



# Efficient rate control scheme using modified inter-layer dependency for spatial scalability

PARUL JADHAV<sup>1,\*</sup> and SHIRISH KSHIRSAGAR<sup>2</sup>

<sup>1</sup>Government College of Engineering, Pune, India

<sup>2</sup>Anchortek Techno Consultancy Pvt. Ltd., Pune, India

e-mail: parul.jadhav@mitpune.edu.in; shirish.kshirsagar@gmail.com

MS received 7 August 2015; revised 10 June 2016; accepted 15 June 2016

**Abstract.** Scalable video coding extension has been added to H.264AVC to support compression and encoding of multiple resolution video sequences, having different frame rates and fidelities in a single bit stream. The motion vectors and the residual data of the enhancement layers are derived from up-sampling the co-located macroblock (MB) of the base layer. The peak signal to noise ratio (PSNR) across the enhancement layers is degraded as up-sampling introduces distortion of high-frequency components. In this paper, a spatial-resolution-ratio-based MB mode decision scheme is proposed for spatially enhanced layers. The scheme uses the motion estimated at the base layer, to encode the respective MBs in the enhancement layers. The spatial-temporal search schemes at the enhancement layers are used to derive motion vectors and residues that are encoded using a quantization parameter obtained using independent rate control (IRC) scheme. The IRC from the prior art is modified to achieve better rate control per layer by recursive updates for mean absolute difference values of each basic unit. Proposed modified inter-layer dependency shows improvement in the PSNR for enhancement layers while the updated IRC enforces better IRC for all the layers.

**Keywords.** Scalable video coding; macroblock modes; inter-layer prediction.

## 1. Introduction

Advancements in the communication technology have enabled consumption of video information of different formats on multiple decoding devices having different spatial resolutions ranging from large plasma televisions to small mobile screens [1]. The current video codecs use single-layer coding or simulcast technique for compression and transmission of video source having same video contents but different resolutions. The basic motivation of scalable video coding (SVC) is to support decoding devices having diversified display resolution, frames rates and quality requirements [2] with a single bit stream. The scalable extension was first proposed by ISO/IEC Moving Picture Experts Group-2 (MPEG-2) compression standard in 2004. In 2005, MPEG and ITU-T's Video Coding Experts Group (VCEG) jointly finalized the SVC as an extension of H.264AVC [3].

H.264/AVC is a hybrid codec using residual coding of the macroblocks (MBs) by motion estimation and compensation using inter-prediction schemes for predictive (P) frames or intra-prediction schemes for intra (I) frames [4, 5] for a single layer. However, the scalable extension of H.264/AVC uses inter-layer-dependency-based encoding.

Each layer implements the basic inter/intra-predictions independently along with the inter-layer motion-compensation prediction. Although the inter-layer residual prediction used in spatial scalability exploits both the spatial as well as temporal correlations between the layers, this prediction is not efficient and fails to provide gain for most of the sequences [6]. The inter-layer motion prediction scheme also has constraints while assigning MB modes for the enhancement layer, since each  $P \times Q$  MB partition in the co-located  $8 \times 8$  block is mapped as  $(2P) \times (2Q)$  MB partition in the enhancement layer. For example, for the inter-predicted MB modes of the base layer such as Inter\_16 $\times$ 16, Inter\_16 $\times$ 8, Inter\_8 $\times$ 16 or Inter\_8 $\times$ 8, the MB modes derived for enhancement layer is Inter\_16 $\times$ 16, which may not be sufficient to represent detail information. These inter-layer prediction schemes also require up-sampling/interpolation of the motion vectors and residue of the base layer's co-located  $8 \times 8$  sub-MB [1]. However, the reconstructed frames obtained through interpolated motion vector information suffer from blocking artifacts [7]. Researchers have focused on the quality of this interpolation, which greatly affects the prediction accuracy and ultimately the coding efficiency [2, 8, 9]. Researchers have proposed different motion compensated approaches for spatially enhanced layers. Some have adaptively chosen pyramid method, sub-band method or independent

\*For correspondence

encoding method to improve the coding performance with extended edge prediction to reduce block artifacts [10]. However, this method demands encoding of the extra side information [11] specifying the optimal motion compensated method being used and creates bit rate overhead when quantization step size of the base layer is larger than that of the enhancement layer.

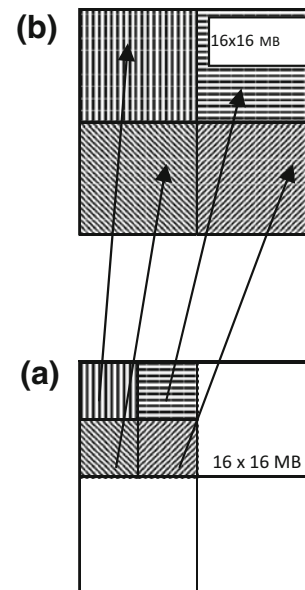
While encoding the enhancement layer, inter-layer inter prediction shows degradation of the peak signal to noise ratio (PSNR) due to (1) limitations of assigning MB modes for enhancement layer, (2) the minimum MB partition as  $\text{Inter}_{8 \times 8}$  and (3) use of up-sampling/interpolation of the motion vectors and residue. The proposed scheme in this paper uses improved MB mode decision scheme for the enhancement layer and avoids the use of up-sampling and multiloop decoding. A spatial-resolution-ratio-based MB mode decision scheme is proposed to estimate MB modes at the enhancement layers. The motion vectors and residue for this derived MB are obtained through a spatial-temporal search scheme at the enhancement layer. These encoded motion vectors and residue bits are then controlled with an efficient independent rate control (IRC) scheme [12] using bit calculation model and the rate controller. This IRC scheme focuses on bit budget allocation and its availability to every basic unit in a group of pictures (GOP) for each frame in the respective layer. Further, this paper proposes modifications to this scheme, to recursively update the mean absolute difference (MAD) parameters and the quadratic rate-distortion model parameters of each basic unit, for all the layers independently.

