

Research Article

Efficient Enhancement for Spatial Scalable Video Coding Transmission

Mayada Khairy,¹ Alaa Hamdy,² Amr Elsayed,² and Hesham Farouk¹

¹Department of Computers and Systems, Electronics Research Institute, ERI, Giza, Egypt

²Electronics, Communications and Computers Department, Helwan University, Cairo, Egypt

Correspondence should be addressed to Mayada Khairy; mayada@eri.sci.eg

Received 2 October 2016; Accepted 22 December 2016; Published 17 January 2017

Academic Editor: Sebastian López

Copyright © 2017 Mayada Khairy et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Scalable Video Coding (SVC) is an international standard technique for video compression. It is an extension of H.264 Advanced Video Coding (AVC). In the encoding of video streams by SVC, it is suitable to employ the macroblock (MB) mode because it affords superior coding efficiency. However, the exhaustive mode decision technique that is usually used for SVC increases the computational complexity, resulting in a longer encoding time (ET). Many other algorithms were proposed to solve this problem with imperfection of increasing transmission time (TT) across the network. To minimize the ET and TT, this paper introduces four efficient algorithms based on spatial scalability. The algorithms utilize the mode-distribution correlation between the base layer (BL) and enhancement layers (ELs) and interpolation between the EL frames. The proposed algorithms are of two categories. Those of the first category are based on interlayer residual SVC spatial scalability. They employ two methods, namely, interlayer interpolation (ILIP) and the interlayer base mode (ILBM) method, and enable ET and TT savings of up to 69.3% and 83.6%, respectively. The algorithms of the second category are based on full-search SVC spatial scalability. They utilize two methods, namely, full interpolation (FIP) and the full-base mode (FBM) method, and enable ET and TT savings of up to 55.3% and 76.6%, respectively.

1. Introduction

With the current boom of network technology, multimedia communication has become a popular subject, especially in information technology circles. The rapid evolution of communication systems technology has been signposted by the emergence of diverse multimedia innovations such as Internet Protocol television, digital television, video telephony telecommunication, video on demand, video conferencing, and web conferencing [1]. This has led to heterogeneous demands for video sequencing with respect to the frame size, frame rate, and bit rate. There has thus been the need to use several criteria for encoding video sequencing to meet the requirements of the diverse systems, as shown in Figure 1 [2].

Scalable Video Coding (SVC) was recently standardized by the Joint Video Team (a group of video coding experts from the ITU-T study group) and ISO/IEC MPEG (Moving Picture Experts Group) [3]. The multiple coding availed by

SVC depends on the reconstruction of the lower resolution or lower quality signals from partial bit streams, as shown in Figure 2 [4, 5].

SVC enables the efficient incorporation of three types of scalabilities, namely, spatial, quality, and temporal scalabilities. The spatial and quality scalabilities can be realized by a layered approach, wherein one base layer (BL) is used to encode the lowest temporal, spatial, and quality representations of the video stream, and one or more enhancement layers (ELs) are used to encode additional information. Using the base layer as a starting point, temporal versions of the video or versions with higher qualities and resolutions can be reconstructed during the decoding process [6]. The block size of the SVC system can be varied to match the motion estimation that is used to reduce temporal redundancy between frames. It has seven modes for interprediction (skip, 16×16 , 16×8 , 8×16 , 8×8 , 8×4 , 4×8 , and 4×4) and two intramodes (4×4 and 16×16) [7, 8].

