Computation Controllable Mode Decision and Motion Estimation for Scalable Video Coding

Liang-Wei Zheng, Gwo-Long Li, Mei-Juan Chen, Chia-Hung Yeh, Kuang-Han Tai, and Jian-Sheng Wu

This paper proposes an efficient computation-aware mode decision and search point (SP) allocation algorithm for spatial and quality scalabilities in Scalable Video Coding. In our proposal, a linear model is derived to allocate the computation for macroblocks in enhancement layers by using the rate distortion costs of the base layer. In addition, an adaptive SP decision algorithm is proposed to decide the number of SPs for motion estimation under the constraint of the allocated computation. Experiment results demonstrate that the proposed algorithm allocates the computation resource efficiently and outperforms other works in rate distortion performance under the same computational availability constraint.

Keywords: Scalable Video Coding, computational scalability, motion estimation, mode decision, computation-aware.

Manuscript received June 24, 2012; revised Jan. 16, 2013; accepted Feb. 16, 2013.

This work was supported by National Science Council, Taiwan, under the Grant NSC-99-2221-E-259-019-MY3.

Liang-Wei Zheng (phone: +886 3 863 4072, m9823013@ems.ndhu.edu.tw), Mei-Juan Chen (cmj@mail.ndhu.edu.tw), Kuang-Han Tai (810123002@ems.ndhu.edu.tw), and Jian-Sheng Wu (m9923025@ems.ndhu.edu.tw) are with the Department of Electrical Engineering, National Dong Hwa University, Hualien, Taiwan, ROC.

Gwo-Long Li (glli@itri.org.tw) is with the Department of Video Coding Core Technology, Industrial Technology Research Institute, Hsinchu, Taiwan, ROC.

Chia-Hung Yeh (corresponding author, yeh@mail.ee.nsysu.edu.tw) is with the Department of Electrical Engineering, National Sun Yat-sen University, Taiwan, ROC.

http://dx.doi.org/10.4218/etrij.13.0112.0421

I. Introduction

To satisfy the application heterogeneities, many consumer electronic devices have been developed with various display, processing, and transmission capabilities. However, traditional video coding standards that provide only a single scalable bitstream will become increasingly inapplicable because they are only capable of supporting a specific level of video quality for a target application. Therefore, a video coding standard called Scalable Video Coding (SVC) [1] was standardized by the Joint Video Team (JVT) in 2007 to provide a scalable bitstream that supports different scalabilities, such as spatial, temporal, and signal-to-noise ratio (SNR). Figure 1 shows the application scenario of SVC in which a single SVC encoded bitstream can provide different resolutions and quality levels of video signals for different applications.

In SVC, a single SVC encoded bitstream consists of a base layer (BL) video signal and several enhancement layer (EL) video signals. In the BL, the H.264/AVC [2] encoder is applied to encode video data; therefore, the bitstreams generated by the BL are H.264/AVC-compatible bitstreams. The ELs execute not only the intrinsic prediction modes of H.264/AVC but also the inter-layer motion, inter-layer residual, and inter-layer intra prediction additionally adopted in SVC to further increase the coding efficiency. As a result, the computational complexity of SVC is much higher than that of H.264/AVC; the mode decision and motion estimation still occupy the majority of computation. Several methods have been proposed to enhance the encoding process, addressing issues of speed, efficiency, and consumption [3]-[6]. Additionally, some studies focused on region of interest, error concealment, and error resilience in SVC [7]-[10].

To accelerate the encoding process, researchers have

presented various ways to reduce the amount of computation required in SVC [11]-[14]. However, these works focused on reducing the number of prediction modes and did not consider the computational complexity constraint; therefore, the methods do not apply to battery-equipped hand-held devices.

In addition to computational complexity reduction, efficient and adaptive allocation of the limited computation for encoding macroblocks (MBs) has recently gained more attention. A computation-aware technique adjusts the computation for each MB to match the computational availability constraint. Therefore, under the computation-limited and computation-variant constraints, the video quality resulting from the computation-aware algorithm provides better video quality than does the encoding system without a computation-aware mechanism. Works [15]-[20] proposed complexity-controlling algorithms; however, these algorithms only apply to H.264/AVC. For SVC, they did not consider the inter-layer information, which is very useful in predicting the relationship between the BL and ELs. Wang and others proposed a computation-aware algorithm (hereafter referred to as "Wang's algorithm") for SVC that uses the rate distortion (RD) costs of the BL to decide the priority of the mode in ELs [21]. The work in [21] focused on reducing the number of prediction modes and searching the range to achieve computational scalability.

This paper accomplishes a new computation-aware algorithm for mode decision and motion estimation in SVC. The correlations between inter-layers are observed and analyzed first. Afterward, the frame-level and MB-level computation allocation algorithms are proposed to efficiently distribute the limited computation for MBs in ELs. Regarding RD, simulation results show that our proposed algorithm performs better than methods that do not account for the amount of computation required in the encoding process.

The rest of this paper is organized as follows. In section II, the computation consumption of different prediction modes and the relationship between layers is analyzed. Details of the proposed computation-aware algorithm are presented in section III. Experiment results and concluding remarks are provided in sections IV and V, respectively.

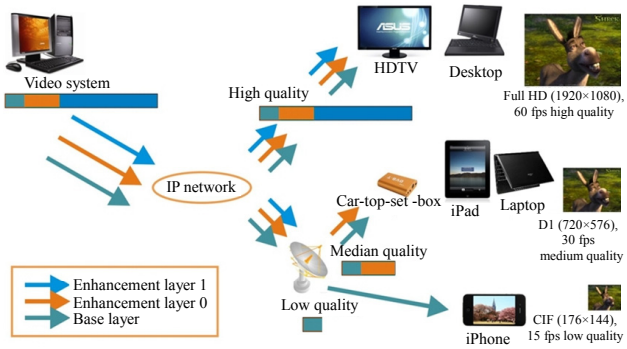


Fig. 1. Application scenario of SVC.

II. Observation and Analysis

1. Computational Complexity Modeling

To develop a computation-aware mode decision algorithm, the computational complexity must be modeled first. This subsection models the computational complexity of each prediction mode by measuring the real coding time. Table 1 exhibits the simulation settings used to obtain our computational complexity modeling results, and those results are shown in Table 2. Table 2 shows the measured computation time for each of the following prediction modes: BLskip, Direct, Mode_16×16, Mode_16×8, Mode_8×16, Mode_8×8, and Intra. The calculation results in "Normalized" by using the coding time of 16×16 as a reference, and the "Approximation" is its rounded value. The computational complexity of each prediction mode is denoted by α. According to Table 2, the respective computation complexity of BLskip and Direct is insignificant compared to that of Mode_16×16, Mode_16×8, and Mode_8×16. The computational complexity of Mode_8×8 is four times that of Mode_16×16. Therefore, hereafter, Approximation (α) represents the basic computational complexity for each prediction mode discussed herein.