This paper attempts to provide a simplified motion search scheme while maintaining or improving the coding efficiency for spatial scalability with better rate control. The resulting encoded bit stream is in the standard SVC syntax.

The contents of the paper are as follows. Section 2 describes the existing concepts of inter/intra-layer prediction. Section 3 includes the proposed modified inter-layer prediction scheme along with IRC scheme for spatial scalability. Section 4 provides the experimental results comparing the existing inter-layer prediction and modified inter-layer prediction schemes. Section 5 includes conclusion and a list of references is provided.

## 2. Inter-layer prediction in scalable video coding

The scalable extension of H.264/AVC specifies a layered structure for spatial, temporal, quality and combination scalabilities as per the requirements of the application. Each layer uses an independent hierarchical motion prediction structure along with the inter-layer motion-compensated predictions [1]. The inter-layer prediction scheme exploits the redundancy between the spatial layers for motion and texture data. H.264 SVC stream indicates the enhancement layer inter-layer dependency by flags like the base mode



**Figure 1.** (a) Co-location of enhancement layer MB to base layer MB partition for dyadic spatial scalability. (b) Enhancement layer MB mode assigned from co-located  $8 \times 8$  block of base layer.

flag, inter-layer residue prediction flag and inter-layer motion prediction flag [1–3, 11]. In such cases, the MB mode, motion vectors and reference picture indices are to be inferred from the co-located  $8 \times 8$  MB from the reference layer as shown in figure 1.

Additional details of inter-layer prediction schemes are as below [1, 8, 13].

### 2.1 Inter-layer intra-prediction

For a  $16 \times 16$  MB in the enhancement layer, if the base mode flag is equal to 1 and if its co-located MB partition ( $8 \times 8$ ) in reference layer (base layer) MB is intra-coded, then it is coded as an INTRA-BL mode. The residue for this MB in the enhancement layer is obtained by taking the difference between the original MB in the enhancement layer and up-sampled decoded and reconstructed co-located MB partition in base layer [1, 2, 8, 13]. However, this increases the complexity of the decoder.

### 2.2 Inter-layer inter-prediction

Similarly, if the co-located MB partition in the reference layer is coded using inter-mode, then the MB in the enhancement layer is coded as  $\text{Inter}_{16 \times 16}$ ,  $\text{Inter}_{16 \times 8}$ ,  $\text{Inter}_{8 \times 16}$  or  $\text{Inter}_{8 \times 8}$  depending on the motion vectors of the co-located  $8 \times 8$  block as shown in table 1. These MB modes assigned to the enhancement layers are based on upsampled MB partition of the co-located  $8 \times 8$  block of the reference layer (base layer) shown in figure 1 and as

**Table 1.** MB mode assignment for enhancement layers with EILD scheme and proposed MILD scheme.

Base layer MB mode	Base layer co-located 8×8 block MB mode	Enhancement layer MB mode	
		EILD prediction	MILD prediction
Inter_16×16	Inter_8×8	Inter_16×16	Inter_16×16
Inter_16×8	Inter_8×8	Inter_16×16	Inter_16×8
Inter_8×16	Inter_8×8	Inter_16×16	Inter_8×16
Inter_8×8	Inter_8×8	Inter_16×16	Inter_8×8
Inter_8×4	Inter_8×4	Inter_16×8	Inter_8×4
Inter_4×8	Inter_4×8	Inter_8×16	Inter_4×8
Inter_4×4	Inter_4×4	Inter_8×8	Inter_4×4

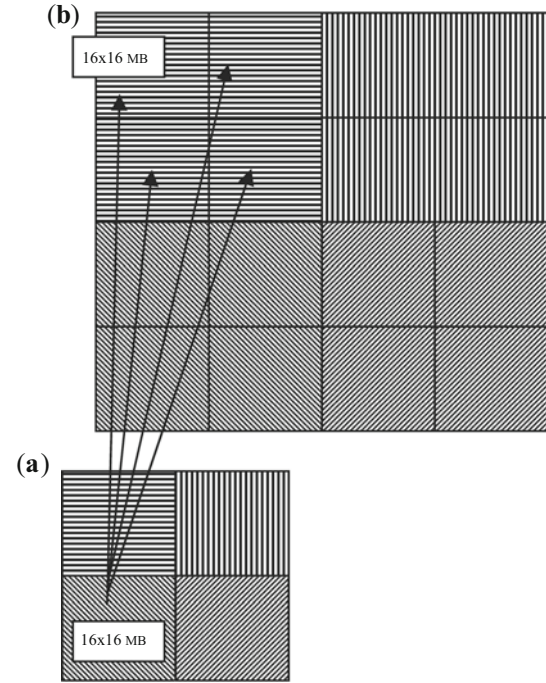
defined by G.8.6.1.3 in [11]. The residue for this MB (MB mode derived as explained above) is obtained by taking the difference between the residues obtained through motion compensation of the MB in the enhancement layer and up-sampled prediction residue of the co-located MB partition in the base layer. This residue and the up-sampled motion vectors of the co-located base layer MB are encoded using reference layer indices [1, 2, 8, 13].

Also since the base layer's quantized motion vectors and residue data are used for encoding the enhanced layers, it demands higher quality for the base layer. When the lower spatial layer has good quality, the enhancement layer's quality is highly affected by the existing up-sampling method within the given constraints of the available bandwidth [14]. This noise may get pronounced under stringent bit-rate constraints as more texture bits are needed for encoding enhancement layers.

### 3. Proposed prediction method

#### 3.1 Modified inter-layer prediction

The motion estimation and compensation can be carried out independently for each enhancement layer by disabling the inter-layer dependency. But this may reduce the speed of encoding, since it is carried out independently for all the layers and all the MBs separately. In the proposed modified inter-layer dependency (MILD) scheme, the inter-layer dependency is disabled, but the MB mode estimated at the base layer along with the intra-modes are used to estimate the motion of the respective MBs for the enhancement layers. The base layer is encoded using single-layer encoding techniques [4, 5]. For encoding the MB for enhancement layers, base layer MB modes are assigned to the enhancement layer MBs by identifying the co-located MB partition in the base layer using (1)–(4) and as shown in figure 2. For dyadic spatial scalability, the enhancement layer MB is mapped with the base layer MB for MB mode decision as shown below.

