Efficient Mode Decision Algorithm Based on Spatial, Temporal, and Inter-layer Rate-Distortion Correlation Coefficients for Scalable Video Coding

Po-Chun Wang, Gwo-Long Li, Shu-Fen Huang, Mei-Juan Chen, and Shih-Chien Lin

The layered coding structure of scalable video coding (SVC) with adaptive inter-layer prediction causes noticeable computational complexity increments when compared to existing video coding standards. To lighten the computational complexity of SVC, we present a fast algorithm to speed up the inter-mode decision process. The proposed algorithm terminates inter-mode decision early in the enhancement layers by estimating the rate-distortion (RD) cost from the macroblocks of the base layer and the enhancement layer in temporal, spatial, and inter-layer directions. Moreover, a search range decision algorithm is also proposed in this paper to further increase the motion estimation speed by using the motion vector information from temporal, spatial, or inter-layer domains. Simulation results show that the proposed algorithm can determine the best mode and provide more efficient total coding time saving with very slight RD performance degradation for spatial and quality scalabilities.

Keywords: Scalable video coding, mode decision, speed up, motion estimation, inter-layer prediction.

Manuscript received Oct. 26, 2009; revised Feb. 11, 2010; accepted Mar. 31, 2010.

Po-Chun Wang (phone: +886 3 8634072, email: m9623035@em96.ndhu.edu.tw), Gwo-Long Li (email: m9323004@em93.ndhu.edu.tw), Shu-Fen Huang (corresponding author, email: m9823002@ems.ndhu.edu.tw), Mei-Juan Chen (email: cmj@mail.ndhu.edu.tw), and Shih-Chien Lin (email: u9523040@ems.ndhu.edu.tw) are with the Department of Electrical Engineering, National Dong-Hwa University, Hualien, Taiwan.

doi:10.4218/etrij.10.0109.0622

I. Introduction

The topic of multimedia communication has received more and more attention with the development of network technology. Recently, the advancement of software and hardware technologies has brought multimedia applications such as video telephony, digital television, video on demand (VOD), Internet protocol television (IPTV), and so on into our daily lives. The development of personal mobile communication systems has introduced the application heterogeneities in video coding. Therefore, the video coding system must encode the video sequence in different frame sizes, frame rates, and bit rates to supply such heterogeneous demands.

To meet the requirements of application heterogeneities, the newest video coding standard called scalable video coding (SVC) [1], [2] or H.264 Scalable Extension was recently standardized by the Joint Video Team of the ITU-T Video Coding Group and ISO/IEC Moving Picture Experts Group. Compared with the previous video coding standards, SVC supports three scalabilities in terms of time, space, and quality. In SVC, the video sources are encoded into one base layer and several enhancement layers. The scalable video coding structure for spatial scalability is shown in Fig. 1. In this figure, the base layer is responsible for coding the smallest size of video sequence, and it is H.264/MPEG-4 Part 10 (AVC) compatible.

To improve the coding efficiency, SVC prefers to remove the redundancies between different frame resolution layers when encoding enhancement layers. Although the coding performance

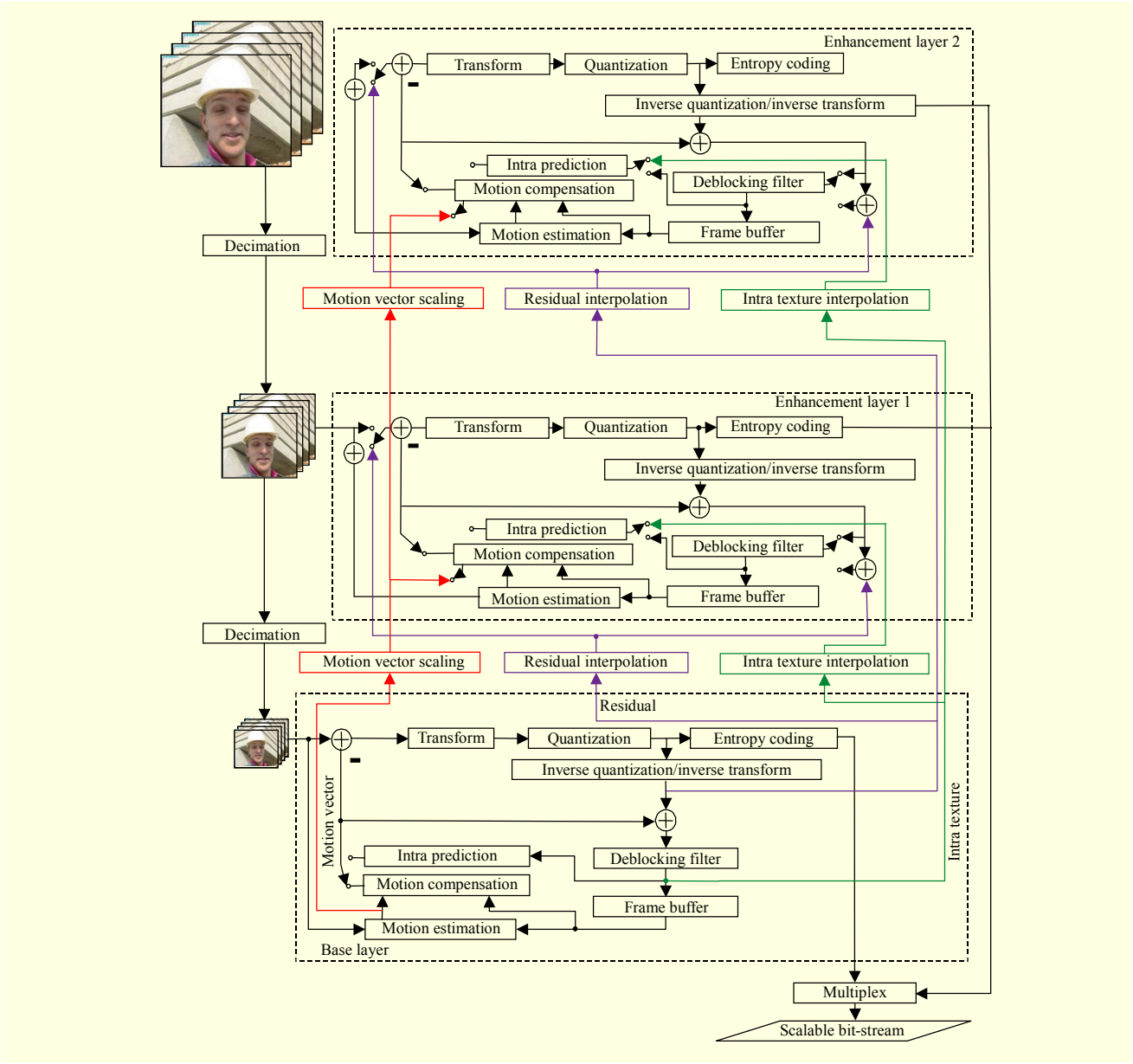

Fig. 1. Spatial scalability architecture of SVC.