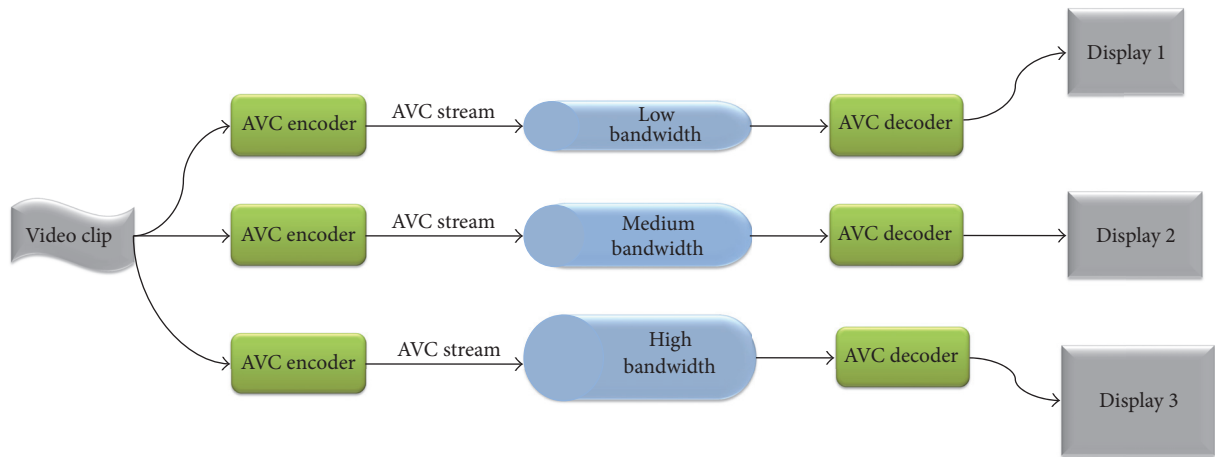


FIGURE 1: Multiple streams/simulcast.

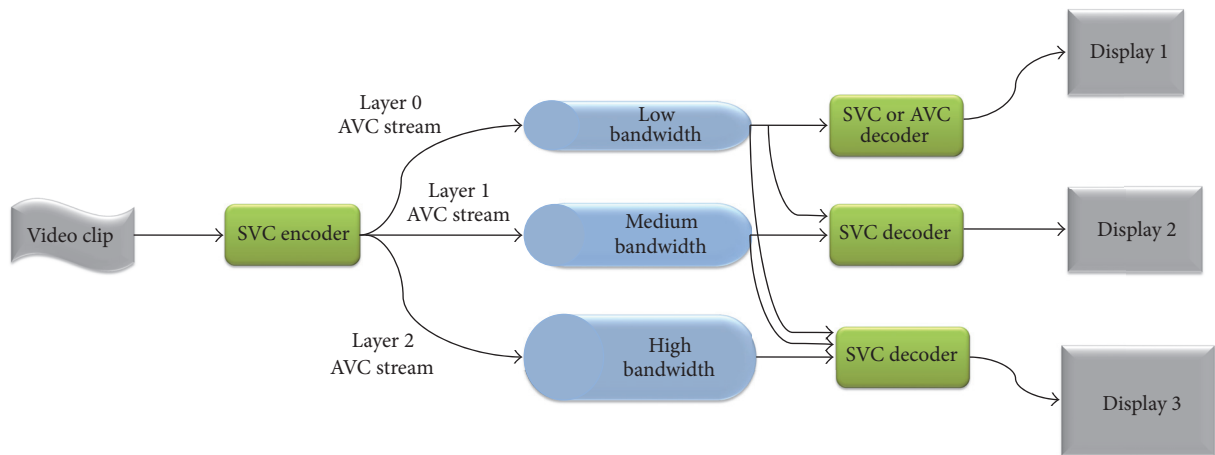


FIGURE 2: Multiple streams/scalable.

To ensure efficient video coding, additional processes are required, including interlayer prediction. The main objective of interlayer prediction is the reuse of the motion and texture information of the BL at the EL. In H.264 AVC, for each EL similar to the BL, an exhaustive search and comparison of the variable block-size partitions are required to minimize the rate-distortion (RD) cost [9].

Interlayer prediction can be done using any of three methods, namely, interlayer motion prediction, interlayer residue prediction, and interlayer intraprediction from a lower layer. Whereas all three methods enable enhancement of the coding efficiency, they increase the computational complexity of video encoding, making it difficult to realize an SVC encoder in a real application [10].