**Figure 2.** (a) Assignment of base layer MB modes to enhancement layer MBs. (b) Enhancement layer MB mode derived from co-located 16×16 MB from base layer.

$$\text{Base}_{\text{MBROW}} = \text{Enh}_{\text{MBROW}} / \text{ScaleSize} \quad (1)$$

$$\text{Base}_{\text{MBCOLUMN}} = \text{Enh}_{\text{MBCOL}} / \text{ScaleSize} \quad (2)$$

$$\text{Base}_{\text{MB}} = \text{Enh}_{\text{MBROW}} * K_{\text{BC}} + \text{Enh}_{\text{MBCOL}} \quad (3)$$

where

$$\text{ScaleSize} = \frac{\text{enhancement layer width}}{\text{base layer width}} \quad (4)$$

$$K_{\text{BC}} = \text{base layer width.}$$

$$\text{Base}_{\text{MB}} = \text{base MB number.}$$

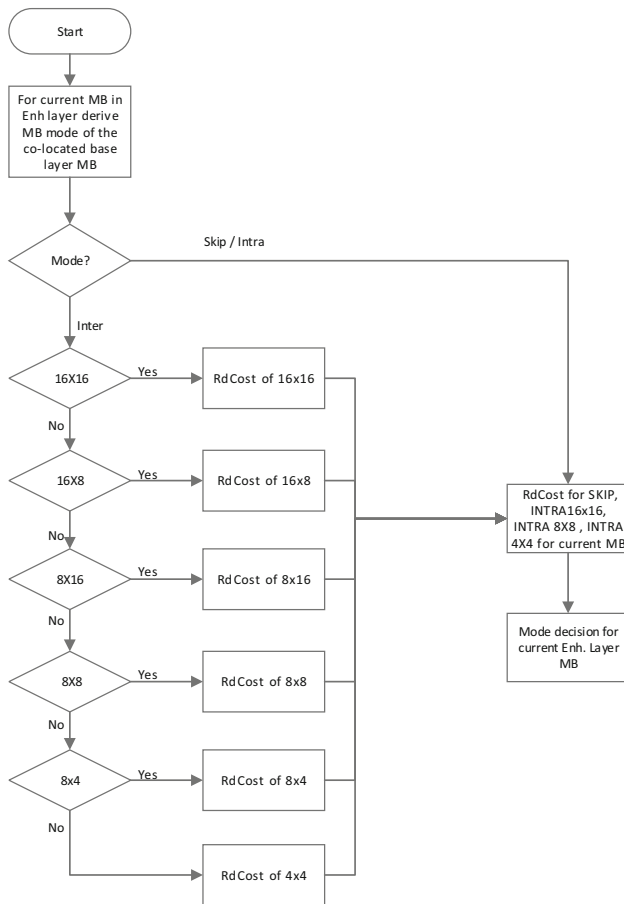
$$\text{Enh}_{\text{MBROW}} = \text{enhancement layer MB row number.}$$

$$\text{Enh}_{\text{MBCOL}} = \text{enhancement layer MB column number.}$$

$$\text{Base}_{\text{MBROW}} = \text{base layer MB row number.}$$

$$\text{Base}_{\text{MBCOLUMN}} = \text{base layer MB column number.}$$

With MILD scheme, the MB mode assignment for the enhancement layer MBs includes Inter\_16×16, Inter\_16×8, Inter\_8×16, Inter\_8×8, Inter\_8×4, Inter\_4×8 and Inter\_4×4. Table 1 compares the MB mode assignment for an enhancement layer MB with existing inter-layer dependency (EILD) and MILD prediction techniques. This scheme at the enhancement layers has enabled the use of smaller MB partitions.



**Figure 3.** Flowchart for mode decision for MB at the enhancement layer.

The enhancement layer MB is assigned the MB mode of the co-located  $16 \times 16$  base layer MB. If this new MB mode is inter-predicted, then the motion is estimated at the enhancement layer using temporal prediction to arrive at the best match with the lowest cost using Lagrange method [15]. This cost is compared with the cost computed using spatial INTRA\_16x16, INTRA\_8x8, INTRA\_4x4 and skip mode to decide the MB mode for the enhancement layer MB as shown in figure 3. If the new MB mode is intra-coded, then the best match is estimated using skip/spatial intra-modes. The MB mode having the least R-D cost among SKIP mode, INTRA\_16x16, INTRA\_8x8 and INTRA\_4x4 is used for encoding. The residue information and the motion vectors thus obtained are encoded using a quantization parameter (QP) obtained by modified IRC scheme as shown in figure 4.

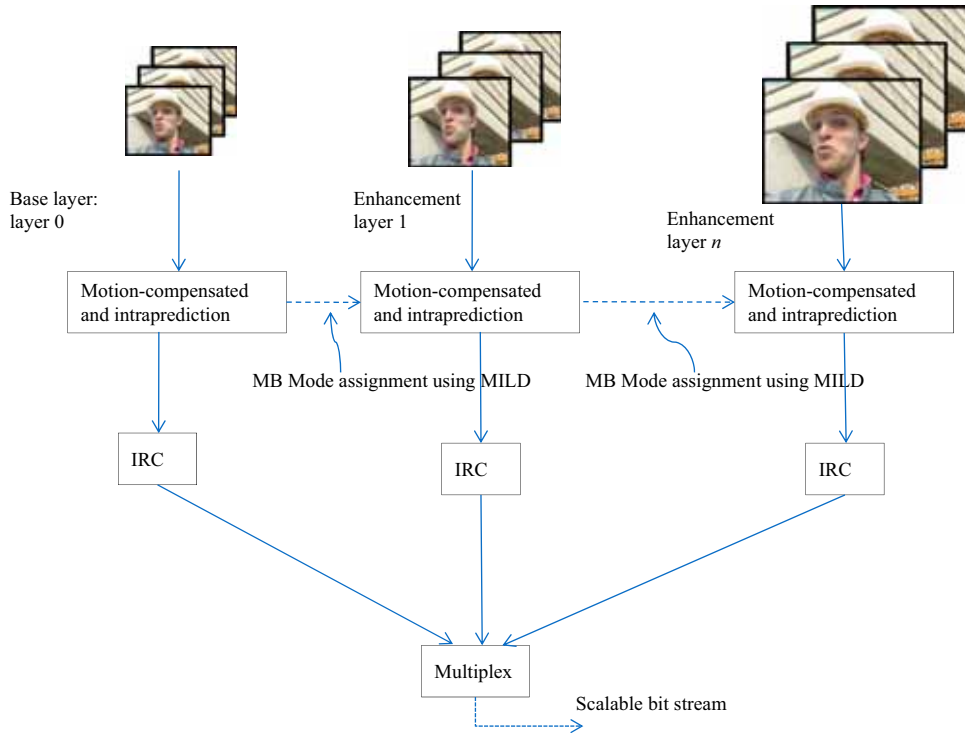
For EILD, the residue of the enhancement layer MB mode (derived from co-located  $8 \times 8$  sub-MB of base layer) is obtained using temporal search at enhancement layer. The constraints mentioned above for MB mode assignment for enhancement layer along with the quality of