2. Analysis of Spatial and Quality Layer Relationship

In SVC, the difference between spatial layers is the frame resolution variation. Therefore, it is expected that there is a strong relationship between spatial layers [11]-[14]. We analyze the relationship between layers for spatial and quality scalabilities by considering the best prediction results and all temporary results during the checking process of all modes. We first define a prediction mode set, which is formed by four BL prediction modes, as shown in (1) through (3).

$$Mode_{\text{Spatial\_BL}} = \left\{ Mode_{\text{Spatial\_}k} \middle| 1 \le k \le 4 \right\},$$
$$Mode_{\text{Quality\_BL}} = \left\{ Mode_{\text{Quality\_}k} \middle| 1 \le k \le 4 \right\}. \quad (1)$$

Table 1. Simulation settings.

Frame resolution	BL: QCIF, EL: CIF
Quantization parameter	BL: 38, EL: 32
# of reference frames	3
Search range	±16
# of coded frames	100

Table 2. Computation time (unit: ms) and related ratio for various coding modes.

Sequence	BLskip	Direct	16×16	16×8	8×16	8×8	Intra	Total_time
Akiyo	2.58	3.73	379.6	388.1	390.7	1,745.3	44.2	2,954.2
Foreman	2.74	4.68	424.7	433.3	434.3	1,687.2	34.6	3,021.5
Football	2.52	5.24	528.9	539.6	541.3	1,919.4	40.6	3,577.6
Bus	1.53	5.65	540.8	551.4	557.0	2,096.2	42.4	3,795.0
City	3.27	5.10	564.5	571.1	578.5	2,128.2	41.7	3,892.3
News	2.85	4.28	576.9	584.5	592.7	2,320.5	40.6	4,122.3
Ice	2.39	6.14	664.7	674.7	687.8	2,458.6	49.0	4,543.3
Mobile	3.13	5.57	496.6	506.9	512.7	2,333.4	44.5	3,902.8
Harbour	3.17	5.87	618.3	629.7	636.4	2,423.4	49.7	4,366.5
Average	2.69	5.14	532.8	542.1	547.9	2,123.6	43.0	3,797.2
Normalized	0.005	0.01	1	1.02	1.03	3.99	0.08	7.12
Approximation (α)	α_{BLskip}	α_{Direct}	$\alpha_{16\times16}$	$\alpha_{16\times8}$	$\alpha_{8\times16}$	$\alpha_{8\times8}$	α_{Intra}	α_{total}
	0	0	1	1	1	4	0	7

For spatial scalability, $Mode_{Spatial\_k}$ is defined as follows.

$$Mode_{Spatial\_k} \in \begin{Bmatrix} Mode\_8\times8, Mode\_8\times4, \\ Mode\_4\times8, Mode\_4\times4 \end{Bmatrix}. \quad (2)$$

For quality scalability, $Mode_{Quality\_k}$ can be expressed as follows.

$$Mode_{Quality\_k} \in \begin{Bmatrix} Mode\_16\times16, Mode\_16\times8, \\ Mode\_8\times16, Mode\_8\times8 \end{Bmatrix}. \quad (3)$$

The defined prediction mode sets $Mode_{Spatial\_BL}$ and $Mode_{Quality\_BL}$ store four prediction modes and their corresponding RD costs in the BL. In addition, the order of the prediction modes inside $Mode_{Spatial\_BL}$ and $Mode_{Quality\_BL}$ must satisfy the following condition:

$$\begin{aligned} J_{Mode_{Spatial\_k}} &\le J_{Mode_{Spatial\_k+1}} \,\middle|\, 1 \le k < 4, \\ J_{Mode_{Quality\_k}} &\le J_{Mode_{Quality\_k+1}} \,\middle|\, 1 \le k < 4, \end{aligned} \quad (4)$$

where J is the RD cost. Equation (4) implies that the prediction modes inside $Mode_{Spatial\_BL}$ and $Mode_{Quality\_BL}$ are sorted in increasing order according to their RD costs.

After defining the prediction mode sets $Mode_{Spatial\_BL}$ and $Mode_{Quality\_BL}$, the prediction results of ELs are first obtained by checking all prediction modes and then cross-checking with the $Mode_{Spatial\_BL}$ and $Mode_{Quality\_BL}$. From the analysis results of the cross-check, we observe some interesting phenomena. First, most MBs in an EL match the same prediction mode order exhibited in $Mode_{Spatial\_BL}$ and $Mode_{Quality\_BL}$ if they are sorted according to their RD costs, as shown in [21]. Therefore, the order of the prediction modes in $Mode_{Spatial\_BL}$ and $Mode_{Quality\_BL}$ can be used to prioritize the mode checking in the EL. However, we also observe that the prediction modes of some MBs in the EL do not reveal the same order exhibited by $Mode_{Spatial\_BL}$ and $Mode_{Quality\_BL}$. In other words, the best prediction mode selected in the BL MB does not imply that its corresponding MB in the EL will select the same prediction mode as the final result. Therefore, we further analyze the relationship between the RD cost of the best prediction mode in the BL and the EL.

Figures 2 and 3 show the analytical results for the spatial scalability, and Fig. 4 exhibits the analytical results of the quality scalability. In these figures, the 0 in the horizontal axis means BLskip, the best inter prediction mode in the EL is $BestMode_{EL}$, the average RD cost of the best prediction modes in the BL is $J_{avg\_Mode\_1}$, and (5) defines $Mode'_{Spatial\_BL}$ and $Mode'_{Quality\_BL}$.

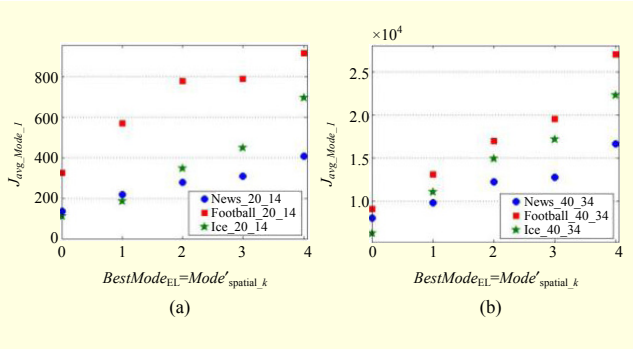


Fig. 2. Relationship between RD cost of BL block and best MB mode in EL (BL: QCIF, EL: CIF): (a) QP_BL=20 and QP_EL=14 and (b) QP_BL=40 and QP_EL=34.

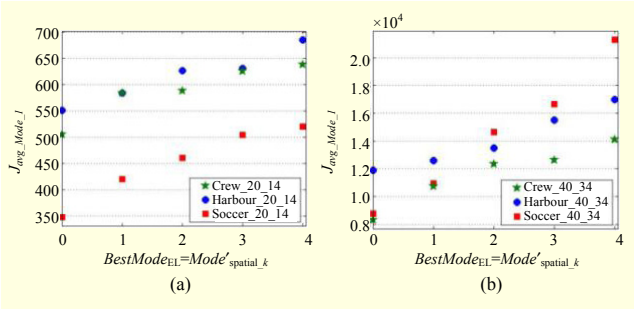


Fig. 3. Relationship between RD cost of BL block and best MB mode in EL (BL: CIF, EL: 4CIF): (a) QP_BL=20 and QP_EL=14 and (b) QP_BL=40 and QP_EL=34.