can be improved by SVC structure, it needs significant computational complexity compared to the original H.264 encoder due to the inclusion of three inter-layer prediction tools. To get better coding performance in H.264/AVC, seven different block sizes and shapes are supported for the inter-mode prediction such as Mode_16×16, Mode_16×8, Mode_8×16, Mode_8×8, Mode_8×4, Mode_4×8, and Mode_4×4. Furthermore, for intra-mode prediction, there are nine prediction directions for Intra_4×4 and four prediction directions for Intra_16×16. However, in addition to the inherent prediction modes supported in H.264/AVC, three more macroblock (MB) prediction modes called inter-layer motion prediction, inter-layer residual prediction, and inter-layer intra prediction are additionally supported to encode the macroblock of enhancement layers in SVC. In these inter-layer prediction modes, the base layer information is used as a reference to further increase the coding performance.

Several fast mode decision algorithms have been proposed for speeding up the SVC encoding process [3]-[12]. In the literature, they attempted to decrease the motion estimation time in the enhancement layer for spatial scalability. In [3], the author only considered DC, horizontal, and vertical prediction modes instead of the nine different modes to decrease the intra-mode coding time in the enhancement layer. Libo [4] proposed a low-complexity intra-prediction algorithm by removing Intra_16×16 and Intra_8×8 modes in the enhancement layer.

Han [5] proposed an intra-prediction scheme for inter-layer intra prediction in the enhancement layer. Ye and others [6] proposed a residual up-sampling scheme for the inter-layer residual prediction. Lange [7], [8] proposed an adaptive motion vector selection method from the base layer for the inter-layer motion prediction. Wang and others [9] proposed the concept of different intra block sizes to improve the performance of inter-layer intra-mode prediction. Li and others [10] observed the intra-mode relationship between the base layer and enhancement layer, and the mode decision rule was introduced to reduce the computational complexity in the enhancement layer by adopting these observations. Lin and others [11] analyzed the mode relationship between the base layer and enhancement layer for different quantization parameters, and thus proposed a mode decision table to decide the best mode in the enhancement layer. In addition to predicting the best mode, the other way is to decide the initial search point and reference frame from the base layer. Lee and others [12] used the bi-predictive zero motion, uni-predictive zero motion, and zero coefficient block of the base layer to decide what prediction modes should be checked in the enhancement layer.

Kim and others [13] proposed an algorithm to accelerate inter-mode decision processes by using the temporal correlation existing between inter frames for H.264. In this method, the MB tracking method was used to get the most correlated block and the rate-distortion (RD) cost of the most correlated block was used as a threshold to speed up the mode decision procedure. Ri and others [14] used the correlation coefficient of rate distortion in spatial and temporal directions to decide the early termination threshold in the mode decision for H.264.

To reduce the computational complexity of SVC, we propose a new fast inter-mode selection scheme by considering the RD cost correlation coefficients in the base layer and the enhancement layer to decide the macroblock mode in the enhancement layer. In addition, according to the maximum correlation coefficient, an adaptive search range decision algorithm is also proposed to further increase the coding speed [15].

The rest of this paper is organized as follows. The proposed inter-mode decision scheme by using the correlation coefficient is described in section II. The experimental results and conclusions are presented in section III and section IV, respectively.

II. Proposed Fast Inter-mode Decision Scheme in SVC

In this section, a fast inter-mode decision algorithm is proposed to enhance the coding efficiency of the SVC encoder. The proposed algorithm includes fast mode decision based on the correlation coefficient of rate-distortion costs for spatial and quality scalabilities. To decide the best mode of the current macroblock in the enhancement layer, the proposed fast mode decision algorithm is described as follows.

1. Rate-Distortion Cost

In H.264 and SVC, the best mode is decided by the rate-distortion cost and it can be expressed as

$$J(s,c,Mode|QP) = SSD(s,c,Mode|QP) + \lambda_{\text{Mode}} \cdot R(s,c,Mode|QP), \quad (1)$$

where s is the original block, c refers to the reconstructed block, QP means the quantization parameter, and λ_{Mode} denotes the Lagrangian multiplier [1]. In (1), the *SSD* is the sum of the square difference that is used to measure the distortion. *Mode* is the encoding mode, and $R(s,c,Mode|QP)$ represents the number of bits associated with the *Mode* and motion vectors. In SVC, different video resolutions and qualities are supported by spatial scalability and quality scalability to satisfy the diversities of user requirements. In the spatial scalability, the video sequence in the base layer is the downscaled version of the enhancement layer. In the quality scalability, the video resolutions are equal for both the base and enhancement layers. This mechanism results in a high correlation between the base layer and the enhancement layer. Hence, we use the correlation coefficient of rate-distortion costs to predict a criterion which is used to decide the best mode of a macroblock in the enhancement layer. Before describing our proposed algorithm, the definition of the correlation coefficient is briefly described as follows. The correlation coefficient is a well-known rule in statistics and probability theory. It is usually utilized to indicate the strength and direction between two variables X and Y as in

$$\text{corr}(X,Y) = \rho_{X,Y}, \quad (2)$$

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y}, \quad (3)$$

where corr is the correlation coefficient, cov refers to the covariance, and σ is the standard deviation. The above formula defines the *population* correlation coefficient. Alternatively, the same idea can be applied to a sample rather than a population, which gives the *sample* correlation coefficient, commonly represented by r. As a result, the sample correlation coefficient can be computed as

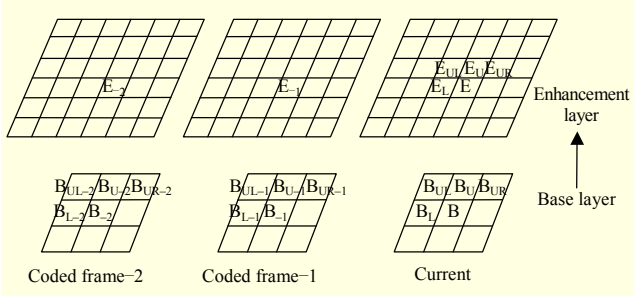

Fig. 2. Illustration of macroblocks used in our proposed scheme for spatial scalability.