upsampling/interpolation limit the performance of EILD. This MILD method uses the best estimated MB partition from base layer or spatial layer and determines the motion vectors and residual data at the enhancement layer, increasing the precision of motion estimation and compensation.

### 3.2 IRC

In video communication, the rate control schemes ensure that the bit stream is encoded with the maximum possible PSNR and is utilizing the available bandwidth. This is rate distortion optimization (RDO). The rate control algorithms focus on maintaining the bit rate constant while improving the quality of the video signal for the available channel bandwidth. The SVC reference software—Joint Scalable Video Model (JSVM) 9.19.15 provided rate control scheme for a base layer. IRC is used to extend the rate control for the enhancement layers.

As per the syntax suggested in the spatially scalable video encoding scheme [11], each frame (spatially enhanced) gets encoded for every layer before encoding the next frame in the sequence. IRC works separately for all the layers, and the bit budget is allocated per frame per GOP as per the bit rate assigned to a layer. The first I and P frame is encoded using initial QP. Each frame utilizes some actual bits as per the complexities involved, and accordingly the target bits are updated after encoding every frame in every GOP for every layer. This update of target bits for a frame conforms to the hypothetical reference decoder (HRD) [12, 16]. It is based on the parameters such as layer bandwidth, remaining bits available considering the GOP bit budget, the current buffer fullness, the complexity of the previous basic unit and the lower and upper bounds of the target bits. Finally, these target bits are used to compute QP using the quadratic rate-distortion model. However, in the earlier art, for computing QP for each frame, the quadratic rate-distortion model parameters  $X1$ ,  $X2$  and  $MAD$  parameters  $C1$ ,  $C2$  controlling the target bits were not updated separately as per the layers. The buffers stored the history of these parameters for the frames in SVC encoding sequence. It led to inaccuracy for determining QP for each frame creating errors in the bit budget allocations per frame per layer. It showed inconsistency for the base layer's performance across both the EILD and MILD algorithms. The EILD and MILD are the techniques used for encoding enhancement layers and hence, irrespective of these implementations, the base layer's PSNR and bit rate performance should be the same. Hence, the IRC was modified to isolate the buffers associated with QP calculations. The modified IRC updates the  $X1$ ,  $X2$ ,  $C1$  and  $C2$  parameter independently for every frame taking into account the history of QP and bits required for encoding the frame for



**Figure 4.** Proposed MILD scheme within SVC.

the respective layer. Thus, the QP is determined independently as explained below in Step 4.

The steps for IRC scheme for  $K$  spatially enhanced layers, for every  $i$ th frame in  $j$ th GOP are given as follows [12].

#### Step 1 Initialization

- The bits assigned per frame are the function of bit rate  $br(n_k)$  and frame rate  $Fr(n_k)$  for the respective  $k$ th layer

$$bfr(n_k) = \frac{br(n_k)}{Fr(n_k)} \quad (5)$$

- Current buffer fullness is initialized as

$$Bc(n_{i,j,k}) = 0 \quad (6)$$

where  $n_{i,j,k}$  is the  $i$ th frame in  $j$ th GOP in  $k$ th layer

- The total bits allocated per GOP are

$$TGOP(n_{i,j,k}) = \frac{br(n_k)}{Fr(n_k)} * N_{GOP} - Bc(n_{N_{GOP}-1,k}) \quad (7)$$

where  $N_{GOP}$  = number of frames in a GOP

- In the beginning, remaining bits are initialized to a value equal to the total bits allocated for the GOP

$$Rembits(n_{1,j,k}) = TGOP(n_{i,j,k}) \quad (8)$$

- To avoid abrupt changes in the visual quality of the frames, parameters such as lower bound  $LB(n_{i,j,k})$  and upper bound  $UB(n_{i,j,k})$  are used which are initialized as

$$LB(n_{1,j,k}) = Rembits(n_{0,1,k}) * \frac{br(n_k)}{Fr(n_k)} \quad (9)$$

$$UB(n_{1,j,k}) = Rembits(n_{0,1,k}) * br(n_k) * \omega \quad (10)$$

where remaining bits of the  $(j-1)$ th GOP  $Rembits(n_{0,1,k}) = 0$  and  $\omega = 0.9$ ;

**Step 2** QP for first I frame: First I frame in the first GOP in every layer is encoded in  $N(n_{1,1,k})$  bits using initial QP, which is defined before

#### Step 3 Update

- From the next frame (i.e.,  $i = 2$ ) onwards the following parameters are calculated and updated

$$\Delta B(n_{i,j,k}) = N(n_{i,j,k}) - bfr(n_k) \quad (11)$$

where  $N(n_{i,j,k})$  = bits used for encoding the  $i$ th frame in  $j$ th GOP in  $k$ th layer