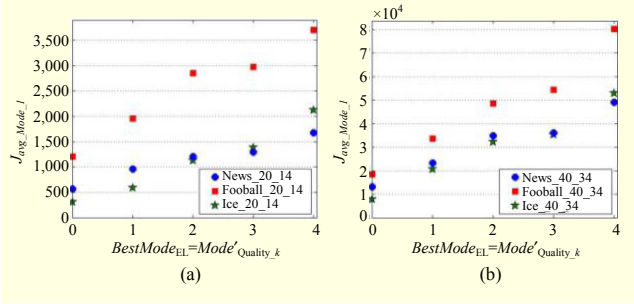


Fig. 4. Relationship between RD cost of BL block and best MB mode in EL (BL: CIF, EL: CIF): (a) QP_BL=20 and QP_EL=14 and (b) QP_BL=40 and QP_EL=34.

$$Mode'_{\text{Spatial\_BL}} = \left\{ Mode'_{\text{Spatial\_}k} \middle| 1 \le k \le 4 \right\},$$
$$Mode'_{\text{Quality\_BL}} = \left\{ Mode'_{\text{Quality\_}k} \middle| 1 \le k \le 4 \right\}. \quad (5)$$

For spatial scalability, the $Mode'_{\text{Spatial\_}k}$ is defined as follows due to the spatial resolution changing.

$$Mode'_{\text{Spatial\_}k} \in \left\{ \begin{matrix} \text{Mode\_16×16, Mode\_16×8,} \\ \text{Mode\_8×16, Mode\_8×8} \end{matrix} \right\}. \quad (6)$$

For quality scalability, the $Mode'_{\text{Quality\_}k}$ can be expressed as

$$Mode'_{\text{Quality\_}k} = Mode_{\text{Quality\_}k}. \quad (7)$$

As shown in Figs. 2 through 4, the best prediction mode of the EL belongs to the latter part of $Mode'_{\text{Spatial\_BL}}$, and $Mode'_{\text{Quality\_BL}}$, its RD cost of the corresponding MB in BL, is much larger. For example, if a prediction mode set is obtained as $Mode_{\text{Spatial\_BL}}$={Mode_4×8, Mode_4×4, Mode_8×4, Mode_8×8}, the $Mode'_{\text{Spatial\_BL}}$ is $Mode'_{\text{Spatial\_BL}}$={Mode_8×16, Mode_8×8, Mode_16×8, Mode_16×16}. However, if the best prediction mode of an MB in the EL is Mode_16×16, which is the last prediction mode in $Mode'_{\text{Spatial\_BL}}$, its RD cost of the corresponding MB in the BL is much larger. Therefore, this observation inspires us to use the RD cost of the BL as the reference of computation allocation for MBs in ELs.

III. Proposed Computation-Aware Mode Decision and Motion Estimation Algorithm

Figure 5 shows a flowchart of the proposed computation-aware algorithm, which consists of two major parts: frame-level computation allocation and MB-level computation allocation. The frame-level computation allocation is in charge of allocating the computation for each MB from the global perspective. Then, the MB-level computation allocation algorithm uses the available computation to allocate the resource of mode decision and motion estimation processes.

1. Frame-Level Computation Allocation

The computation allocation in this level is mainly composed of three steps. The first step is the computation allocation that calculates the computation for each current MB in the EL by considering the relationship of the RD costs in the BL. In the second step, the computation regulation process is applied to regulate the calculated computation in the first step so that the allocated computation for each MB does not exceed the maximum available computation allowance. Therefore, the computation process can be further applied to the MBs, which is helpful in improving the prediction results.

A. Computation Allocation

Regarding video content, it is expected that the higher RD costs are associated with MBs with complex or high motion video content. Allocating additional computation to such MBs usually results in better coding performance. Therefore, we can make use of the relationship between RD costs to guide the computation allocation process. Our observation in section II shows us that a linear relationship exists between the RD costs of the BL and the EL. Therefore, in our proposed computation allocation approach, the computation of each current MB in the EL is calculated by a linear equation that considers the RD cost relationship between the blocks of its corresponding BL MB, as (8) and (9) for spatial scalability and for quality scalability, respectively:

$$C_{\text{EL\_}i} = \frac{J_{\text{Blk\_}i}}{\sum_{j=0}^{N_{\text{BL}}-1} J_{\text{Blk\_}j}} \times adj \times N_{\text{EL}} \times \alpha_{\text{total}}, \quad (8)$$

$$C_{\text{EL\_}i} = \frac{J_{\text{MB}_{\text{BL\_}i}}}{\sum_{j=0}^{N_{\text{EL}}-1} J_{\text{MB}_{\text{BL\_}j}}} \times adj \times N_{\text{EL}} \times \alpha_{\text{total}}, \quad (9)$$

where $C_{\text{EL\_}i}$ is the allocated computation of the i-th MB in the EL, J stands for the RD cost, $\text{MB}_{\text{BL\_}i}$ refers to the corresponding BL MB of the current i-th MB in the EL, Blk_j is the BL block with index j in the BL, N_{EL} stands for the number of MBs in the EL, and N_{BL} is the number of 8×8

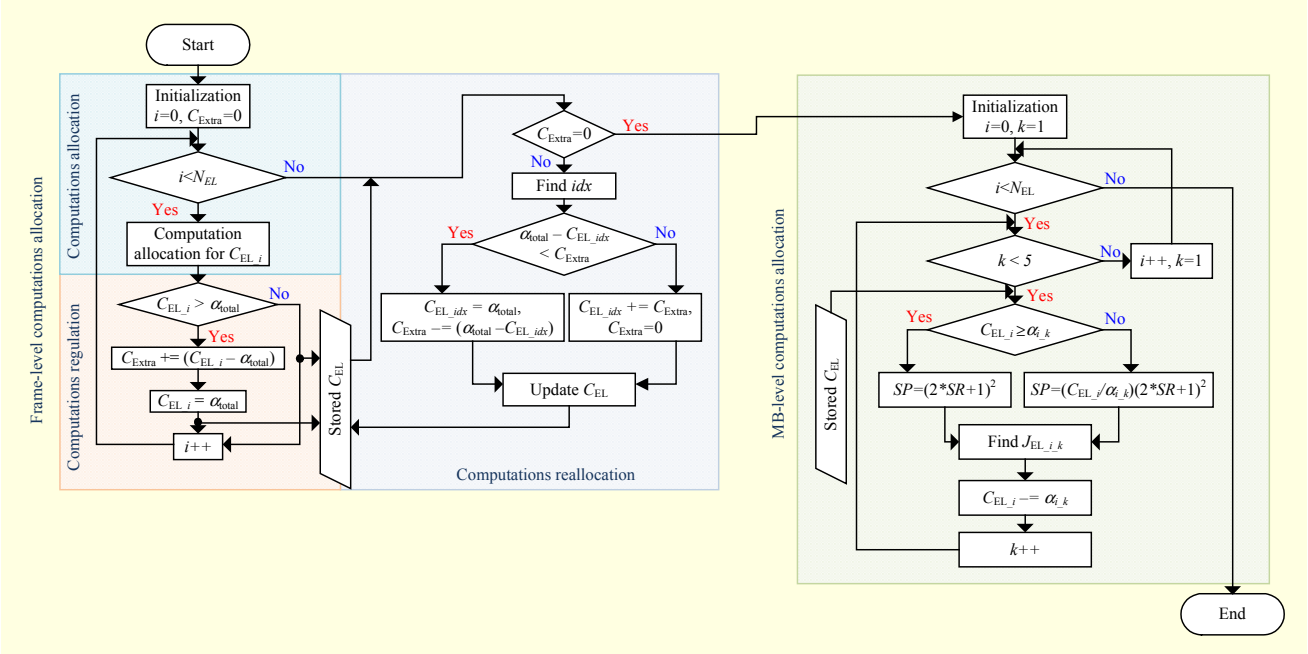


Fig. 5. Flowchart of proposed computation-aware mode decision algorithm.

