) If none of the above is true and if the 8×8 find a best match in the full pixel ME that does not change after performing the fractional pixel ME (Integer pixel Motion), then:
 - disable the sub pixel ME for the 8×4 , 4×8 and 4×4 .

4.4.2 Spatial & SNR Scalability:

In addition to stages 1-3 for the temporal scalability, the following three steps are also incorporated in the case of Spatial and SNR scalability:

- (4) If the 16×16 macroblock in the base layer find a best match in the full pixel ME that does not change after performing the fractional pixel ME (Integer pixel Motion), then:
 - limit the mode to 16×16 , 16×8 and 8×16 ,

Fig. 7 Algorithm proposed in this work.

- disable the sub pixel ME for the 16×8 and 8×16 .
- (5) If the best mode for the corresponding block in the base layer is < 4 , then:
 - limit the mode selection to 16×16 , 16×8 and 8×16 .
 - (6) If none of the above is true and if the corresponding 8×8 partition in the base layer finds a best match in the full pixel ME that does not change after performing the fractional pixel ME (Integer pixel Motion), then:
 - disable the sub pixel ME for the 8×4 , 4×8 and 4×4 .
- The algorithm is summarised using the flowchart in **Fig. 7**.

5. Experimental Results

To evaluate the proposed algorithm, a comprehensive set of experiments have been carried out. The scheme is implemented on a JSVM 9.15 encoder [6]. The test platform uses an Intel Core 2 CPU 6420 @ 2.13 GHz with 2.0 GB RAM. The Intel VTune performance analyzer was used to measure the number of machine cycles differences which reflects the total encoding Time Saving (TS).

Additionally, Bjontegaard delta PSNR (BDPSNR), and Bjontegaard delta bit rate (BDBR) [12] have been used to evaluate the proposed algorithm performance versus the JSVM encoder and recent work in the area. The test conditions are shown in **Table 4**.

Table 4 Encoder experiment conditions.

Parameter		Value	Parameter	Value
Frame rate	Base	15/30 Hz	GOP size	16
in/out	Enhancement	15/30 Hz	MV resolution	$\frac{1}{4}$ Pel
QP Setting	Base	14-22 30-38	No. of frames	100
	Enhancement	12-20 28-36	Motion Search range	16
Spatial	Base	QCIF/CIF	Reference Picture	5
Resolution	Enhancement	CIF/4CIF	Search Function	SAD
SNR	Base	QCIF/CIF/4CIF	Temporal Resolution	QCIF/CIF/
Resolution	Enhancement	QCIF/CIF/4CIF		640 × 480/720p

Table 5 Comparison between the proposed algorithm and the JSVM 9.15 software.

Sequence	TS (%)	BD PSNR (dB)	BD BR (%)
Resolution QCIF/CIF			
Foreman	36.90	-0.15	+2.50
Mobile	39.20	-0.16	+1.50
Bus	34.69	-0.05	+0.60
Silent	56.30	-0.11	+2.30
Akiyo	69.40	-0.04	+1.14
Football	42.68	-0.08	+1.80
News	60.20	-0.10	+2.00
Hall	62.50	-0.14	+2.60
Resolution CIF/4CIF			
City	43.40	-0.03	+1.94
Crew	35.76	-0.06	+1.47
Soccer	37.95	-0.14	+1.66
Harbour	39.70	-0.09	+1.16
Resolution 640 × 480			
Akko & Kayo	55.6	-0.08	+0.90
Rena	52.60	-0.11	+1.20
Resolution 720p			
Shields	57.20	-0.06	+1.18
Stockholm	54.61	-0.02	+0.57
Average	48.67	-0.09	+1.53

5.1 Temporal Scalability

The experimental results are given in **Table 5**, negative values in the table denote PSNR degradation and bit rate savings respectively. It can be seen that the proposed scheme achieves an average of 48.67% time saving with negligible losses in PSNR and increments in bit rate.