- $\Delta B(n_{i,j,k})$  are the error bits, which are further used to compute target bits for the current coding frame.
- The video encoding standards have mentioned a virtual buffering mechanism to control the bit rate

fluctuations. The HRD would then decode the video frames without overflow or underflow. The current buffer fullness, the remaining bits with the GOP, the lower bound, the upper bound parameters and the target buffer level are updated as follows.

$$Bc(n_{i,j,k}) = Bc(n_{i-1,j,k}) - \Delta B(n_{i-1,j,k}) \quad (12)$$

$$Bc(n_{0,j+1,k}) = Bc(n_{i,N_{GOP},k}) \quad (13)$$

$$Rembits(n_{i,j,k}) = Rembits(n_{i-1,j,k}) - N(n_{i-1,j,k}) \quad (14)$$

$$LB(n_{i,j,k}) = LB(n_{i-1,j,k}) - \Delta B(n_{i-1,j,k}) \quad (15)$$

$$UB1(n_{i,j,k}) = UB1(n_{i-1,j,k}) - \Delta B(n_{i-1,j,k}) \quad (16)$$

$$UB2(n_{i,j,k}) = UB2(n_{i-1,j,k}) * \omega \quad (17)$$

$$Tbl(n_{i,j,k}) = Tbl(n_{i-1,j,k}) - \frac{Tbl(n_{2,j,k})}{N_p - 1} \quad (18)$$

where  $Tbl(n_{i,j,k})$  is the target buffer level  
 $N_p$  = number of P frames in the GOP.

- After coding the first P frame in  $j$ th GOP, the target buffer level is initialized as

$$Tbl(n_{2,j,k}) = Bc(n_{2,j,k}) \quad (19)$$

- Target bits are the weighted combination of target bits allocated  $\tilde{R}(n_{i,j,k})$ , and computed target bits  $\hat{R}(n_{i,j,k})$ ,

$$R(n_{i,j,k}) = \beta * \hat{R}(n_{i,j,k}) + (1 - \beta) * \tilde{R}(n_{i,j,k}) \quad (20)$$

where  $\beta$  is a constant with a typical value of 0.75.

- Target bits allocated are

$$\tilde{R}(n_{i,j,k}) = bfr(n_k) + \Upsilon * (Tbl(n_{i,j,k}) - Bc(n_{i,j,k})) \quad (21)$$

where  $\Upsilon$  is a constant with a typical value of 0.5.

- Target bits computed are

$$\hat{R}(n_{i,j,k}) = \frac{Rembits(n_{i,j,k})}{N_{p,r,i-1}} \quad (22)$$

where  $N_{p,r,i-1}$  = number of P frames remaining.

- The target bits are further modified to achieve smooth visual effect using

$$R(n_{i,j,k}) = (1 - LB(n_{i,j,k}) * 0.05) * R(n_{i,j,k}) \quad (23)$$

- The target bits are further bounded using the lower bound and upper bound limits to conform to HRD

$$R(n_{i,j,k}) = \max((LB(n_{i,j,k}) - R(n_{i,j,k})) \quad (24)$$

$$R(n_{i,j,k}) = \min((LB(n_{i,j,k}) - R(n_{i,j,k})) \quad (25)$$

#### Step 4 To compute QP

In the prior art, the parameters such as remaining bits, buffer fullness, lower bound limits, upper bound limits and target bits are updated iteratively for all the basic units, in all the GOPs in the respective layers ( $i$ th frame in  $j$ th GOP in the  $k$ th layer) as explained above.

As per the proposed modification in IRC technique, actual bits assigned to a basic unit using MAD of an  $i$ th frame in  $j$ th GOP is computed independently for all the layers. The quadratic rate-distortion model parameters  $X_1$ ,  $X_2$  for a frame are computed as per Eqs. (27) and (28). They are a function of previous bits required, the previous QP assigned and window size (buffer size). These  $X_1$ ,  $X_2$  parameters are updated using a linear regression method [17] as shown below. These updates are done independently for all the layers to determine the  $Q_{step}$ , i.e., quantization step size.

$$R(n_{i,j,k}) = \frac{X_1 * MAD_{i,j,k}}{Qstep^2} + \frac{X_2 * MAD_{i,j,k}}{Qstep} \quad (26)$$

$X_2$

$$X_2 = \frac{w \sum_{v=1}^w R(n_{i,j,k})^v - \left( \sum_{v=1}^w (Qstep_{i,j,k}^v)^{-1} \right) \left( \sum_{v=1}^w (Qstep_{i,j,k}^v * R(n_{i,j,k})^v) \right)}{w \sum_{v=1}^w (Qstep_{i,j,k}^v)^{-2} - \left( \sum_{v=1}^w (Qstep_{i,j,k}^v)^{-1} \right)^2} \quad (27)$$

$$X_1 = \frac{\sum_{v=1}^w \left( Qstep_{i,j,k}^v * R(n_{i,j,k})^v - X_2 * (Qstep_{i,j,k}^v)^{-1} \right)}{w} \quad (28)$$

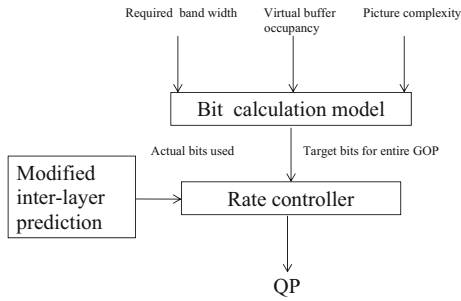
where  $w$  = buffer size.  $R(n_{i,j,k})$  are the actual bits used for encoding the basic unit,  $Qstep_{i,j,k}^v$  is the quantization step size of an  $i$ th frame for the  $j$ th GOP of the  $k$ th layer. While computing  $Qstep$  from the above equation,  $MAD_{i,j,k}$  is not known and is predicted using previous frame in the  $j$ th GOP

$$MAD_{i,j,k} = C1 * MAD_{i,j-1,k} + C2 \quad (29)$$

where  $C1$  and  $C2$  are the two coefficients of the prediction model with initial values as 1 and 0, respectively. They are updated through linear regression method after coding each frame independently in the layers. Figure 5 shows the integration of IRC with MILD scheme.