Several algorithms for fast mode decision schemes have been proposed to reduce the computational complexity. Shen et al. [4] suggested two techniques for mode decision, namely, skip mode decision and adaptive early termination, which utilize both coding information on spatial neighbor macroblocks (MBs) within the same frame and neighbor MBs

in the BL. Kim et al. [11] described an algorithm based on a 16×16 coded block pattern (CBP) mode in the current frame and the best CBP mode in the selected reference frame. Li et al. [12] developed an algorithm based on the distribution relationship between the BL and EL. The algorithm enabled reduction of the number of candidate modes. Kim et al. [13] explained that the MB in the EL layer could be predicted using the modes of a collocated MB and its neighboring MBs in the BL. Wang et al. [14] proposed an algorithm that terminated the intermode decision early in the EL by estimating the RD cost from the MB of the BL and the enhancement layer, from temporal and spatial perspectives. Kim et al. [15] suggested a fast mode search for combined scalability based on the correlation between the search directions for bidirectional motion prediction of neighboring MBs and the current MB at the encoder. Dawei et al. [16] introduced an algorithm based on early termination of the modes used in the collocated reference macroblock of all the layers. This was done by calculating the mean peak signal-to-noise ratio (PSNR) for all the frames at the temporal level for a specific mode.

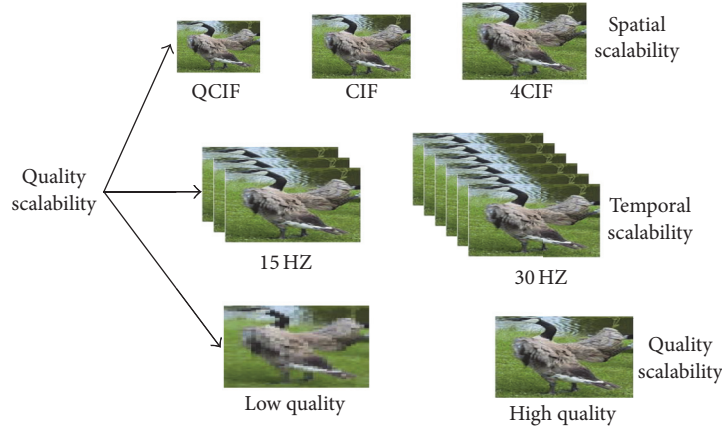


FIGURE 3: Types of scalability.

The present paper introduces four algorithms based on the sending of partial EL data, thus reducing the number of sending bits. They also shorten the transmission time (TT) across the network and reduce the amount of required EL encoding, thereby also shortening the encoding time (ET).

The rest of this paper is organized as follows. Section 2 presents an overview of the different types of scalabilities and their underlying concepts. The proposed fast mode decision is introduced in Section 3, while Section 4 describes the experiments performed using the proposed algorithms and also presents and discusses the results. The paper is ended in Section 5 with a brief presentation of the conclusions of the study.

2. Overview of Types of Scalabilities and Their Underlying Concepts

2.1. Types of Scalabilities. There are three types of scalabilities in video coding, namely, spatial scalability, quality scalability, and temporal scalability. Figure 3 shows the differences among the three types of scalabilities. Spatial scalability involves the coding of a video using multiple spatial resolutions. As shown in Figure 4, the data decoded at lower resolutions can be used to predict the data of higher resolutions to reduce the bit rate. Quality scalability is considered as a special case of spatial scalability because the generated stream can be used to predict and decode the video with different qualities, as shown in Figure 5 [17]. Conversely, in temporal scalability coding, structures containing bidirectional (B) and prediction (P) pictures in the BL are decoded as B and P frames, respectively. The EL frames are predicted using the lower temporal layer frames as the references frames. The frames in the BL and EL are used to build a group of pictures (GOP), as shown in Figure 6 [18].

2.2. Rate-Distortion Optimized Mode Selection. SVC uses the rate-distortion optimization (RDO) technique to select the best coding mode for each MB, and this enables the

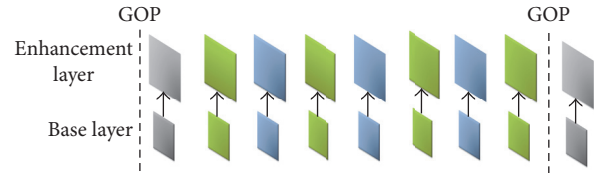


FIGURE 4: Spatial scalability.

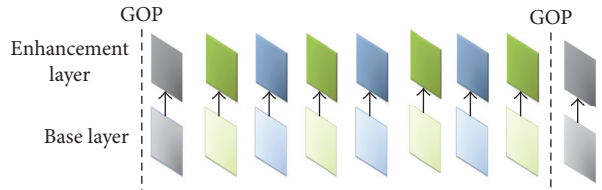


FIGURE 5: Quality scalability.