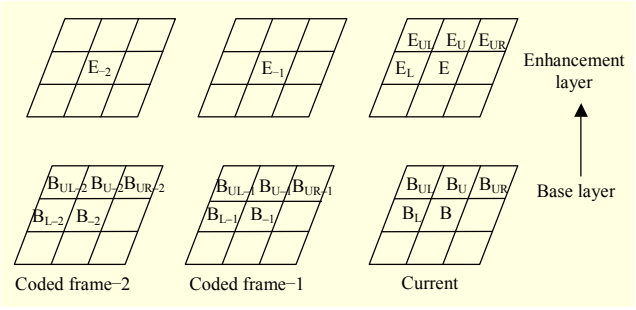

Fig. 3. Illustration of macroblocks used in our proposed scheme for quality scalability.

$$r_{XY} = \frac{\Sigma x_i y_i - \dfrac{\Sigma x_i \Sigma y_i}{n}}{\sqrt{\Sigma x_i - \dfrac{(\Sigma x_i)^2}{n}}\sqrt{\Sigma y_i^2 - \dfrac{(\Sigma y_i)^2}{n}}}, \quad (4)$$

where $x_i \in \{X\}$ and $y_i \in \{X\}$. Therefore, the relationship in spatial, temporal, and inter-layer domains is taken into account for computing the correlation coefficients, and consequently, constructing our fast mode decision algorithm.

In this paper, the rate-distortion costs are used as the samples for computing the correlation coefficients. Figures 2 and 3 show the corresponding macroblocks (samples) that have been used in our proposed algorithm for deriving the correlation coefficients in spatial and quality scalabilities, respectively. As shown in Figs. 2 and 3, the subscript index number of –1 and –2 indicate the collocated blocks in the (coded frame–1)th frame and the (coded frame–2)th frame, respectively. The definitions of various samples for spatial, temporal, and inter-layer domains are shown in (5) to (10), respectively.

Spatial samples:

$$X_{\mathrm{S}} = \{RDcost_{\mathrm{B}}, RDcost_{\mathrm{B}_{-1}}, RDcost_{\mathrm{B}_{-2}}\}, \quad (5)$$

$$Y_{\mathrm{S}\alpha} = \{RDcost_{\mathrm{B}_{\alpha}}, RDcost_{\mathrm{B}_{\alpha-1}}, RDcost_{\mathrm{B}_{\alpha-2}} \,|\, \alpha \in \{\mathrm{L,U,UL,UR}\}\}, \quad (6)$$

Temporal samples:

$$X_{\mathrm{T}} = \{RDcost_{\mathrm{B}}, RDcost_{\mathrm{B_L}}, RDcost_{\mathrm{B_{UL}}}, RDcost_{\mathrm{B_U}}\}, \quad (7)$$

$$Y_{\mathrm{T}\beta} = \left\{ \begin{matrix} RDcost_{\mathrm{B}_{-\beta}}, RDcost_{\mathrm{B_{L-\beta}}}, \\ RDcost_{\mathrm{B_{UL-\beta}}}, RDcost_{\mathrm{B_{U-\beta}}} \end{matrix} \middle| \, \beta \in \{1,2\} \right\}, \quad (8)$$

Inter-layer samples:

$$X_{\mathrm{I}} = \{RDcost_{\mathrm{B_L}}, RDcost_{\mathrm{B_{UL}}}, RDcost_{\mathrm{B_U}}\}, \quad (9)$$

$$Y_{\mathrm{I}} = \{RDcost_{\mathrm{E_L}}, RDcost_{\mathrm{E_{UL}}}, RDcost_{\mathrm{E_U}}\}, \quad (10)$$

where *RDcost* refers to the rate-distortion cost of a certain macroblock. X_{S}, X_{Y}, and X_{I} are the samples in spatial, temporal, and inter-layer domains, respectively. They will be used to substitute for the X component in (4) to calculate the correlation coefficients. Similarly, $Y_{\mathrm{S}\alpha}$, $Y_{\mathrm{T}\beta}$, and Y_{I} are the samples which will be used to substitute for the Y component in (4) to obtain the correlation coefficients. Once the samples of spatial, temporal, and inter-layer domains are completely defined, the correlation coefficients of these domains are calculated by (4) separately. Consequently, the maximum correlation coefficient is selected by (11) to (14) since the most related macroblock should be extracted to be further used as a prediction reference.

$$C_{\mathrm{S}} = \max\{\gamma_{X_{\mathrm{S}}Y_{\mathrm{S}\alpha}}\}, \quad (11)$$

$$C_{\mathrm{T}} = \max\{\gamma_{X_{\mathrm{T}}Y_{\mathrm{T}\beta}}\}, \quad (12)$$

$$C_{\mathrm{L}} = \gamma_{X_{\mathrm{L}}X_{\mathrm{L}}}. \quad (13)$$

$$C_{\mathrm{Pred}} = \max\{C_{\mathrm{S}}, C_{\mathrm{T}}, C_{\mathrm{L}}\}. \quad (14)$$

With the assistance of (11) to (14), the direction which has the maximum correlation coefficient can be obtained, and consequently, used as the prediction reference for the current encoding macroblock in the enhancement layer. Hence, the prediction reference derived from the previous step is further adopted to decide an early termination threshold (TH) to avoid exhaustive mode testing. The *TH* is based on the maximum correlation coefficient, C_{Pred}, and it can be adaptively adjusted during the encoding process of the enhancement layer because of the high relationship between the base and the enhancement layers. As a result, the *TH* can be derived by (15). After obtaining the *TH*, it will be used during the mode decision making process to check whether the modes waiting for examination should be further tested or not. If the rate distortion cost of the current testing mode is smaller than *TH*, the mode decision making process will be terminated immediately.

$$TH = \begin{cases} RDcost_{E_{-1}} : C_{pred} = \gamma_{X_T Y_{T1}}, \\ RDcost_{E_{-2}} : C_{pred} = \gamma_{X_T Y_{T2}}, \\ RDcost_{E_L} : C_{pred} = \gamma_{X_S Y_{SL}}, \\ RDcost_{E_U} : C_{pred} = \gamma_{X_S Y_{SU}}, \\ RDcost_{E_{UL}} : C_{pred} = \gamma_{X_S Y_{SUL}}, \\ RDcost_{E_{UR}} : C_{pred} = \gamma_{X_S Y_{SUR}}, \\ RDcost_B : C_{pred} = \gamma_{X_L Y_L}. \end{cases} \quad (15)$$

2. Adaptive Search Range

As mentioned, the motion estimation of SVC consumes many computational complexities. To further decrease the coding time, the exhaustive search point checking is expected to be avoided during the motion search in the enhancement layer. To achieve this goal, the relationship between the motion vectors of the current encoding macroblock and the macroblock with the maximum correlation coefficient are analyzed. In our analysis, seven sequences including Akiyo, Foreman, Garden, News, Soccer, Stefan, and Table Tennis (Table for short) with the frame resolution of the base layer in QCIF (176×144) and the enhancement layer in CIF (352×288) are tested. Table 1 shows the probability that the motion vector magnitude of the current encoding macroblock is smaller than or equal to the motion vector magnitude of the macroblock with the maximum correlation coefficient. From Table 1, it can be observed that the motion vector magnitude between the current encoding macroblock and the macroblock with maximum correlation coefficient is very similar (at least 82.50% for high motion sequence Stefan).