## 4. Experimental results

The proposed technique was validated using SVC reference software—Joint Scalable Video Model (JSVM) 9.19.15. Rate control for the enhancement layer was implemented using the modified IRC scheme. The performance evaluation of the proposed MILD was tested on various benchmark video sequences shown in table 2.



**Figure 5.** System structure of modified IRC algorithm with MILD.

**Table 2.** Video sequences and motion type.

Test sequence	Motion
Four people	Head and shoulder
Kristen and Sara	
Stockholms Shields	Panoramic
Ice	
Basketball drive	Fast motion
Kimono	
Park run	
	Background and object moving in opposite direction

**Table 3.** Summary of simulation parameters for configuration file.

Simulation parameter	Setting
Base layer mode	AVC compatible
GOP size	16
Base mode flag	1/0
Inter-layer residue prediction flag	1/0
Inter-layer motion prediction flag	1/0

**Table 4.** Performance comparison for modified inter-layer dependency and the existing inter-layer dependency.

Test sequences	Layer 0 PSNR – dB		Layer 1 PSNR – dB		Layer 2 PSNR – dB		Layer 3 PSNR – dB	
	Existing	Proposed	Existing	Proposed	Existing	Proposed	Existing	Proposed
Four people	32.72	32.72	37.86	37.85	36.25	38.47	36.25	36.31
Kristen and Sara	35.01	35.01	39.85	41.19	38.15	40.17	38.28	39.54
Stockholm	32.87	32.87	32.75	34.43	29.27	30.59	28.65	29.65
Shields	29.46	29.46	29.63	31.20	26.07	28.48	22.94	25.99
Ice	28.50	28.50	34.05	36.89	35.25	37.96	36.94	39.06
Basketball drive	25.99	25.99	28.73	31.12	29.13	30.35	29.19	29.28
Kimono	28.29	28.29	30.48	32.30	30.47	30.90	30.80	30.84
Park run	25.54	25.54	24.72	26.62	22.50	23.54	19.97	20.81

These video sequences are indicative, and analysis can be further extended for different test sequences. The following video test sequences with varying spatial resolution and carrying different motion were tested.

Implementation was carried for four spatial layers:

- enhancement layer3 – L3;HD-Resolution: 1280×720; target bit rate: 1984 kbps
- enhancement layer2 – L2;4CIF-Resolution: 704×576; target bit rate: 1344 kbps
- enhancement layer1 – L1;CIF-Resolution: 324×288; target bit rate: 704 kbps
- base layer – L0; QCIF Resolution: 176×144; Target Bit rate: 64 kbps.

For each layer, the difference between the target bit rates is set at 640 kbps for experimentation purpose. Analysis can further be carried out for different combinations of bit rates.

The input parameters in the configuration file for the JSVM 9.19.15 reference software were modified from their default values as shown in table 3.

The base mode flag, inter-layer residue flag, and inter-layer motion flag are set to 1 to enable EILD and are set to 0 to enable MILD—the proposed scheme. The base layer is encoded using H.264AVC mode for both the techniques.

Table 4 compares the PSNR achieved by the MILD and EILD for four spatially enhanced layers. The PSNR of the base layer is the same for both the techniques since it is encoded as H.264AVC single layer. However, PSNR for the layers 1–3 using the proposed technique is greater than the existing technique in general. Also, PSNR for fast moving video sequences is on a lower side for both the techniques, which can be improved by allocating more target bits.

The fluctuations in bit rate for the video sequences mentioned above are as shown in figure 6. M-Lx represents MILD and E-Lx represents EILD with  $x$  representing layer number. The IRC scheme used for both these techniques adaptively changes the QP. For base layer both these