Additionally the proposed algorithm has been compared with the recently proposed algorithm in Ref. [11]. The comparison is shown in **Table 6**, from the table it can be seen that for a very similar RD performance the proposed algorithm achieves an average of 19.78% time saving. It is important to notice that if the result in Table 5 is compared to the result in Ref. [11] the result is not directly reflected in Table 6. The reason for this is in the paper the authors used a high range of quantisation parameters (28, 32, 36 and 40) which makes the saving more significant. In contrast, in Table 6 the QP range were (14, 22, 30 and 38), this range gives a better evaluation of both methods.

5.2 Spatial & CGS Scalability

In this experiment, initially our algorithm is compared to the mode decision algorithm in JSVM9.15 software; the comparison result is shown in **Table 7**. It can be observed from this table that our algorithm provides a time reduction in the range of 32–62% and 28–64%, depending on the video content, while generating the same video quality for scalable spatial and scalable quality video coding respectively.

Table 6 Comparison between the proposed algorithm and the results from Ref. [11].

Sequence	TS (%)	BD PSNR (dB)	BD BR (%)
Resolution QCIF/CIF			
Foreman	13.40	-0.01	+0.05
Mobile	14.70	-0.01	-0.06
Bus	8.90	+0.01	+0.04
Silent	22.05	-0.02	+0.01
Akiyo	31.80	+0.05	-0.02
Football	23.65	-0.08	+0.03
News	24.30	-0.10	+0.20
Hall	27.80	+0.14	+0.01
Resolution CIF/4CIF			
City	9.80	+0.02	+0.05
Crew	15.70	+0.03	-0.12
Soccer	19.68	-0.01	+0.01
Harbour	18.79	-0.05	+0.17
Resolution 640 × 480			
Akko & Kayo	21.91	-0.09	+0.01
Rena	24.43	-0.01	+0.02
Resolution 720p			
Shields	27.51	-0.02	-0.01
Stockholm	25.67	+0.01	+0.03
Average	19.78	-0.01	+0.04

Table 7 Comparison between the proposed algorithm and the JSVM 9.15 software.

Sequence	Spatial			SNR		
	TS (%)	BD PSNR (dB)	BD BR (%)	TS (%)	BD PSNR (dB)	BD BR (%)
Resolution QCIF/CIF						
Foreman	36.00	-0.10	2.40	36.50	-0.10	3.06
Mobile	35.40	-0.10	2.00	39.10	0.0	1.70
Bus	39.67	-0.09	1.97	42.19	-0.01	1.96
Silent	48.90	-0.10	2.30	56.10	-0.14	2.00
Akiyo	62.85	-0.02	1.03	64.75	-0.02	1.20
Football	37.40	-0.08	2.30	39.95	-0.16	2.30
News	51.50	-0.09	2.00	58.10	-0.12	1.80
Hall	56.60	-0.05	2.10	62.20	-0.07	2.20
Resolution CIF/4CIF						
City	37.20	-0.08	2.16	41.50	-0.09	1.77
Crew	32.60	-0.08	1.47	28.83	-0.06	1.24
Soccer	29.67	-0.01	1.67	31.20	-0.09	1.37
Harbour	36.80	-0.04	1.10	42.22	-0.03	0.90
Average	42.14	-0.07	1.80	44.48	-0.08	1.80

Table 8 provide comparison results between a recently developed fast mode selection algorithm [9] and our method. It shows that our method achieves an average of 18.67% and 14.98% time saving for scalable spatial and scalable quality video coding respectively, while maintaining better PSNR and lower bit rate.

Table 9 provides comparison results between a recently developed fast mode selection algorithm [10] and the proposed method. It shows that proposed method achieves an average of 22.26% time saving while maintaining similar PSNR and bit rate.

Table 8 Comparison between the proposed algorithm and Ren's method [9].