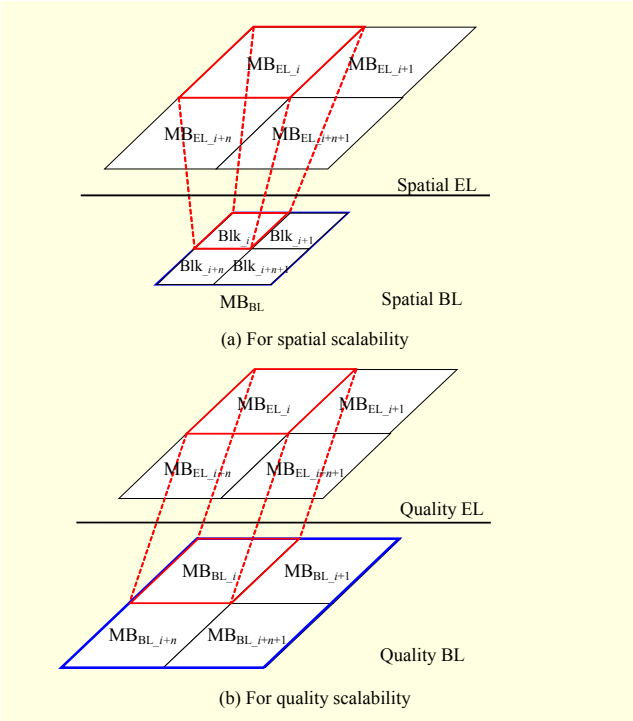


Fig. 6. Illustration of calculating C_{EL_i} (a) spatial scalability and (b) quality scalability.

subblocks in the BL. Since the frame resolution ratio between successive spatial layers is 2 in our proposed algorithm, the number of its corresponding 8×8 subblocks in the BL is 4 for the current MB in the spatial EL. The parameter *adj* refers to the computation adjusting parameter, the value of which ranges from 0 to 1; if the value is 1, there is no computational constraint. Figure 6 illustrates the required elements for calculating C_{EL_i}. The allocation of C_{EL_i} means that a portion of RD costs of the corresponding BL block is used to calculate the required computation of the current MB in the EL.

B. Computation Regulation

After calculating C_{EL_i}, C_{EL_i} may exceed α_{total}, the actual required computation units. In this case, the over allocated computation of C_{EL_i} must be regulated and gathered for the further computation distribution. Therefore, the gathered computation C_{Extra} can be obtained as follows.

$$C_{Extra} = \sum_{i=0}^{N_{EL}-1} C_{ext_i} \Big| C_{ext_i} = \begin{cases} C_{EL\_i} - \alpha_{total}, & \text{if } C_{EL\_i} > \alpha_{total}, \\ 0, & \text{otherwise.} \end{cases} \quad (10)$$

Afterward, the regulated computation of C_{EL_i} becomes

$$C_{EL\_i} = \begin{cases} \alpha_{total}, & \text{if } C_{EL\_i} > \alpha_{total}, \\ C_{EL\_i}, & \text{otherwise.} \end{cases} \quad (11)$$

C. Computation Reallocation

So far, the computation of each C_{EL_i} has been calculated by the previous two steps. However, since the computation process can be further applied, this step tries to distribute the computation of C_{Extra} to other MBs that require additional computation to perform better. According to section II, it is

beneficial to allocate computation to the MBs whose corresponding BL blocks have larger RD costs. As a result, the block with the largest RD cost in the BL is selected first, and the free computation of C_{Extra} is allocated to the corresponding MBs in the EL. The following equations (12) and (13) for spatial scalability and for quality scalability, respectively, show that the MB with *idx* is selected to reallocate computation.

$$idx = \underset{i}{\operatorname{argmax}} \left\{ J_{\text{Blk}\_i} \middle| C_{\text{EL}\_i} < \alpha_{\text{total}} \right\}, \tag{12}$$

$$idx = \underset{i}{\operatorname{argmax}} \left\{ J_{\text{MB}_{\text{BL}\_i}} \middle| C_{\text{EL}\_i} < \alpha_{\text{total}} \right\}. \tag{13}$$

The selected MB with *idx* in this iteration is not considered in the next iteration. After the selection of (12) or (13), the computation of the selected MB in the EL can be recalculated:

$$C_{\text{EL}\_idx} = \begin{cases} \alpha_{\text{total}}, & \text{if } \alpha_{\text{total}} - C_{\text{EL}\_idx} < C_{\text{Extra}}, \\ C_{\text{EL}\_idx} + C_{\text{Extra}}, & \text{otherwise.} \end{cases} \tag{14}$$

The extra free computation is updated as follows:

$$C_{\text{Extra}} = \begin{cases} C_{\text{Extra}} - \left(\alpha_{\text{total}} - C_{\text{EL}_{idx}}\right), & \text{if } \alpha_{\text{total}} - C_{\text{EL}\_idx} < C_{\text{Extra}}, \\ 0, & \text{otherwise.} \end{cases} \tag{15}$$

Equations (12) through (15) are executed iteratively until the free computation C_{Extra} becomes empty. Once all free computation is released to the MBs, the MB-level computation allocation and motion estimation are initiated for each MB along with the allocated computation C_{EL_i}.