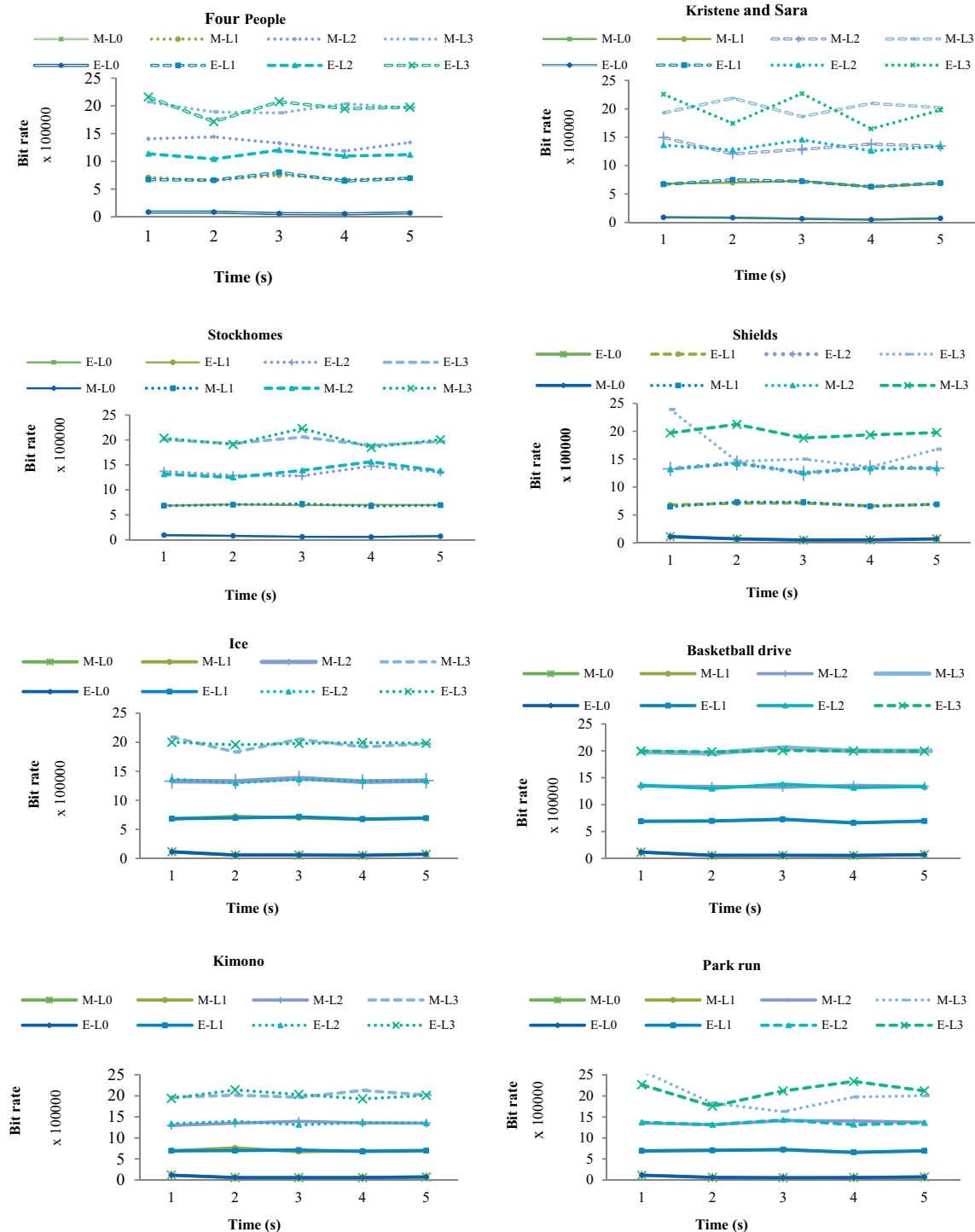


Figure 6. Comparative analysis of achieved bit rate using existing and proposed modified inter-layer dependency scheme.

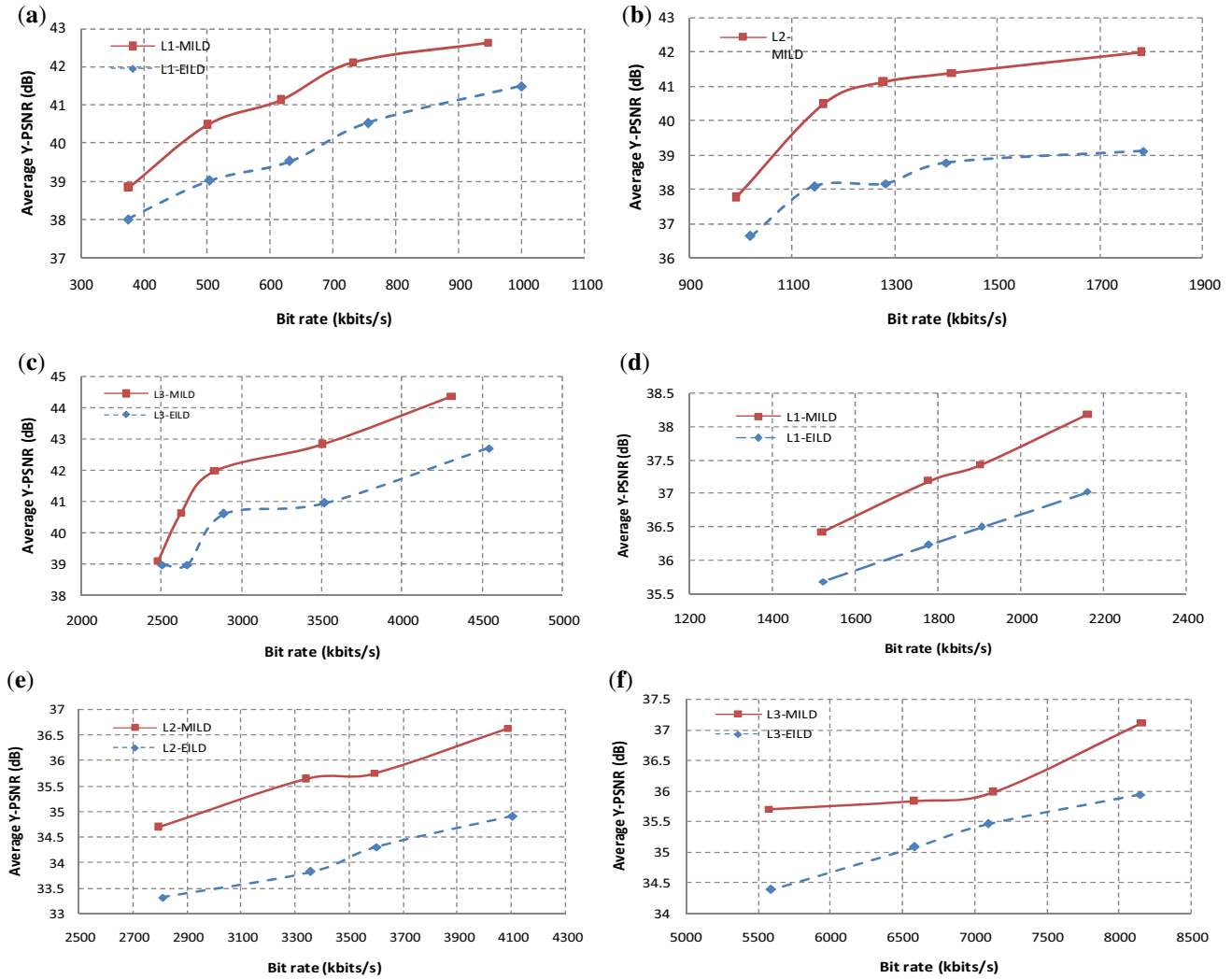
schemes provide similar bit rate control. With the increase in enhancement layers, the bits required for encoding the up-sampled motion vectors and residue data also increases, especially for video sequences having slow motion.

Figure 7 shows the coding efficiency for L1, L2 and L3 enhancement layers using MILD and EILD schemes with modified IRC. The base layer’s bit rate was set to 96 kbps

for Kristen and Sara and 384 kbps for basketball drive video sequence as suggested by JVT-test conditions [2]. It can be observed that coding efficiency is maintained or improved using MILD for slow sequence – Kristen and Sara and fast sequences – basketball drive.