Sequence	Spatial			SNR		
	TS (%)	BD PSNR (dB)	BD BR (%)	TS (%)	BD PSNR (dB)	BD BR (%)
Resolution QCIF/CIF						
Foreman	7.75	+0.14	-2.27	4.40	+0.13	-1.70
Mobile	18.40	+0.07	-0.58	8.60	+0.08	-0.87
Bus	19.70	+0.66	-0.92	9.80	+0.16	-0.99
Silent	10.70	+0.07	+0.11	13.10	+0.09	-1.30
Akiyo	13.50	+0.58	-0.81	16.74	+0.27	-0.19
Football	17.56	+0.61	-1.42	14.90	+0.08	-0.15
News	9.50	+0.34	-2.09	10.20	+0.21	-2.07
Hall	30.10	+0.17	-0.60	11.70	+0.15	-0.69
Resolution CIF/4CIF						
City	25.70	+0.15	-1.60	20.40	+0.13	-1.65
Crew	22.50	+0.93	-2.10	18.95	+0.21	-1.80
Soccer	21.34	+0.08	-1.07	23.40	+0.35	-1.07
Harbour	27.40	+0.05	-1.01	27.60	+0.08	-1.91
Average	18.67	+0.32	-1.19	14.98	+0.16	-1.20

Table 9 Comparison between the proposed algorithm and Dawei's method [10].

Sequence	TS (%)	BD PSNR (dB)	BD BR (%)
Resolution QCIF/CIF			
Foreman	19.72	+0.10	-1.70
Mobile	16.37	-0.08	+0.87
Bus	18.94	-0.01	+0.69
Silent	25.70	+0.03	-1.40
Akiyo	34.60	+0.07	+1.02
Football	15.86	-0.25	+0.78
News	29.40	+0.01	+0.07
Hall	32.80	-0.15	-0.69
Resolution CIF/4CIF			
City	21.98	-0.04	+1.27
Crew	19.83	-0.02	+0.95
Soccer	17.06	-0.01	-0.07
Harbour	14.91	+0.05	+0.68
Average	22.26	-0.05	+0.40

6. Conclusion

In the scalable video coding extension of the H.264/AVC standard, an exhaustive search technique is used to select the best coding mode for each macroblock. This technique achieves the highest possible coding efficiency, but it demands a higher video encoding computational complexity which constrains its use in many practical applications. This paper proposes a combined fast sub-pixel motion estimation and a fast mode decision algorithm for inter-frame coding for temporal, spatial, and coarse grain signal-to-noise ratio scalability. It makes use of correlation between the macroblock and its enclosed partitions at different layers. Experimental results show that the scheme reduces the computational complexity significantly with negligible coding loss and bit-rate increases when compared to JSVM 9.15 and recently reported fast mode decision algorithms. This saved computation can advance the progress in the realisation of the H.264 scalable extension in real-time applications and low complexity coding systems.

Reference

- [1] Wiegand, T., Sullivan, G.J., Bjontegaard, G. and Luthra, A.: Overview of the H.264/AVC video coding standard, *IEEE Trans. Circuits Syst. Video Technol.*, Vol.13, No.7, pp.560–576 (2003).
- [2] Schwarz, H., Marpe, D. and Wiegand, T.: Overview of the Scalable