2. MB-Level Computation Allocation

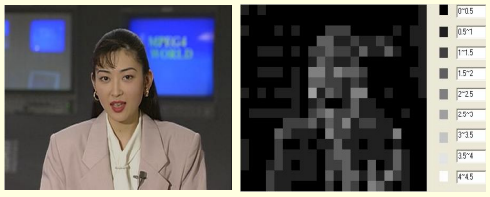
After frame-level computation allocation, MB-level computation allocation is applied for each current MB in the EL. The main idea of MB-level computation allocation is to determine the number of search points that can be used for each current MB for motion estimation under the constraint of the allocated computation C_{EL_i}. Figure 7 shows the pseudocode of the proposed MB-level computation allocation algorithm. In the proposal, the spiral search order [22] is adopted in the motion estimation process. The main idea of our proposed algorithm is briefly described as follows. First, the prediction mode set of the *i*-th MB in the EL is derived as [21] described, which indicates the mode prediction order. Afterward, for each prediction mode *k* within a prediction mode set, the associated computation unit α_{i_k} is compared to C_{EL_i} to determine if the remaining C_{EL_i} is sufficient to complete the operation of the mode. If the remaining C_{EL_i} is sufficient, the *SP* is assigned to $(2\times SR+1)^2$. Otherwise, *SP* is assigned to $C_{\text{EL}\_i}/\alpha_{i\_k}\times(2\times SR+1)^2$, which stands for the remaining computation allowance. The remaining C_{EL_i} is allocated according to the amount of computation required by α_{i_k} and the search range size. Once *SP* has been decided, the motion estimation is executed with the *SP*, and C_{EL_i} is updated. If C_{EL_i} is not empty, the next prediction mode *k*+1 is examined by the previous steps. Otherwise, the prediction of the current *i*-th MB is terminated immediately.

```
Data: C_EL_i, allocated computation of the i-th MB in EL
Data: α_i_k, basic computation unit of the k-th prediction mode within
      prediction mode list corresponding to i-th MB in EL
Data: N_EL, number of MBs in EL
Data: SR, default search range size
Data: SP, determined number of search points
for i→0 to N_EL−1 do
   k←1;
   do
      if (C_EL_i≥α_i_k)
         SP ← (2×SR+1)^2;
         find the best result with SP;
         C_EL_i ← C_EL_i−α_i_k;
         k ← k+1;
      else
         SP ← C_EL_i/α_i_k×(2×SR+1)^2;
         find the best result with SP;
         break;
      end
   while (k<5)
end
```

Fig. 7. Pseudocode of proposed MB-level computation allocation algorithm.

IV. Simulation Results

The proposed computation-aware mode decision algorithm is implemented on the H.264/SVC reference software JSVM 9.18 [23]. Figure 8 shows the distribution of the allocated computation of our proposed algorithm for different test sequences. In Fig. 8, the original frame is shown on the left-hand side, and the allocated computation is shown on the other; our proposed algorithm efficiently allocates a larger amount of



(a) Akiyo (32nd frame)

(b) Foreman (12th frame)

Fig. 8. Distribution of allocated computation in EL (10% computation, BL QCIF, QP=38; EL CIF, QP=25, *adj*=0.7) for (a) Akiyo and (b) Foreman sequences.

Table 3. RD performance comparison of spatial scalability: (a) BL: QCIF and EL: CIF (JSVM, BLskip, and Proposed with 10% computation, BL QCIF, QP=38; EL CIF, QP=25) and (b) BL: CIF and EL: 4CIF (JSVM, BLskip, and Proposed with 10% computation, BL CIF, QP=38; EL 4CIF, QP=25).

(a)

Sequence	JSVM		BLskip		Proposed	
	PSNR (dB)	Bitrate (kbps)	ΔPSNR (dB)	ΔBitrate (%)	ΔPSNR (dB)	ΔBitrate (%)
Akiyo	43.947	151.046	–0.855	76.86%	–0.121	3.44%
Bus	37.922	1777.94	–1.373	74.92%	–0.194	6.22%
City	39.519	628.50	–2.496	215.25%	–0.245	10.37%
Football	39.036	2174.20	–1.153	31.59%	–0.123	5.63%
Foreman	39.816	752.35	–1.759	102.57%	–0.216	6.94%
Harbour	37.283	1893.07	–0.365	23.48%	–0.116	2.58%
ICE	42.811	625.24	–1.595	59.01%	–0.283	6.85%
Mobile	37.549	1865.89	–1.173	81.67%	–0.180	4.98%
News	42.616	361.19	–0.901	46.02%	–0.167	4.75%
Average	40.055	1136.60	–1.297	79.04%	–0.183	5.75%

(b)

Sequence	JSVM		BLskip		Proposed	
	PSNR (dB)	Bitrate (kbps)	ΔPSNR (dB)	ΔBitrate (%)	ΔPSNR (dB)	ΔBitrate (%)
City	37.893	3548.26	–0.841	107.79%	–0.070	2.93%
Crew	39.385	4740.75	–1.120	58.85%	–0.037	2.64%
Harbour	37.764	7256.19	–0.610	22.45%	–0.082	1.54%
ICE	42.247	1821.73	–1.264	66.06%	–0.103	3.72%
Soccer	39.292	4185.36	–1.415	91.17%	–0.062	3.22%
Average	39.316	4310.46	–1.050	69.26%	–0.071	2.81%

the computation allowance to the MBs that belong to the moving area.

Table 3 exhibits a comparison of the RD of combinations of different spatial layer resolutions for spatial scalability. In Table 3, the column labeled "BLskip" contains values reflecting cases in which only BLskip is used for the MBs in ELs; BLskip has the lowest computational complexity, but RD performance drops significantly. For low resolution sequences, as shown in Table 3(a), the average PSNR degradation of BLskip and our proposed algorithm is 1.297 dB and 0.183 dB, respectively. Regarding bitrate, our proposed algorithm only results in a 5.75% increase, but the BLskip produces a 79.04% increase. For high resolution sequences, as shown in Table 3(b), the RD performance drop is slightly less than that of low resolution sequences. Using our proposed algorithm results in an average PSNR decrease of only 0.071 dB and an average bitrate increase of 2.81%. However, although high resolution sequences have less of an RD performance drop in general, using BLskip results in a PSNR decrease of 1.050 dB and a bitrate increase of 69.26%. Regarding quality scalability, the RD performances are shown in Table 4. As shown in this table, the average PSNR degradation and bitrate increase of our proposed algorithm is 0.126 dB and 1.42%, respectively. In summary, although BLskip has the lowest computation complexity, the RD performance drops significantly compared to that of our proposed algorithm.

Table 4. RD performance comparison of quality scalability (JSVM, BLskip, and Proposed with 10% computation, BL CIF, QP=38; EL CIF, QP=25).

Sequence	JSVM		BLskip		Proposed	
	PSNR (dB)	Bitrate (kbps)	ΔPSNR (dB)	ΔBitrate (%)	ΔPSNR (dB)	ΔBitrate (%)
Akiyo	43.796	264.62	–0.885	33.64%	–0.048	0.33%
Bus	37.834	1934.77	–0.468	24.91%	–0.144	2.43%
City	39.377	782.19	–1.228	41.27%	–0.163	1.25%
Football	38.907	2358.13	–1.037	20.24%	–0.119	2.66%
Foreman	39.757	896.67	–1.225	36.17%	–0.171	2.45%
Harbour	37.199	2005.92	–0.162	10.75%	–0.093	0.66%
ICE	42.718	704.28	–0.754	34.75%	–0.179	1.46%
Mobile	37.472	2054.16	–0.458	15.12%	–0.121	1.66%
News	42.604	505.92	–0.646	21.56%	–0.099	0.56%
Average	39.963	1278.52	–0.763	26.49%	–0.126	1.42%

Table 5. RD performance comparison (26% computation, BL QCIF, QP=38; EL CIF, QP=25).