Table 5 indicates that the computational performance for MILD scheme is at par or better for some of the video





**Figure 7.** Comparison of MILD and EILD for enhancement layer coding efficiency. (a) L1-Kristene and Sara, CIF. (b) L2-Kristene and Sara, 4CIF. (c) L3-Kristene and Sara, HD. (d) L1-Basketball drive, CIF. (e) L2-Basketball drive, 4CIF. (f) L3-Basketball drive, HD.

**Table 5.** Analysis of reduction in time.

Video sequence	Time (ms) EILD – $T_{EILD}$	Time (ms) MILD – $T_{MILD}$	Reduction in time= $((T_{EILD} - T_{MILD})/T_{EILD}) * 100(\%)$
Four people	1725090	1555548	9.828009
Kristen and Sara	1490108	1630361	-9.41227
Stockholm	1712573	1821318	-6.3498
Shields	1768882	1950281	-10.255
Ice	3570789	2077062	41.83185
Basketball drive	2219020	2148077	3.197042
Kimono	1774351	2070942	-16.7155
Parkrun	2423424	2245781	7.330248

sequences than EILD scheme. It is seen that the performance is improved between 3% and 41% for four video sequences, whereas it marginally degrades between 6% and 9% for two sequences and for the remaining two it degrades between 10% and 16%. The video sequences with more homogeneous

regions and having fast motions show improvement in performance. Hence, the base layers MB modes are the best suitable in such cases. These illustrative results are based on experiments conducted on SVC reference software – Joint Scalable Video Model (JSVM) 9.19.15.

## 5. Conclusion

This paper proposes a MILD scheme for efficiently determining MB modes of enhancement layers in dyadic scalability. The spatial-resolution-ratio-based MB mode decision scheme computes the motion vectors and residual data at the enhancement layer using estimated MB partition from base layer. The proposed spatial-temporal search scheme has increased the precision of motion estimation and compensation, leading to improved results as compared with the existing inter-layer dependency scheme in spatial scalability. It is observed that coding efficiency is improved or maintained for fast and slow motion sequences. This scheme is computationally efficient for fast motion video sequences.

In addition, we have improved our prior work on independent rate control scheme for bit budget allocation by recursively updating the MAD and the quadratic rate-distortion model parameters. Using this improved scheme, modified inter-layer dependency provides better rate control than when it is applied to the existing inter-layer dependency.

## References

- [1] Schwarz H and Marpe D 2007 Overview of the scalable video coding extension of the H.264/AVC Standard. *IEEE Trans. Circuits Syst. Video Technol.* 17(9): 1103–1120
- [2] Segall C and Sullivan G 2007 Spatial scalability within the H.264/AVC scalable video coding extension. *IEEE Trans. Circuits Syst. Video Technol.* 17(9): 1121–1135
- [3] Reichel J, Schwarz H and Wein M 2005 *Scalable video coding-working draft 1*. Joint Video Team of ITU-T Vceg and ISO/IEC MPEG, Doc.JVT-N020, Hong Kong, CN
- [4] Richardson E G 2002 *Video codec design*. West Sussex PO19 IUD, England, Wiley
- [5] Richardson E G 2003 *H.264 and MPEG-4 Video Compression*. West Sussex, England: Wiley
- [6] Xiong R, Xu J and Wu F 2008 In-scale motion compensation for spatially scalable video coding. *IEEE Trans. Circuits Syst. Video Technol.* 18(2): 145–158
- [7] Li M, Chandrasekhar P, Dane G and Nguyen T 2007 Inter-layer motion vector interpolation for low-complexity and very low bitrate scalable video coding. In: *Proceedings of the Forty-First IEEE Asilomar Conference on Signals, Systems and Computers*, pp. 231–234, doi:10.1109/ACSSC.2007.4487202
- [8] Han J, Melkote V and Rose K 2014 An estimation-theoretic framework for spatially scalable video coding. *IEEE Trans. Image Process.* 23(8): 3684–3697
- [9] Segall C and Katsageslos A 2006 Resampling for spatial scalability. In: *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, pp. 181–184, doi:10.1109/ICIP.2006.312364
- [10] Zhang R and Comer M 2008 Efficient inter-layer motion compensation for spatially scalable video coding. *IEEE Trans. Circuits Syst. Video Technol.* 18(10): 1325–1334
- [11] ITU-T H.264 Series H 2010 Audio Visual and Multimedia Systems Infrastructure of Audiovisual services—coding of moving video
- [12] Jadhav P and Kshirsagar S 2013 Independent rate control scheme for spatial scalable layer in H.264SVC. In: *Proceedings of the Fifth IEEE International Conference on Advances in Recent Technologies in Communication and Computing (ARTCom 2013)*, pp. 52–58, doi:10.1049/cp.2013.2234
- [13] Schwarz H, Hinz T, Marpe D and Weigand T 2005 Constrained inter-layer prediction for single loop decoding in spatial scalability. In: *Proceedings of ICIP(2)*, IEEE, pp. 870–873, doi:10.1109/ICIP.2005.1530194
- [14] Shin I and Park H 2009 Adaptive up-sampling method using DCT for spatial scalability of scalable video coding. *IEEE Trans. Circuits Syst. Video Technol.* 19(2): 206–214
- [15] Li X, Amon P, and Hutter A 2009 Lagrange multiplier selection for rate-distortion optimization in SVC. In: *Proceedings of the Thirtieth Meeting*, Geneva, CH, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG(ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6)
- [16] Li Z, Gao W, Pan F, Ma S, Lim K P, Feng G, Lin X, Rahardja S, Lu H and Lu Y 2003 Adaptive rate control with HRD consideration. In: *Proceedings of the eighth Meeting*, Geneva, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG(ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6)
- [17] Lu X and Martin G 2014 An improved rate control algorithm for SVC with optimised MAD prediction. In: *Proceedings of the Sixteenth IEEE International Workshop on Multimedia Signal Processing (MMSP)*