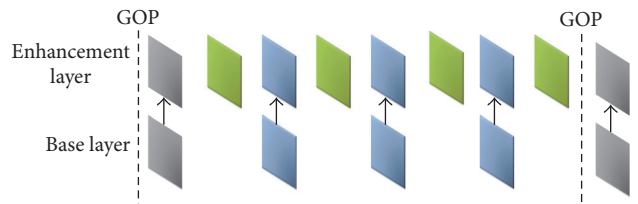


FIGURE 6: Temporal scalability.

achievement of the given bit rate with minimum distortion [19, 20]. The RDO formula is as follows:

$$RD = D + \lambda R, \tag{1}$$

where RD is the distortion cost, which is the difference between the current MB and the encoded MB; D is the sum of absolute differences (SAD) or the sum of squared differences (SSD); and R is the number of bits required to encode the current MB based on the selected mode. The

Lagrange multiplier λ is used to achieve the minimum RD through a trade-off between R and D .

2.3. Zero-Mean Normalized Cross Correlation (ZNCC). The ZNCC video analysis method was used in this study to detect the motion type of the video [21], whether fast motion, medium motion, or slow motion. It was one of the important factors considered in the trade-off among the benefits of the different proposed methods. The ZNCC equation is as follows:

$$\text{ZNCC}(F_{t_1}, F_{t_2}) = \frac{\sum_i ((H_{t_1}^i - \bar{H}_{t_1}) \times (H_{t_2}^i - \bar{H}_{t_2}))}{\sqrt{\sum_i (H_{t_1}^i - \bar{H}_{t_1})^2 \times \sum_i (H_{t_2}^i - \bar{H}_{t_2})^2}}, \quad (2)$$

where F_{t_1} and F_{t_2} , respectively, denote video frames obtained at times t_1 and t_2 , H_{t_1} and H_{t_2} denote the corresponding frame histograms, H^i is the i th. bin of the color histogram H , and \bar{H} is the mean value of all the entries of H . The ZNCC equation produces a real value between -1 and 1 , with -1 indicating no similarity whatsoever between the histograms and $+1$ indicating that they are identical [22].

2.4. Phase-Based Frame Interpolation for Video. Frame interpolation is one of the main concepts used in the development of the algorithms proposed in this work. The two most popular methods employed in frame interpolation algorithms are the Lagrangian and Eulerian methods.

Lagrangian methods are based on motion models [14, 23], while Eulerian methods are based on color change per pixel over time [24, 25]. The interpolation algorithms proposed in this paper may be considered to utilize an extension of Eulerian methods [26].

Considering input frames F_1 , F_2 , and F_{out} and denoting the steerable pyramid decompositions by P_1 and P_2 , the steps of the proposed interpolation algorithms are as follows:

- (1) Calculate $(P_1, P_2) \leftarrow$ decomposed (F_1, F_2) , $(\phi_1, \phi_2) \leftarrow$ phase (P_1, P_2) ($A_1, A_2) \leftarrow$ amplitude (P_1, P_2) [27].
- (2) Calculate $\phi_{\text{diff}} \leftarrow$ phasedifference (ϕ_1, ϕ_2) as the following equation:
$$\phi_{\text{diff}} = a \tan 2(\sin(\phi_1 - \phi_2), \cos(\phi_1 - \phi_2)). \quad (3)$$
- (3) For all $l = L - 1 : 1$ do $\tilde{\phi}_{\text{diff}}^l =$ shiftcorrelation $(\tilde{\phi}_{\text{diff}}^l)$
- (4) Calculate $\hat{\phi}_{\text{diff}} \leftarrow$ adjustphase $(\phi_{\text{diff}}, \tilde{\phi}_{\text{diff}})$ as the following equation:
$$\hat{\phi}_{\text{diff}} = \tilde{\phi}_{\text{diff}} + 'Y^* 2\pi. \quad (4)$$
- (5) Calculate $\phi_\alpha \leftarrow$ interpolate $(\phi_\alpha, \hat{\phi}_{\text{diff}}, \alpha)$ as the following equation:
$$\phi_\alpha = \phi_1 + \alpha \hat{\phi}_{\text{diff}}. \quad (5)$$
- (6) Calculate $A_\alpha =$ blend (A_1, A_2, α) .
- (7) Calculate $P_\alpha =$ recombine (ϕ_α, A_α) .
- (8) Calculate $I_\alpha =$ reconstruct (P_α) [28].