Sequence	JSVM		Wang [21] (only PBMD)		Proposed	
	PSNR (dB)	Bitrate (kbps)	ΔPSNR (dB)	ΔBitrate (%)	ΔPSNR (dB)	ΔBitrate (%)
Akiyo	43.947	151.05	–0.166	6.36%	–0.046	1.31%
Bus	37.922	1777.94	–0.212	8.10%	–0.118	2.64%
City	39.519	628.50	–0.267	12.40%	–0.126	5.13%
Football	39.036	2174.20	–0.137	5.82%	–0.083	2.95%
Foreman	39.816	752.35	–0.315	13.06%	–0.117	6.92%
Harbour	37.283	1893.07	–0.121	3.22%	–0.074	1.24%
ICE	42.811	625.24	–0.353	12.32%	–0.124	1.98%
Mobile	37.549	1865.89	–0.194	6.48%	–0.107	2.40%
News	42.616	361.19	–0.203	8.25%	–0.077	1.61%
Average	40.055	1133.60	–0.219	8.45%	–0.097	2.76%

To show the efficiency of the proposed computation

Table 6. RD performance comparison (17% computation, BL QCIF, QP=38; EL CIF, QP=25).

Sequence	JSVM		Wang [21] (PBMD+ABME)		Proposed	
	PSNR (dB)	Bitrate (kbps)	ΔPSNR (dB)	ΔBitrate (%)	ΔPSNR (dB)	ΔBitrate (%)
Akiyo	43.947	151.05	–0.167	6.99%	0.076	2.09%
Bus	37.922	1777.94	–0.289	11.35%	–0.153	4.07%
City	39.519	628.50	–0.387	13.64%	–0.180	7.43%
Football	39.036	2174.20	–0.187	8.72%	–0.103	4.10%
Foreman	39.816	752.35	–0.365	13.49%	–0.162	8.58%
Harbour	37.283	1893.07	–0.121	3.29%	–0.095	1.83%
ICE	42.811	625.24	–0.373	13.48%	–0.189	3.76%
Mobile	37.549	1865.89	–0.247	7.82%	–0.143	3.50%
News	42.616	361.19	–0.240	12.44%	–0.114	2.78%
Average	40.055	1133.60	–0.264	10.14%	–0.118	4.24%

allocation algorithm, we compare our proposed algorithm with Wang's algorithm [21] under the same computational availability constraint. In [21], priority-based mode decision (PBMD) and activity-based motion estimation (ABME) were compared regarding RD. Table 5 shows a comparison of the RD performances of PBMD and our proposed algorithm under the constraint of 26% computational availability. From this table, the average PSNR degradations of PBMD and our proposed algorithm are 0.219 dB and 0.097 dB, respectively. Regarding bitrate, our proposed algorithm results in a 2.76% increase, and PBMD results in an 8.45% increase. As shown in Table 6, the average PSNR degradations of PBMD+ABME and our proposed algorithm are 0.264 dB and 0.118 dB, respectively, under the constraint of 17% computational availability. Regarding bitrate, our proposed algorithm results in a 4.24% increase, and PBMD+ABME results in a 10.14% increase.

Figure 9 exhibits a comparison of PBMD, PBMD+ABME, and our proposed algorithm regarding coding complexity versus PSNR, bitrate, and RD costs for spatial scalability. According to the results, our proposed algorithm only needs 10% of the computation allowance to reach RD performance stabilization as opposed to the 20% needed by PBMD and PBMD+ABME. The comparison featured in Fig. 10 is similar to that of Fig. 9 but specific to quality scalability. For a very low computational availability condition, our proposed algorithm has significant RD performance improvement over the BLskip approach. Furthermore, to achieve high RD performance, our proposed algorithm only needs approximately 10% computation support. Figure 11 exhibits a

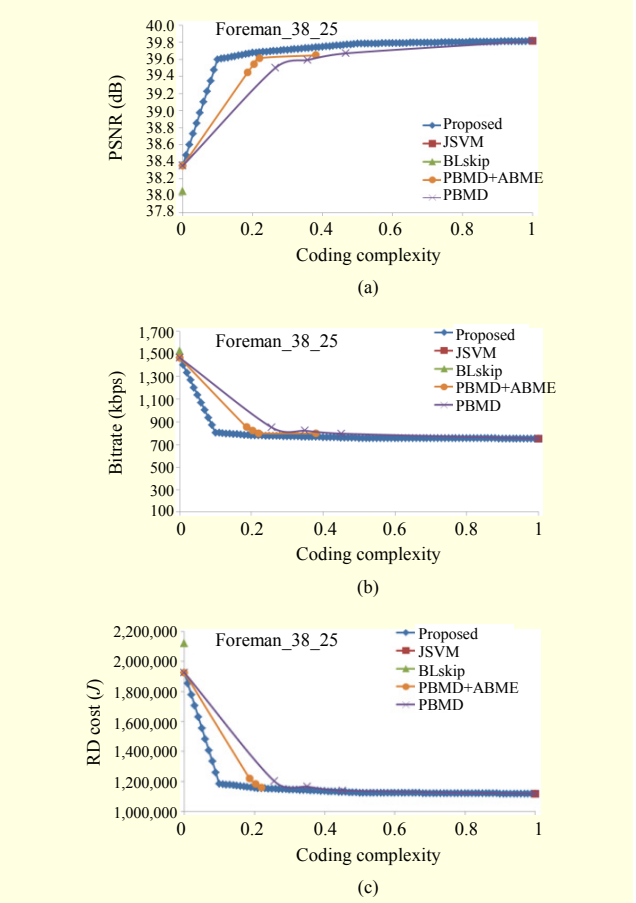


Fig. 9. Coding complexity vs. RD performance comparison of spatial scalability for Foreman sequence: (a) PSNR, (b) bitrate, and (c) rate distortion cost (BL QP=38 and EL QP=25).

RD-complexity comparison for different algorithms in spatial scalability, which reveals that our proposed algorithm can achieve much better RD performance results than other algorithms under the same computational availability constraint. Additionally, our proposed algorithm requires much less computation than other algorithms.

Tables 7 through 9 show the Structural Similarity Index Metric (SSIM) [24] comparisons of all algorithms discussed in this paper. SSIM is one of the objective quality assessment metrics. The main difference between SSIM and PSNR is that SSIM does not estimate the perceived errors for the quantification of image degradations but considers them as changes in structural information variation. The higher the SSIM value is, the more consistent the perception of the human eye is. In evaluating coding performance, the results obtained from considering both PSNR and SSIM together appear to be better than the results obtained from considering only one of these values. Tables 7 and 8 show the SSIM comparisons of various algorithms for spatial scalability. In SSIM, our

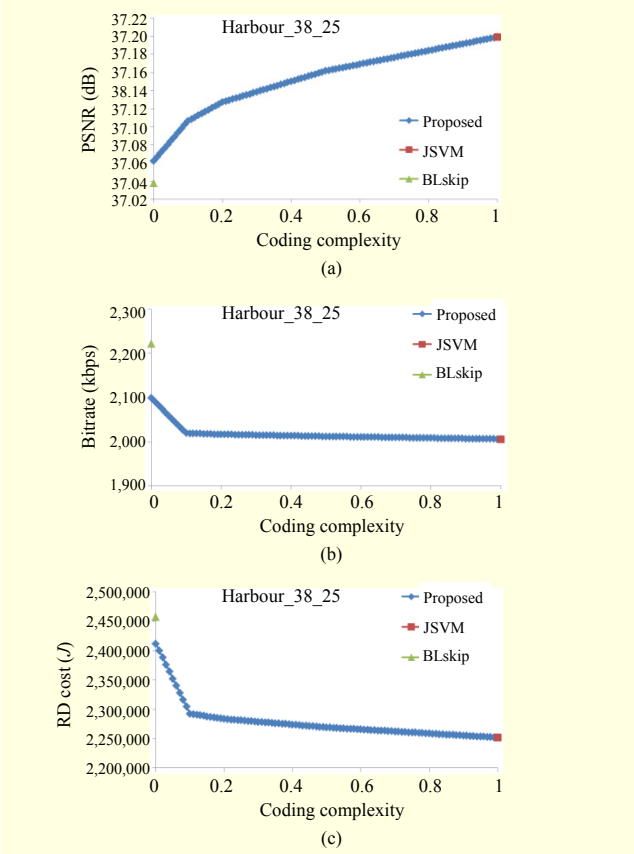


Fig. 10. Coding complexity vs. RD performance comparison of quality scalablity for Harbour sequence: (a) PSNR, (b) bitrate, and (c) rate distortion cost (BL QP=38 and EL QP=25).