This situation shows us that the best motion vector of the current encoding macroblock in the enhancement layer can be obtained within a restricted search area which can be decided by the motion vectors of the macroblock with the maximum correlation coefficient. As a result, the search range can be adaptively adjusted with the change of the macroblock which has the maximum correlation coefficient. The search range for the current encoding macroblock can be calculated as follows.

$$SR = \begin{cases} \left|MV_{E_{-1}}\right| : C_{pred} = \gamma_{X_T Y_{T1}}, \\ \left|MV_{E_{-2}}\right| : C_{pred} = \gamma_{X_T Y_{T2}}, \\ \left|MV_{E_L}\right| : C_{pred} = \gamma_{X_S Y_{SL}}, \\ \left|MV_{E_U}\right| : C_{pred} = \gamma_{X_S Y_{SU}}, \\ \left|MV_{E_{UL}}\right| : C_{pred} = \gamma_{X_S Y_{SUL}}, \\ \left|MV_{E_{UR}}\right| : C_{pred} = \gamma_{X_S Y_{SUR}}, \\ \left|MV_B\right| \times m : C_{pred} = \gamma_{X_L Y_L}, \end{cases} \quad (16)$$

where $|MV|$ indicates the motion vector magnitude of the corresponding macroblock. The m refers to the ratio of the frame resolution between layers. If the maximum correlation coefficient belongs to inter-layer prediction, the search range should be multiplied by a factor of m to adjust for search range size since the motion vectors of the base layer are up-sampled for the prediction of the enhancement layer in SVC.

Table 1. Probability that motion vector magnitude of current macroblock is smaller than or equal to motion vector magnitude of macroblock with maximum correlation coefficient.

Sequence	Probability (%)
Akiyo	99.20
Foreman	87.16
Garden	83.14
News	98.21
Soccer	83.88
Stefan	82.50
Table	89.63

3. Proposed Fast Inter-mode Decision Algorithm

The flowchart of the overall proposed algorithm is shown in Fig. 4 and the coding procedure is described as follows. First, the base layer is encoded by the H.264 compatible encoder. Afterward, for encoding the enhancement layers, we calculate seven correlation coefficients from spatial, temporal and inter-layer domains by (4) to (10), and the maximum correlation coefficient is obtained by (11) to (14). After obtaining the maximum correlation coefficient, the threshold, *TH*, based on the C_{Pred}, is derived by (15) to define an early termination criterion for increasing the mode decision speed. Meanwhile, the search range of the current encoding macroblock is dynamically decided by C_{Pred} in (16). The first test mode includes the C_{Pred}'s mode, SKIP, DIRECT, and Mode_16×16. For an upcoming test mode, the position with the minimum rate-distortion cost is found out within the decided search range. If the minimum rate distortion cost of the current testing mode is smaller than *TH*, the test for the unchecked modes is terminated immediately and the best mode is selected from the previous checked modes. Otherwise, the other modes are checked in turns.

III. Experimental Results

The proposed fast mode decision algorithm is implemented

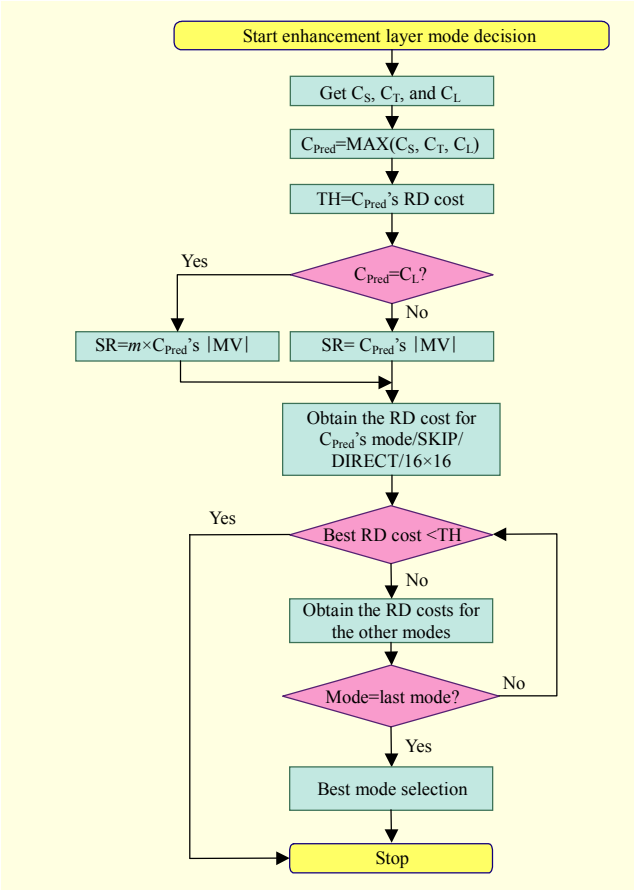

Fig. 4. Flowchart of proposed scheme.

Table 2. Simulation environment.

Parameter	Value
GOP size	8, 16, 32
Search range	±16
Number of reference frame	3
Intra period	–1
Motion vector accuracy	1/4
Frame rate (frames per second)	30

in the JSVM 9.10 encoder. The computer used for the simulation has 3.0 GHz Intel Pentium 4 CPU, 512 MB RAM with Windows XP professional operating system.

Our simulations are performed with two test combinations. The first combination is to test thirteen sequences including Akiyo, City, Coastguard, Crew, Garden, Foreman, Harbour, Ice, Mobile, News, Stefan, Table Tennis (Table for short), and Soccer with the frame resolution of the base layer in QCIF (176 × 144) and the enhancement layer in CIF (352 × 288) so that the m is 2. Another test combination is to test four sequences including the City, Crew, Harbour, and Ice test

Table 3. PSNR and ΔPSNR (dB) comparisons for spatial scalability (GOP8, BL QP=38, 2-layer, reference frame number = 3).