2.5. The Mode-Distribution Correlation between Macroblocks in BL and EL. SVC utilizes a BL and one or more ELs. There is a relationship between the frames in the different layers. The layers are all identical with the exception of certain parameters, which depend on the type of SVC. This similarity can be used to minimize the number of selected modes of the macroblocks. In the present study, only a limited number of macroblocks modes were selected for testing for intra- and interprediction [29], resulting in reduced encoding time. In the case of spatial SVC, the presently proposed macroblock mode of the EL has the same mode as the corresponding macroblocks in the BL, which is selected by exhaustive search [30].

3. The Proposed Fast Mode Decision Algorithms

This paper proposes four algorithms for full SVC and interlayer residual SVC spatial scalability. The goal of the proposed algorithms is the reduction of the processing time (encoding time and transmission time). Two concepts are employed for this purpose. The main concept involves sending only a part of the EL frame data, with the missing frame data generated at the decoder by interpolation using the information on the surrounding frames, as mentioned in Section 2.4. The second concept is based on the mode-distribution correlation between the BL and EL, and this is affected by the spatial scalability, as mentioned in Section 2.5. Using these two concepts, the four proposed algorithms enable significant shortening of both the TT and ET. The first two steps of the four algorithms are the same. In the first step, the BL frames are encoded through an exhaustive search. In the second step, the up sample technique is used to make the BL frames the same size as the EL frames. The subsequent steps differ among the algorithms.

3.1. Proposed Algorithms for Improving Interlayer Residual Spatial SVC

3.1.1. Interlayer Interpolation (ILIP) Algorithm. Outlined in Figure 7, this algorithm utilizes the interlayer residual concept for spatial SVC. To encode the EL frames, only the odd frames of the interlayer residuals are transmitted after being encoded by exhaustive search to achieve a high quality. This is because the transmitted frames depend on RDO, as mentioned in Section 2.2. At the decoder, the decoded odd interlayer residual frames and the even frames are derived by interpolation between the decoded odd frames. To obtain the full frames for the EL, all the interlayer residual frames are added to the BL decoded frames. This sequence reduces the sending bits, resulting in high performance with respect to the TT and ET. The results of this algorithm are described in Section 4. If shortening the TT is the most important reason for using SVC, the ILIP algorithm would be preferable.

3.1.2. Interlayer Base Mode (ILBM) Algorithm. This algorithm particularly shortens the ET relative to the TT. The main difference between this algorithm and the previous one is the use of the mode-distribution correlation between the