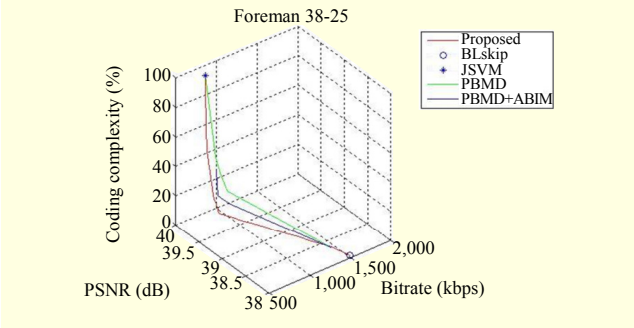


Fig. 11. RD-complexity comparison of spatial scalability for Foreman sequence (BL QCIF and EL CIF).

proposed method is very similar to JSVM, and all the outcomes of the proposed algorithm are better than those of BLskip and Wang's algorithm. Similar results of quality scalability can be seen in Table 9.

A subjective quality comparison of BLskip, JSVM, and our proposed algorithm under different computation contraints is featured in Fig. 12. For the lowest computational complexity case, BLskip, the image is notably blurry, as seen in Fig. 12(a).

Table 7. SSIM comparison of spatial scalability for various computations and algorithms (BL QCIF, QP=38; EL CIF, QP=25).

Sequence	JSVM	BLskip	10%	17%		26%	
			Pro.	Wang's	Pro.	Wang's	Pro.
Akiyo	0.98105	0.97930	0.98084	0.98071	0.98092	0.98076	0.98099
Bus	0.97376	0.96049	0.97220	0.97115	0.97257	0.97218	0.97290
City	0.97434	0.95119	0.97334	0.97263	0.97355	0.97321	0.97389
Football	0.96109	0.95230	0.95945	0.95775	0.95983	0.95981	0.95997
Foreman	0.96686	0.95250	0.96537	0.96475	0.96564	0.96500	0.96601
Harbour	0.97884	0.97613	0.97818	0.97804	0.97828	0.97805	0.97842
ICE	0.98216	0.97971	0.98138	0.98114	0.98168	0.98118	0.98190
Mobile	0.98489	0.97888	0.98411	0.98380	0.98429	0.98396	0.98448
News	0.97985	0.97714	0.97937	0.97909	0.97955	0.97932	0.97967
Average	0.97587	0.96752	0.97492	0.97434	0.97515	0.97483	0.97536

Table 8. SSIM comparison of spatial scalability for various algorithms (BL CIF, QP=38; EL 4CIF, QP=25).

Sequence	JSVM	BLskip	Proposed (10%)
City	0.95815	0.94727	0.95405
Crew	0.93929	0.93743	0.93859
Harbour	0.97125	0.96903	0.97022
ICE	0.96765	0.96583	0.96663
Soccer	0.94286	0.94233	0.94285
Average	0.95584	0.95238	0.95447

Table 9. SSIM comparison of quality scalability for various algorithms (BL CIF, QP=38; EL CIF, QP=25).

Sequence	JSVM	BLskip	Proposed (10%)
Akiyo	0.97988	0.97895	0.97975
Bus	0.97208	0.96328	0.97097
City	0.97325	0.96085	0.97228
Football	0.96032	0.95340	0.95869
Foreman	0.96626	0.95686	0.96499
Harbour	0.97826	0.97647	0.97770
ICE	0.98111	0.97938	0.98033
Mobile	0.98421	0.98144	0.98365
News	0.97933	0.97755	0.97898
Average	0.97497	0.96980	0.97415

However, for our proposed algorithm, the image quality is improved with an increase of the computational allowance. Regarding subjective quality in the case that only 10% of the computation allowance is used, the result obtained using

Fig. 12. Subjective quality comparison: (a) BLskip with 0.005% coding complexity and PSNR=32.40 dB, (b) proposal with 0.1% coding complexity and PSNR=32.90 dB, (c) proposal with 10% coding complexity and PSNR=33.35 dB, and (d) JSVM with 100% coding complexity and PSNR=33.40 dB.

JSVM is indistinguishable from that obtained using our proposed algorithm.

V. Conclusion

An efficient computation controllable algorithm was proposed in this paper. Through the RD costs obtained from the BL, a linear computation allocation according to the RD costs was derived to perform the mode prediction of the EL. The mode was prioritized according to the results in the BL, and the search range was adjusted by the remaining computation. Simulation results demonstrate that our proposed algorithm has less degradation in RD performance under certain computational availability constraints compared with algorithms proposed in other works. Also, under the condition of the same RD performance, our proposed algorithm significantly decreases the amount of computation required.

References

[1] Joint Video Team (JVT), *Recommendation ITU-T H.264 | International Standard ISO/IEC 14496-10 / Amendment 3: Scalable Video Coding*, JVT-X201, 2007.

[2] Joint Video Team (JVT), *Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification (ITU-T Rec. H.264 | ISO/IEC 14496-10 AVC)*, ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, JVT-G050r1, 2003.

[3] B.-G. Kim and S.-K. Song, "Enhanced Inter Mode Decision Based on Contextual Prediction for P-Slices in H.264/AVC Video Coding," *ETRI J.*, vol. 28, no. 4, Aug. 2006, pp. 425-434.

[4] B.-G. Kim, J.-H. Kim, and C.-S. Cho, "A Fast Intra Skip Detection Algorithm for H.264/AVC Video Encoding," *ETRI J.*, vol. 28, no. 6, Dec. 2006, pp. 721-731.

[5] B.-G. Kim, J.-H. Kim, and C.-S. Cho, "Fast Inter Mode Decision Algorithm Based on Macroblock Tracking in H.264/AVC Video," *ETRI J.*, vol. 29, no. 6, Dec. 2007, pp. 736-744.

[6] H.-S. Kang, S.-W. Lee, and H.G. Hosseini, "Probability Constrained Search Range Determination for Fast Motion Estimation," *ETRI J.*, vol. 34, no. 3, June 2012, pp. 369-378.

[7] T.M. Bae et al., "Multiple Region-of-Interest Support in Scalable Video Coding," *ETRI J.*, vol. 28, no. 2, Apr. 2006, pp. 239-242.