Sequence	JSVM		Lin's [11]		Lee's [12]		Proposed	
EL QP	18	32	18	32	18	32	18	32
Akiyo	47.05	39.90	–0.03	–0.04	–0.13	–0.10	–0.02	–0.02
City	43.35	34.95	–0.04	–0.06	–0.12	–0.13	–0.06	–0.07
Coastguard	42.76	32.91	–0.04	–0.05	–0.15	–0.11	–0.08	–0.05
Crew	44.03	36.01	–0.03	–0.10	–0.22	–0.26	–0.09	–0.13
Garden	43.85	33.15	–0.05	–0.06	–0.16	–0.24	–0.04	–0.10
Foreman	43.81	36.16	–0.06	–0.11	–0.22	–0.28	–0.06	–0.10
Harbour	42.59	32.69	–0.02	–0.03	–0.10	–0.11	–0.02	–0.04
Ice	46.87	38.84	–0.11	–0.27	–0.46	–0.52	–0.10	–0.19
Mobile	42.73	32.54	–0.06	–0.05	–0.16	–0.17	–0.03	–0.06
News	46.14	38.01	–0.03	–0.05	–0.25	–0.13	–0.03	–0.04
Stefan	43.42	35.01	–0.05	–0.07	–0.17	–0.10	–0.06	–0.06
Table	43.36	34.50	–0.07	–0.19	–0.23	–0.23	–0.08	–0.15
Soccer	43.71	34.65	–0.04	–0.13	–0.14	–0.14	–0.10	–0.11
Average	44.13	35.33	–0.05	–0.09	–0.19	–0.19	–0.06	–0.09

sequences with the frame resolution of base layer in QCIF (176 × 144), the first enhancement layer in CIF (352 × 288), and the second enhancement layer in 4CIF (704 × 576). The group of picture (GOP) is set to 8, 16, and 32 for temporal scalability. The simulation settings are listed in Table 2. The maximum search range is set to ±16 pixels, and the number of reference frames is set to 3 for motion estimation. The quantization parameter is set to 38 in the base layer and 18 and 32 for the first enhancement layer for 2-layer scalability. Each test sequence contains a total of 100 frames, with one I frame followed by 99 P and B frames. For the performance comparison, the three methods of Lin [11], Lee [12], and the original JSVM 9.10 reference software [16] are compared with our proposed algorithm. We implemented Lin's algorithm without MODE_SR in JSVM 9.10 since the MODE_SR has been removed in reference software JSVM 9.10. We compare the proposed method with original JSVM 9.10, Lin's and Lee's methods in terms of the PSNR, and mean structural similarity (MSSIM) [17], bitrate, and time saving (see (17)) to measure the performance.

$$\text{Time saving} = \frac{T_{\text{JSVM}} - T_{\text{Reference}}}{T_{\text{JSVM}}} \times 100\%. \qquad (17)$$

Tables 3 through 6 show PSNR and MSSIM comparisons with different GOPs of 8 and 16 in different quantization parameters (QPs) for 2-layer spatial scalability. In these tables,

Table 4. MSSIM comparison for spatial scalability (GOP8, BL QP=38, 2-layer, reference frame number = 3).

Sequence	JSVM		Lin's [11]		Lee's [12]		Proposed	
EL QP	18	32	18	32	18	32	18	32
Akiyo	0.989	0.967	0.989	0.967	0.989	0.966	0.989	0.967
City	0.989	0.941	0.989	0.940	0.989	0.939	0.989	0.940
Coastguard	0.988	0.908	0.988	0.907	0.988	0.906	0.988	0.907
Crew	0.982	0.923	0.982	0.921	0.981	0.919	0.982	0.921
Garden	0.993	0.971	0.993	0.970	0.993	0.969	0.993	0.970
Foreman	0.984	0.934	0.983	0.933	0.983	0.932	0.983	0.933
Harbour	0.994	0.955	0.994	0.954	0.994	0.953	0.994	0.954
Ice	0.988	0.968	0.988	0.967	0.987	0.966	0.988	0.967
Mobile	0.993	0.966	0.993	0.965	0.993	0.964	0.993	0.965
News	0.989	0.962	0.989	0.962	0.988	0.962	0.989	0.962
Stefan	0.991	0.971	0.991	0.971	0.990	0.970	0.991	0.971
Table	0.981	0.900	0.981	0.896	0.980	0.896	0.981	0.896
Soccer	0.985	0.906	0.985	0.902	0.985	0.905	0.985	0.904
Average	0.988	0.944	0.988	0.943	0.988	0.942	0.988	0.943

Table 5. PSNR and ΔPSNR (dB) comparisons for spatial scalability (GOP16, BL QP=38, 2-layer, reference frame number = 3).

Sequence	JSVM		Lin's [11]		Lee's [12]		Proposed	
EL QP	18	32	18	32	18	32	18	32
Akiyo	47.60	40.48	–0.04	–0.06	–0.14	–0.12	–0.01	–0.03
City	43.41	35.54	–0.04	–0.09	–0.12	–0.24	–0.06	–0.11
Coastguard	42.90	33.03	–0.07	–0.14	–0.19	–0.14	–0.16	–0.09
Crew	44.10	36.02	–0.02	–0.09	–0.21	–0.25	–0.13	–0.15
Garden	43.94	33.14	–0.06	–007	–0.16	–0.25	–0.06	–0.13
Foreman	43.91	36.51	–0.06	–0.14	–0.20	–0.32	–0.09	–0.14
Harbour	42.67	32.70	–0.02	–0.03	–0.10	–0.12	–0.03	–0.05
Ice	47.03	38.98	–0.11	–0.23	–0.42	–0.47	–0.12	–0.22
Mobile	42.82	32.61	–0.05	–0.06	–0.16	–0.20	–0.03	–0.08
News	46.57	38.70	–0.05	–0.09	–0.28	–0.20	–0.04	–0.08
Stefan	43.47	35.02	–0.05	–0.37	–0.15	–0.09	–0.07	–0.07
Table	43.41	34.68	–0.07	–0.21	–0.22	–0.30	–0.10	–0.23
Soccer	43.78	35.19	–0.04	–0.13	–0.15	–0.20	–0.10	–0.16
Average	44.28	35.58	–0.05	–0.66	–0.19	–0.22	–0.08	–0.12

Table 6. MSSIM comparison for spatial scalability (GOP16, BL QP=38, 2-layer, reference frame number = 3).

Sequence	JSVM		Lin's [11]		Lee's [12]		Proposed	
EL QP	18	32	18	32	18	32	18	32
Akiyo	0.991	0.971	0.991	0.970	0.991	0.970	0.991	0.970
City	0.990	0.949	0.989	0.948	0.989	0.947	0.989	0.948
Coastguard	0.988	0.911	0.988	0.910	0.988	0.908	0.988	0.909
Crew	0.982	0.923	0.982	0.922	0.981	0.919	0.982	0.921
Garden	0.993	0.972	0.993	0.971	0.993	0.970	0.993	0.971
Foreman	0.984	0.940	0.984	0.938	0.983	0.937	0.983	0.938
Harbour	0.994	0.954	0.994	0.954	0.994	0.953	0.994	0.954
Ice	0.989	0.969	0.988	0.967	0.987	0.966	0.988	0.968
Mobile	0.994	0.968	0.993	0.967	0.993	0.966	0.993	0.967
News	0.990	0.967	0.990	0.966	0.989	0.966	0.990	0.966
Stefan	0.991	0.971	0.991	0.972	0.990	0.970	0.991	0.971
Table	0.981	0.900	0.981	0.900	0.980	0.896	0.981	0.898
Soccer	0.985	0.921	0.985	0.918	0.985	0.919	0.985	0.918
Average	0.989	0.947	0.988	0.946	0.988	0.945	0.988	0.946

Table 7. Bitrate (bits per second) and bitrate increase ratio (%) comparisons for spatial scalability (GOP8, BL QP=38, 2-layer, reference frame number = 3).