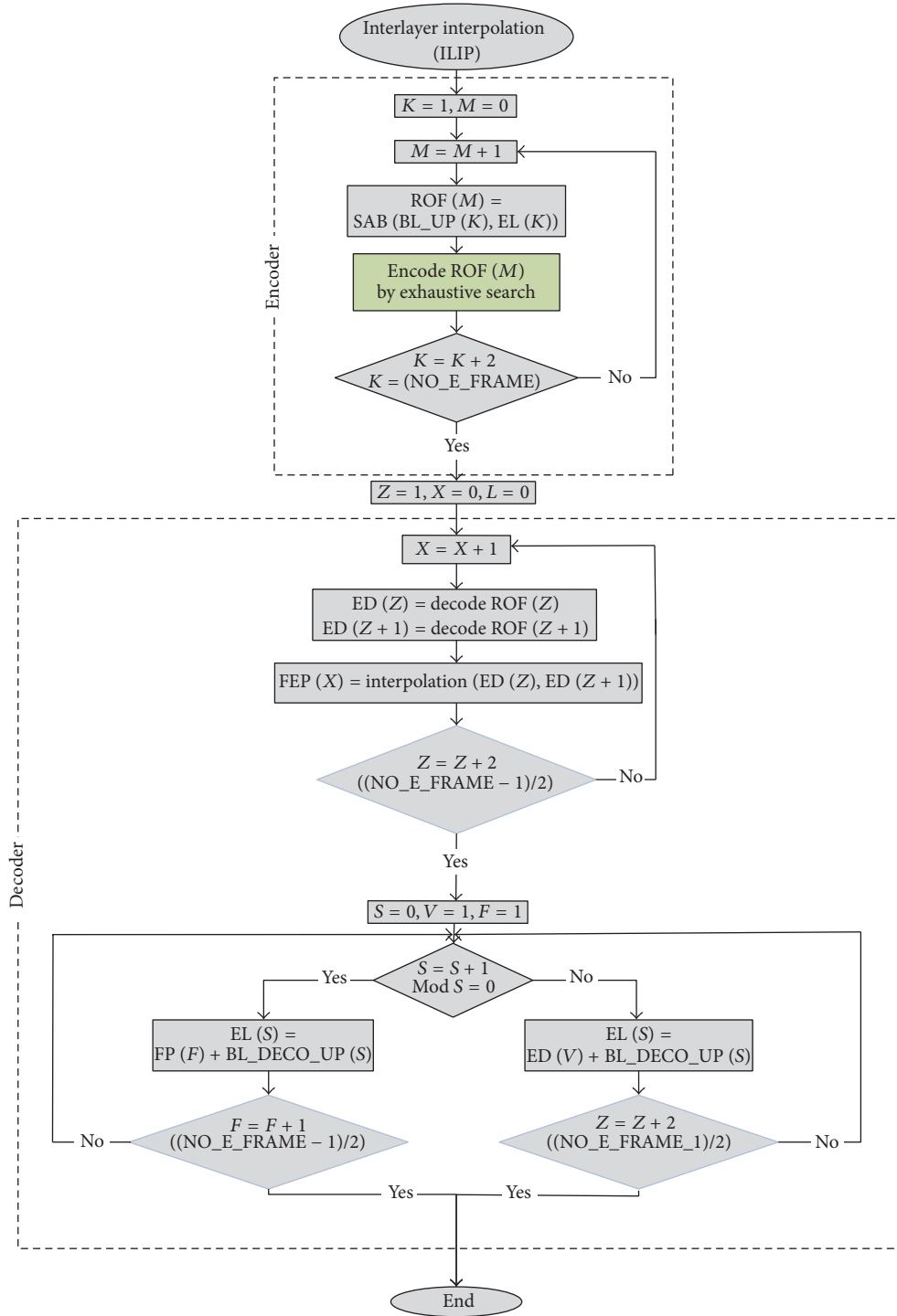


FIGURE 7: Interlayer interpolation. ROF = interlayer residual odd frame. ED = decoded odd frames. FEP = interpolation interlayer residual even frames.

BL and the EL, instead of an exhaustive search, to encode the odd frames of the interlayer residual. The use of the mode-distribution correlation saves much ET with negligible increase in the number of sending bits compared to the ILIP algorithm. However, compared to the Joint Scalable Video Model (JSVM), the ILBM algorithm significantly reduces the

number of sending bits. The flow chart of this algorithm is shown in Figure 8. When using the ILIP and ILBM algorithms, the video sequence should be encoded with a quality level lower than a certain threshold of quality (THQ). The application of the concept of the interlayer residual to SVC thus places a quality limitation. If the video sequence must be