[8] G.H. Park and K. Kim, "Adaptive Scanning Method for Fine Granularity Scalable Video Coding," *ETRI J.*, vol. 26, no. 4, Aug. 2004, pp. 332-343.

[9] C.S. Park, T.S. Wang, and S.J. Ko, "Error Concealment Using Inter-layer Correlation for Scalable Video Coding," *ETRI J.*, vol. 29, no. 3, June 2007, pp. 390-392.

[10] X.-F. Li, N. Zhou, and H.-S. Liu, "Joint Source/Channel Coding Based on Two-Dimensional Optimization for Scalable H.264/AVC Video," *ETRI J.*, vol. 33, no. 2, Apr. 2011, pp. 155-162.

[11] C.H. Yeh et al., "Fast Mode Decision Algorithm for Scalable Video Coding Using Bayesian Theorem Detection and Markov Process," *IEEE Trans. Circuits and Syst. Video Technol.*, vol. 20, no. 4, Apr. 2010, pp. 563-574.

[12] P.C. Wang et al., "Efficient Mode Decision Algorithm Based on Spatial, Temporal, and Inter-Layer Rate Distortion Correlation Coefficients for Scalable Video Coding," *ETRI J.*, vol. 32, no. 4, Aug. 2010, pp. 577-587.

[13] B.S. Lee et al., "A Fast Mode Selection Scheme in Inter-Layer Prediction of H.264 Scalable Extension Coding," *Proc. IEEE Int. Symp. Broadband Multimedia Syst. Broadcast.*, 2008, pp. 1-5.

[14] H.C. Lin, W.H. Peng, and H.M. Hang, "Fast Context-Adaptive Mode Decision Algorithm for Scalable Video Coding with Combined Coarse-Grain Quality Scalable (CGS) and Temporal Scalability," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 5, May 2010, pp. 732-748.

[15] M.Y. Chiu and W.C. Siu, "Computationally-Scalable Motion Estimation Algorithm for H.264/AVC Video Coding," *IEEE Trans. Consum. Electron.*, vol. 56, no. 2, July 2010, pp. 895-903.

[16] C.S. Kannangara, I.E. Richardson, and A.J. Miller, "Computational Complexity Management of a Real-Time H.264/AVC Encoder," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 9, Sept. 2008, pp. 1191-1200.

[17] C.Y. Chen et al., "One-Pass Computation-Aware Motion Estimation with Adaptive Search Strategy," *IEEE Trans. Multimedia*, vol. 8, no. 4, Aug. 2006, pp. 698-706.

[18] P.L. Tai et al., "Computation-Aware Scheme for Software-Based Block Motion Estimation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 9, Sept. 2003, pp. 901-913.

[19] W. Lin et al., "A Computation Control Motion Estimation Method

for Complexity-Scalable Video Coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 11, Nov. 2010, pp. 1533-1543.
[20] C.H. Chen et al., "Multipath Flatted-Hexagon Search for Block Motion Estimation," *J. Inf. Hiding Multimedia Signal Process.*, vol. 1, no. 2, Apr. 2010, pp. 110-131.
[21] T.H. Wang et al., "Computation-Scalable Algorithm for Scalable Video Coding," *IEEE Trans. Consum. Electron.*, vol. 57, no. 3, Aug. 2011, pp. 1194-1202.
[22] R.W. Hall, "Efficient Spiral Search in Bounded Spaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-4, no. 2, Mar. 1982, pp. 208-215.
[23] J. Reichel, H. Sclwarz, and M. Wien, "Joint Scalable Video Model JSVM-9," Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6), JVT-V203, 2007.
[24] Z. Wang et al., "Image Quality Assessment: from Error Visibility to Structural Similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, Apr. 2004, pp. 600-612.

Liang-Wei Zheng received his BS from the Department of Communication Engineering, National Penghu University of Science and Technology, Penghu, Taiwan, in 2009 and his MS from the Department of Electrical Engineering, National Dong Hwa University, Hualien, Taiwan, in 2011. His research interests include computation-aware mode decision and motion estimation for video coding.

Gwo-Long Li received his BS from the Department of Computer Science and Information Engineering, Shu-Te University, Kaohsiung, Taiwan, in 2004, his MS from the Department of Electrical Engineering, National Dong Hwa University, Hualien, Taiwan, in 2006, and his PhD from the Department of Electronics Engineering, National Chiao-Tung University, Hsinchu, Taiwan, in 2011. In 2006, he received the Excellent Master Thesis Award from the Institute of Information and Computer Machinery. He is currently an engineer at the Industrial Technology Research Institute (ITRI), Hsinchu, Taiwan. His research is focused on video signal processing and its VLSI architecture design.

Mei-Juan Chen received her BS, MS, and PhD in electrical engineering from National Taiwan University, Taipei, Taiwan, in 1991, 1993, and 1997, respectively. She was an assistant professor (1997-2000) and an associate professor (2000-2005) in the Department of Electrical Engineering, National Dong Hwa University, Hualien, Taiwan. Since August 2005, she has been a professor in the Department of Electrical Engineering, National Dong Hwa University. Her research topics include image/video processing, video compression, motion estimation, error concealment, and video transcoding. She was the recipient of the 2005 K.T. Li Young Researcher Award from the ACM Taipei/Taiwan Chapter for her contribution to video signal codec technique. This award is given annually to only one person under the age of 36 conducting research in Taiwan. In 2006, she received the Distinguished Young Engineer Award from the Chinese Institute of Electrical Engineering, Taiwan. In 2005 and 2012, she received the Jun S. Huang Memorial Foundation Best Paper Award.

Chia-Hung Yeh received his BS and PhD in electrical engineering from the National Chung-Cheng University, Chiayi, Taiwan, in 1997 and 2002, respectively. He has been with the Department of Electrical Engineering, National Sun Yat-sen University (NSYSU), as an associate professor since February 2010. He served on the editorial board of the *Journal of Multimedia* from 2006 to 2008 and has been on the editorial boards of the *EURASIP Journal on Advances in Signal Processing* and the *Journal of Visual Communication and Image Representation*. Dr. Yeh has authored more than 120 published technical international conference and journal papers and holds 40 patents in the USA, Taiwan, and China. His research interests include video compression, video communication, multimedia database management, and image/audio/video signal processing. Dr. Yeh received the 2002 Outstanding Student Award from CCU, the 2007 Young Investigator Award, the 2010–2011 Outstanding Teaching Award, and the 2011 Excellent Teaching Award from NSYSU. He also received the Distinguished Young Engineer Award from the Chinese Institute of Electrical Engineering in 2011.

Kuang-Han Tai received his MS from the Department of Electrical Engineering, National Dong Hwa University, Hualien, Taiwan, in 2012. He is now pursuing his PhD in the same department. His research interests include video/image compression and motion estimation.

Jian-Sheng Wu received his BS from the Department of Photonics and Communication Engineering, Asia University, Taichung, Taiwan, in 2010, and his MS from the Department of Electrical Engineering, National Dong Hwa University, Hualien, Taiwan, in 2012. His research interests include mode decision and motion estimation for Scalable Video Coding.