Sequence	JSVM		Lin's [11]		Lee's [12]		Proposed	
EL QP	18	32	18	32	18	32	18	32
Akiyo	459.07	78.40	1.18	1.88	5.20	3.60	0.51	0.00
City	1892.16	318.12	2.10	2.87	7.31	6.30	1.84	0.80
Coastguard	4034.64	689.67	0.42	1.56	2.60	3.12	0.46	0.74
Crew	2629.57	462.14	2.22	4.80	6.44	2.58	1.70	0.84
Garden	4091.46	870.47	0.98	1.71	4.32	3.96	1.37	1.76
Foreman	1888.31	317.56	3.50	4.89	9.81	4.52	2.16	0.78
Harbour	4664.94	819.36	0.00	1.70	1.57	2.50	0.46	1.05
Ice	1121.24	312.86	5.17	5.07	11.3	4.32	4.16	2.73
Mobile	5029.86	834.77	0.67	1.71	3.30	3.71	0.56	0.29
News	857.06	181.91	2.00	3.03	7.40	3.91	1.13	0.84
Stefan	4014.29	872.65	1.29	3.88	3.12	4.32	1.75	2.42
Table	2988.52	538.17	4.36	6.44	8.22	6.63	3.07	1.54
Soccer	2252.73	472.64	3.39	5.48	6.09	4.15	3.14	2.42
Average	2763.37	520.67	2.10	3.46	5.90	4.12	1.72	1.25

it can be found that the PSNR performance of our proposed method outperforms to Lee's method and is very close to Lin's algorithm and JSVM for different QPs.

The reason why our proposed method can achieve better performance than other methods is that Lin's and Lee's methods used the statistical approach to predict the mode in the enhancement layer. However, the proposed method predicts the best mode for the enhancement layer by the real RD cost. The performance of the proposed approach can be higher than

Table 8. Bitrate (bits per second) and bitrate increase ratio (%) comparisons for spatial scalability (GOP16, BL QP=38, 2-layer, reference frame number = 3).

Sequence	JSVM		Lin's [11]		Lee's [12]		Proposed	
EL QP	18	32	18	32	18	32	18	32
Akiyo	468.46	80.75	1.65	2.14	5.02	2.96	0.75	0.30
City	1929.98	333.53	2.53	2.78	7.57	4.89	2.79	1.42
Coastguard	4110.26	704.51	0.17	0.98	2.26	1.65	0.84	0.75
Crew	2687.32	466.39	2.28	4.93	5.78	2.26	2.22	1.33
Garden	4135.33	873.02	1.41	2.79	5.02	5.83	2.22	2.95
Foreman	1865.79	316.23	4.02	4.96	10.0	3.70	3.14	1.19
Harbour	4723.52	814.52	0.06	1.74	1.59	1.98	0.57	1.13
Ice	1155.26	317.09	5.40	5.76	10.7	4.48	4.71	3.08
Mobile	5034.89	817.80	1.13	2.70	3.62	3.66	1.08	1.06
News	878.57	190.54	2.29	3.20	6.81	2.91	1.71	1.41
Stefan	4091.81	908.31	1.44	4.23	3.03	4.43	2.10	3.00
Table	3053.36	561.23	4.47	5.53	8.18	5.18	3.97	2.43
Soccer	2297.34	501.45	3.52	5.27	6.27	3.71	3.74	2.77
Average	2802.45	529.64	2.34	3.62	5.84	3.66	2.30	1.76

Table 9. Time saving (%) comparison for spatial scalability (GOP8, BL QP=38, 2-layer, reference frame number = 3).

Sequence	Lin's [11]		Lee's [12]		Proposed	
EL QP	18	32	18	32	18	32
Akiyo	64	65	75	76	76	77
City	42	42	71	72	72	75
Coastguard	44	45	70	71	72	74
Crew	54	54	66	67	72	74
Garden	49	50	69	69	73	73
Foreman	48	48	69	70	67	71
Harbour	45	45	65	65	71	74
Ice	56	57	68	69	68	72
Mobile	40	41	69	70	70	73
News	63	64	73	74	73	74
Stefan	37	37	57	57	53	52
Table	43	43	63	62	56	57
Soccer	47	47	66	67	68	69
Average	49	49	68	68	69	70

the other methods. The PSNR decrease of the proposed method is below 0.12 dB on average compared to JSVM 9.10. From Tables 4 and 6, we can observe that all test algorithms have similar MSSIM results. Tables 7 and 8 show the bitrate comparison for each algorithm with GOP8 and GOP16 in BL (QP=38) and EL (QP=18, 32) in the enhancement layer for 2-layer spatial scalability.

Table 10. Time saving (%) comparison for spatial scalability (GOP16, BL QP=38, 2-layer, reference frame number = 3).

Sequence	Lin's [11]		Lee's [12]		Proposed	
EL QP	18	32	18	32	18	32
Akiyo	63	63	74	75	71	71
City	37	38	71	71	70	70
Coastguard	42	42	69	70	71	71
Crew	51	52	65	66	72	72
Garden	47	47	68	68	71	71
Foreman	45	46	68	68	70	71
Harbour	41	41	64	64	71	71
Ice	54	54	68	68	71	71
Mobile	37	37	69	69	71	71
News	61	61	72	72	71	71
Stefan	38	38	57	57	50	51
Table	44	44	63	62	55	58
Soccer	43	43	65	65	72	72
Average	46	47	67	67	68	69

Table 11. Enhancement layer time saving (%) comparison for spatial scalability (GOP16, BL QP=38, 2-layer, reference frame number = 3).

Sequence	Lin's [11]		Lee's [12]		Proposed	
EL QP	18	32	18	32	18	32
Akiyo	81	79	90	90	89	88
City	76	76	88	88	88	87
Coastguard	72	70	85	85	89	90
Crew	60	60	80	80	89	91
Garden	68	68	84	84	90	90
Foreman	68	67	84	84	90	90
Harbour	57	57	79	79	88	90
Ice	61	64	82	82	90	90
Mobile	72	70	85	85	90	90
News	74	75	87	87	87	87
Average	69	69	84	84	89	89

From these results, we can observe that the bitrate of the proposed method is lower than that of Lin's and Lee's methods but slightly higher than JSVM. Tables 9 through 11 demonstrate the coding time saving compared to JSVM 9.10 for the proposed method and other algorithms for 2-layer spatial scalability. From these tables, we can observe that our proposed method can save a large amount of encoding time

Table 12. Mode hit rate for spatial scalability (GOP16, BL QP=38, EL QP=32, reference frame number = 3).