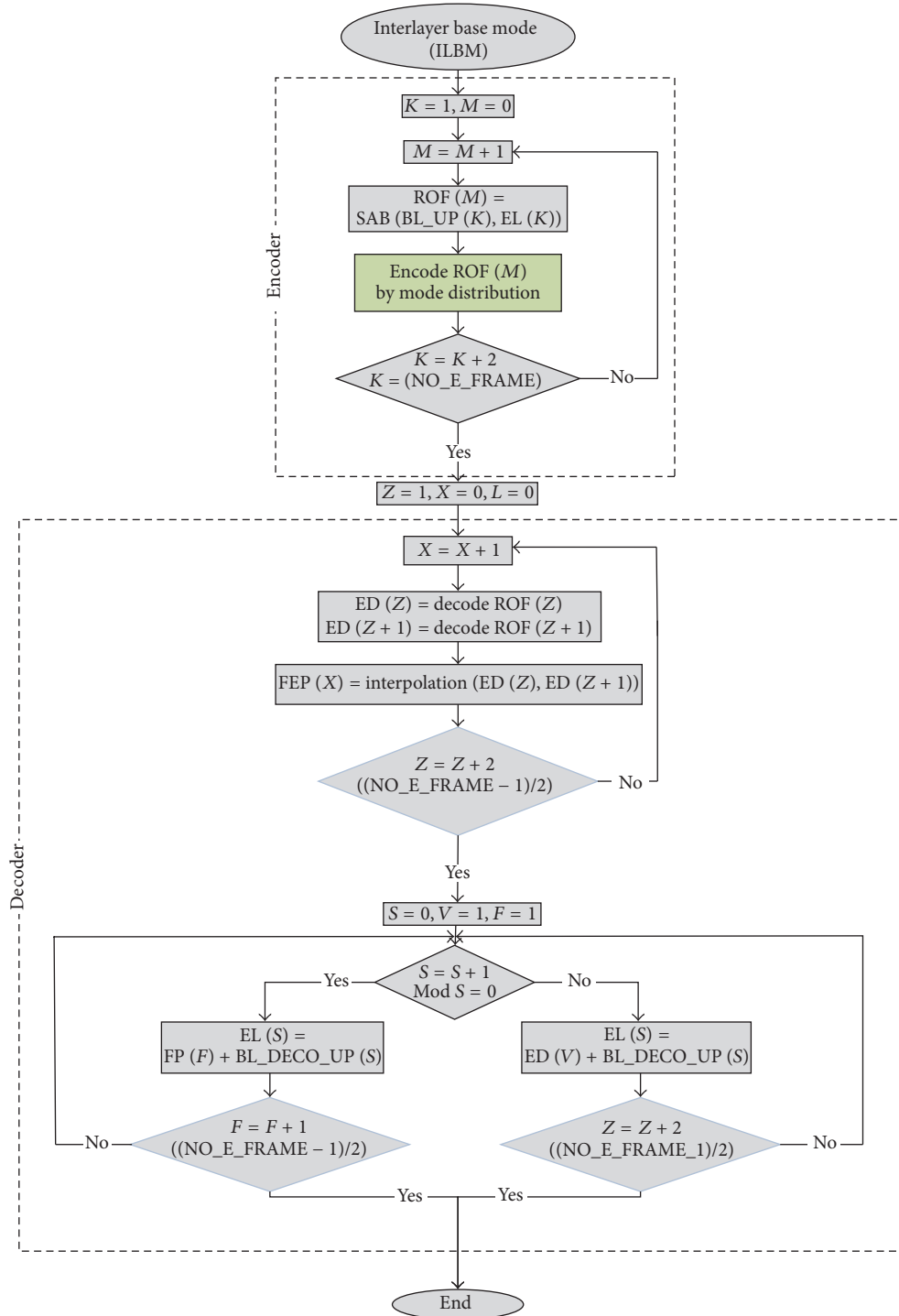


FIGURE 8: Interlayer base mode.

encoded with a quality higher than the THQ threshold, the SVC full-search mode should be employed. However, for a given video speed, depending on which between TT and ET has higher shortening priority, the two algorithms to consider are the full interpolation (FIP) and full-base mode (FBM) algorithms described in Section 3.2.

3.2. Proposed Algorithms for Improving Full Spatial SVC

3.2.1. Full Interpolation (FIP) Algorithm. As shown in Figure 9, the main difference here is that we do not skip all the even frames but send a part of them as will be clarified. At the encoder, this algorithm completely encodes the odd