- Video Coding Extension of the H.264/AVC Standard, *IEEE Trans. Circuits Syst. Video Technol.*, Vol.17, No.9, pp.1103–1120 (2007).
- [3] Rusanovskyy, D., Ugur, K. and Lainema, J.: Adaptive Interpolation with Directional Filters, *Video Coding Experts Group (VCEG) 33rd Meeting*, Shenzhen, China (2007).
- [4] Ostermann, J. and Narroschke, M.: Motion compensated prediction with 1/8-pel displacement vector resolution, *Video Coding Experts Group (VCEG) 30th Meeting*, Hangzhou, China (2006).
- [5] Abdelazim, A., Yang, M.Y. and Grecos, C.: Fast Sub-Pixel Motion Estimation based on the interpolation effect on different block sizes for H264/AVC, *Optical Engineering Letters*, Vol.48, No.3, pp.030504–1–030504–3 (2009).
- [6] Reichel, J., Schwarz, H. and Wien, M.: Joint Scalable Video Model 9.15 (JSVM 9.15), Technical report, Joint Video Team (2009). Doc. JVT-V202.
- [7] Li, H., Li, Z. and Wen, C.: Fast Mode Decision Algorithm for Inter-Frame Coding in Fully Scalable Video Coding, *IEEE Trans. Circuits Syst. Video Technol.*, Vol.16, No.7, pp.889–895 (2006).
- [8] Goh, G., Kang, J., Cho, M. and Chung, K.: Fast Mode Decision for Scalable Video Coding based on Neighboring Macroblock Analysis, *ACM Symposium on Applied Computing*, pp.1845–1846 (2009).
- [9] Ren, J. and Kheirnavaz, N.: Fast Adaptive Early Termination for Mode Selection in H.264 Scalable Video Coding, *IEEE International Conference on Image Processing (ICIP'08)*, San Diego, California (2008).
- [10] Dawei, Y., Hou, B. and Sotudeh, R.: A Fast Mode Decision Algorithm in Spatial and Temporal Scalable Video Coding, *Proc. 3rd International Symposium on Intelligent Information Technology Application*, pp.716–719 (2009).
- [11] Lin, H., Hang, H. and Peng, W.: Fast Temporal Prediction For H.264/AVC Scalable Video Coding, *IEEE International Conference on Image Processing (ICIP'09)*, Cairo, Egypt, pp.3425–3428 (2009).
- [12] Bjontegaard, G.: Calculation of Average PSNR Difference between RD-Curves, Technical report, Video Coding Expert Group of ITU-T (2001). Doc. VCEG-M33.

Abdelrahman Abdelazim received a B.Eng. (Hons) degree in Digital Communication in 2007 and a Ph.D. degree in 2011, both from the University of Central Lancashire (UCLAN), Preston, England. His research interests are in the area of reducing the complexity of Video Coding Encoders in real-time Scalable and Multi-

view applications. He has published several conference and journal papers in the subject of video compression and transmission. From 2007 to 2008, he worked on a number of industrial digital communication projects. Since 2008 he has been teaching at the school of Computing, Engineering and Physical Sciences, UCLAN. He is also a member of the Applied Digital Image and Signal Processing (ADSIP) Research Centre.

Stephen Mein holds a B.Eng. (Hons) degree in Electronic Engineering (2002), and a Ph.D. (2009). Since 2007 he has worked as a Lecturer in Electronics within the School of Computing, Engineering and Physical Sciences (CEPS) at the University of Central Lancashire. His research interests focus on the application of

image processing techniques and probabilistic methods to problems in the field of ultrasonic non-destructive testing.

Martin R. Varley holds a B.Sc. (Hons) degree in Electrical and Electronic Engineering (1985), and a Ph.D. (1990). He has worked in UK higher education for over 25 years, and is currently a Principal Lecturer in Electronic and Computer Engineering, and Academic Lead for Engineering, in the School of Computing, Engineering and Physical Sciences (CEPS) at the University of Central Lancashire. He is a Member of the Institution of Engineering and Technology (MIET), and has authored and co-authored over 60 journal and conference papers. His research interests include applications of digital signal and image processing including compression, immersive visualisation and robotics; artificial neural networks; and engineering education.

Djamel Ait-Boudaoud is currently the Dean of Faculty of Technology at the University of Portsmouth. With an academic career spanning over 20 years, he has held several senior academic appointments in the UK and Overseas. Prior to joining the University of Portsmouth, he was the head of School of Computing, Engineering and Physical Sciences at the University of Central Lancashire. He is a Chartered Engineer (CEng) and a Fellow of the Institution of Engineering and Technology (FIET). He holds a Ph.D. (1991) from the University of Nottingham where he conducted research in system and circuit integration for Digital Signal Processing (DSP) applications. He has authored and co-authored several research papers and his research interests include VLSI architectures, Image Processing and Computer vision, and Evolutionary algorithms.

Physical Sciences at the University of Central Lancashire. He is a Chartered Engineer (CEng) and a Fellow of the Institution of Engineering and Technology (FIET). He holds a Ph.D. (1991) from the University of Nottingham where he conducted research in system and circuit integration for Digital Signal Processing (DSP) applications. He has authored and co-authored several research papers and his research interests include VLSI architectures, Image Processing and Computer vision, and Evolutionary algorithms.

(Communicated by *Kai-Kuang Ma*)