Sequence	Mode hit rate (%)
Akiyo	97.02
City	92.86
Coastguard	88.36
Crew	86.52
Garden	88.39
Foreman	90.93
Harbour	84.97
Ice	92.78
Mobile	85.44
News	95.49
Average	90.28

Table 13. PSNR and ΔPSNR (dB) comparisons for spatial scalability (GOP16, BL QP=38, 3-layer, reference frame number=3).

Sequence	JSVM		Lin's [11]		Lee's [12]		Proposed	
EL1 QP	18	32	18	32	18	32	18	32
EL2 QP	18	26	18	26	18	26	18	26
City	37.55	37.47	–0.06	–0.13	–0.06	–0.07	–0.05	–0.06
Crew	43.72	39.18	–0.09	–0.26	–0.17	–0.17	–0.10	–0.07
Harbour	42.99	37.46	–0.04	–0.07	–0.10	–0.10	–0.02	–0.06
Ice	45.53	42.07	–0.16	–0.31	–0.27	–0.31	–0.11	–0.24
Average	42.45	39.05	–0.09	–0.19	–0.15	–0.16	–0.07	–0.11

Table 14. Bitrate (bits per second) and bitrate increase ratio (%) comparisons for spatial scalability (GOP16, BL QP=38, 3-layer, reference frame number = 3).

Sequence	JSVM		Lin's [11]		Lee's [12]		Proposed	
EL1 QP	18	32	18	32	18	32	18	32
EL2 QP	18	26	18	26	18	26	18	26
City	4221.4	2951.3	2.12	9.61	4.73	4.41	1.92	3.84
Crew	11933.8	2513.1	0.93	8.70	1.24	5.64	0.80	6.35
Harbour	17869.1	5210.0	–0.54	3.01	0.66	2.51	–0.45	2.81
Ice	4788.7	1359.1	4.18	11.4	8.61	11.3	2.71	9.50
Average	9703.3	3008.4	1.67	8.18	3.81	5.97	1.25	5.63

than other algorithms, especially for JSVM. We further analyze the mode hit rate. In a mode hit rate calculation, if the mode of the current macroblock predicted by our proposal equals to the final decided mode of JSVM, we call this situation a hit. Table 12 shows the comparison of hit rate of our proposed algorithm for various sequences. The proposed algorithm can achieve around a 90% correct rate. This situation implies that our proposed algorithm can accurately predict the coding mode and consequently reduce the computational complexity. Tables 13 through 15 show the performances of PSNR, bitrate, and time saving in different QPs for 3-layer spatial scalability.

Table 15. Time saving (%) comparison for spatial scalability (GOP16, BL QP=38, 3-layer, reference frame number = 3).

Sequence	Lin's [11]		Lee's [12]		Proposed	
EL1 QP	18	32	18	32	18	32
EL2 QP	18	26	18	26	18	26
City	74	75	86	85	87	86
Crew	80	83	86	85	86	85
Harbour	74	74	86	84	87	87
Ice	80	80	85	85	85	85
Average	77	78	86	85	86	86

Table 16. PSNR and ΔPSNR (dB) comparisons for quality scalability (GOP32, BL QP=38, 2-layer, reference frame number = 3).

Sequence	JSVM		Lin's [11]		Proposed	
EL QP	18	32	18	32	18	32
Akiyo	48.29	39.94	–0.08	–0.12	–0.01	–0.01
City	43.76	35.62	–0.08	–0.35	–0.03	–0.02
Coastguard	42.99	33.60	–0.05	–0.13	–0.04	–0.03
Crew	44.25	35.18	0.01	–0.17	–0.01	–0.02
Garden	44.10	33.98	–0.03	–0.12	–0.02	–0.01
Foreman	44.70	36.05	–0.08	–0.20	–0.02	–0.00
Harbour	42.47	32.51	–0.00	–0.11	–0.00	–0.00
Ice	46.55	37.13	–0.04	–0.02	–0.04	–0.02
Mobile	42.58	32.86	–0.00	–0.15	–0.00	–0.00
News	47.09	37.80	–0.07	–0.11	–0.01	–0.01
Soccer	44.31	34.93	–0.09	–0.09	–0.06	–0.03
Average	44.64	35.42	–0.05	–0.14	–0.02	–0.01

For the time saving comparison, our proposed method can individually save about 69% and 85% coding time for 2-layer and 3-layer spatial scalability with very slight rate-distortion performance degradation when compared to JSVM.

For quality scalability, the enhancement layer resolution is equal to the base layer for QCIF. For the performance comparisons, two methods of Lin's and the original JSVM are compared with our proposed method. Tables 16 through 19 show the performance comparisons, including PSNR, MSSIM, bitrate, and time saving with GOP32. The proposed method can save about 60% total coding time. The PSNR and bitrate are very close to JSVM and outperform Lin's.

Table 17. MSSIM comparison for quality scalability (GOP32, BL QP=38, 2-layer, reference frame number = 3).

Sequence	JSVM		Lin's [11]		Proposed	
EL QP	18	32	18	32	18	32
Akiyo	0.995	0.978	0.995	0.977	0.995	0.978
City	0.992	0.956	0.992	0.952	0.992	0.955
Coastguard	0.988	0.913	0.988	0.911	0.988	0.913
Crew	0.989	0.930	0.989	0.927	0.989	0.929
Garden	0.995	0.974	0.994	0.972	0.995	0.973
Foreman	0.993	0.962	0.993	0.961	0.993	0.962
Harbour	0.995	0.959	0.995	0.958	0.995	0.959
Ice	0.992	0.974	0.992	0.974	0.992	0.974
Mobile	0.996	0.974	0.996	0.973	0.996	0.974
News	0.995	0.973	0.995	0.973	0.995	0.973
Soccer	0.989	0.932	0.989	0.929	0.989	0.931
Average	0.993	0.957	0.993	0.955	0.993	0.956

Table 18. Bitrate (bits per second) and bitrate ratio (%) comparisons for quality scalability (GOP32, BL QP=38, 2-layer, reference frame number = 3).

Sequence	Lin's [11]		Lee's [12]		Proposed	
EL QP	18	32	18	32	18	32
Akiyo	153.28	36.87	1.28	1.30	–0.14	–0.00
City	455.08	95.34	4.68	–3.15	0.50	0.83
Coastguard	817.37	137.38	1.61	–0.71	0.37	–0.49
Crew	773.32	149.63	2.01	0.44	0.21	–0.27
Garden	833.57	173.62	1.49	–0.20	0.17	–0.15
Foreman	498.41	113.54	3.12	0.26	0.26	0.06
Harbour	1101.30	187.15	0.88	0.06	0.14	–0.01
Ice	435.40	125.90	1.13	1.93	0.37	–0.12
Mobile	1165.94	189.64	1.26	–0.69	0.30	–0.04
News	287.42	75.61	1.33	1.12	–0.02	–0.08
Soccer	622.44	155.89	1.41	1.15	0.89	0.06
Average	649.41	130.96	1.84	0.14	0.28	–0.02

IV. Conclusion

In this paper, an efficient inter-mode decision algorithm is proposed to speed up the encoding process of SVC. The proposed fast inter-mode decision is based on the rate-distortion cost correlation coefficients of base layer and enhancement layer to determine the mode of macroblocks in the enhancement layer. In addition, according to the maximum correlation coefficient direction, the search range for the current encoding macroblock can be dynamically decided. Experimental results demonstrate that the proposed algorithm can provide much time saving with slight PSNR degradation and bitrate increase when compared to JSVM 9.10.

Table 19. Time saving comparison (%) for quality scalability (GOP32, BL QP=38, 2-layer, reference frame number = 3).

Sequence	Lin's [11]		Proposed	
EL QP	18	32	18	32
Akiyo	51	51	60	60
City	49	49	59	59
Coastguard	50	50	59	59
Crew	49	49	59	59
Garden	48	48	58	58
Foreman	48	48	59	60
Harbour	49	49	59	59
Ice	49	49	60	60
Mobile	49	48	60	60
News	51	51	60	60
Soccer	44	44	61	61
Average	49	49	59	60

References

[1] T. Wiegand et al., "Joint Draft ITU-T Rec. H.264 | ISO/IEC 14496-10 / Amd.3 Scalable Video Coding," Joint Video Team (JVT) JVT-X201, 2007.

[2] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard," *IEEE Trans.Circuits Syst. Video Technol.*, vol. 17, no. 9, Sept. 2007, pp. 1103-1120.

[3] L. Xiong, "Reducing Enhancement Layer Directional Intra-prediction Modes," Joint Video Team (JVT) JVT-P041, 2005.

[4] Y. Libo, "Low Complexity Intra Prediction for Enhance-ment Layer," Joint Video Team (JVT) JVT-Q084, 2005.

[5] W.J. Han, "Modified IntraBL Design Using Smoothed Reference," Joint Video Team (JVT) JVT-R091, 2006.

[6] Y. Ye and Y. Bao, "CE2: Improved Residual Upsampling for ESS," Joint Video Team (JVT) JVT-W117, 2007.

[7] R. Lange, "Extended Inter-layer Motion Vectors Prediction in Scalable Video Coding–Case Study and Improvement Proposal," Joint Video Team (JVT) JVT-R094, 2006.

[8] R. Lange, "Extended Inter-layer Motion Vectors Prediction in JSVM–Case Study and Experimental Results," Joint Video Team (JVT) JVT-S063, 2006.

[9] T. S. Wang et al., "Improved Inter-layer Intra Prediction for

Scalable Video Coding," *Proc. IEEE TENCON*, 2007, pp. 1-4.
[10] H. Li et al., "Fast Mode Diction for Spatial Scalable Video Coding," *Proc. IEEE Int. Symp. Circuits Syst.*, May 2006, pp. 3305-3308.
[11] H.C. Lin et al., "Layer-adaptive Mode Decision and Motion Search for Scalable Video Coding with Combined Coarse Granular Scalability (CGS) and Temporal Scalability," *Proc. IEEE Int. Conf. Image Processing,* vol. 2, Sept. 2007, pp. 289-292.
[12] 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. Broadcasting,* Mar. 2008, pp. 1-5.
[13] 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.
[14] S.H. Ri, Y. Vatis, and J. Ostermann, "Fast Inter-mode Decision in an H.264/AVC Encoder Using Mode and Lagrangian Cost Correlation," *IEEE Trans. Circuits Syst. Video Technol.,* vol. 19, no. 2, Feb. 2009, pp. 302-306.
[15] P.C. Wang et al., "An Efficient Mode Decision Scheme by using RD Cost Correlation Coefficients in Scalable Video Coding," *Proc. APSIPA Annual Summit Conference*, Sapporo, Japan, Oct. 2009, pp. 57-63.
[16] J. Reichel, H. Schwarz, 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 Q6), JVT-V203, 2007.
[17] 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.

Po-Chun Wang received his BS from the Department of Electronic Engineering, Lunghwa University of Science and Technology, Taoyuan, Taiwan, in 2007, and MS from the Department of Electrical Engineering, National Dong-Hwa University, Hualien, Taiwan, in 2009. His research interests are the scalable video coding and mode decision.

Gwo-Long Li received his BS from the Department of Computer Science and Information Engineering, Shu-Te University, Kaohsiung, Taiwan, in 2004, and MS from the Department of Electrical Engineering, National Dong-Hwa University, Hualien, Taiwan, in 2006. He is currently working toward his PhD in the Department of Electronics Engineering, National Chiao-Tung University, Hsinchu, Taiwan. In 2006, he received the Excellent Master Thesis Award from Institute of Information and Computer Machinery. He is a student member of IEEE, and his research interests are the video signal processing and its VLSI architecture design.

Shu-Fen Huang received her BS from the Department of Electrical Engineering, National Dong-Hwa University, Hualien, Taiwan, in 2009. She is currently working toward the MS in the Department of Electrical Engineering, National Dong-Hwa University, Hualien, Taiwan. Her research interests include video coding and transcoding.

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. She also served as the chair of her department from 2005 to 2006. Her research topics include video compression, motion estimation, error concealment, and video transcoding. She received the Dragon Paper Awards in 1993 and 1997 and the Xeror Paper Award in 1993. She was also the recipient of the 2005 K.T. Li Young Researcher Award from ACM Taipei/Taiwan Chapter for her contribution on video signal codec technique. In 2006, she received the Distinguished Young Engineer Award from The Chinese Institute of Electrical Engineering, Taiwan. In 2006, she also received the Excellent Master Thesis Supervision Award from Institute of Information and Computer Machinery. Since August 2008, she has served as an associate editor for *EURASIP Journal on Advances in Signal Processing*.

Shih-Chien Lin received her BS from the Department of Electrical Engineering, National Dong-Hwa University, Hualien, Taiwan, in 2010. Currently, she is working toward the MS in the Department of Communication Engineering, National Chiao-Tung University, Hsinchu, Taiwan. Her research interests mainly lie in scalable video coding.