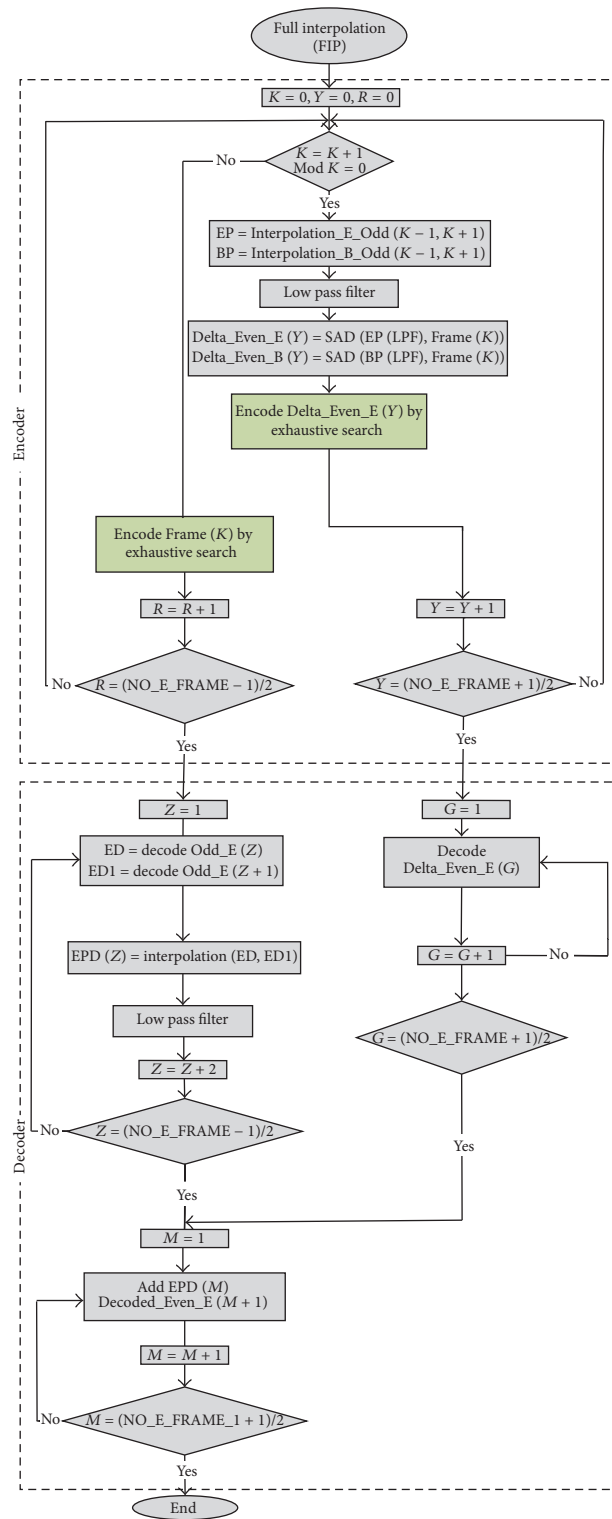


FIGURE 9: Full interpolation. EP = interpolation even EL frame at encoder. BP = interpolation even BL frame at encoder. Delta_Even_E (Y) = SAD between (EP and corresponding even frame in EL). Delta_Even_B (Y) = SAD between (BP and corresponding even frame in BL).

frames in the EL by exhaustive search. The even frames are obtained by interpolation. The interpolated frames are then subtracted from the targeted even frames and encoded using the SAD formula in (6). The difference is passed through a low-pass filter (7) and the output (Delta_Even_E) is encoded as the P frame. The foregoing procedure is also applied to the corresponding even frames in the BL, and the output (Delta_Even_B) is used as the reference frames for (Delta_Even_E) to encode the last as P frame with a high quality.

$$\text{SAD} = \sum_{ij} IC_{ij} - R_{ij}I, \quad (6)$$

where C_{ij} is a pixel of the current even frame, and R_{ij} is the corresponding pixel of the interpolation frame.

Assuming the frame to be of size $M \times N$ with a weighted averaging filter of size $m \times n$,

$$G(x, y) = \frac{\sum_{s=-a}^a \sum_{t=-b}^b w(s, t) f(X + s, y + t)}{\sum_{s=-a}^a w(s, t)}, \quad (7)$$

where $x = (1 : M - 1)$, $y = (1 : N - 1)$, $a = (m - 1)/2$, $b = (n - 1)/2$, w is the mask coefficient, and f denotes the frame.

All the data sent from the encoder are decoded at the decoder. The full odd frames for the EL layer are obtained by the previous step, as well as parts of all the even frames. The missing parts of the even frames are obtained by interpolation between every two successive odd decoding frames and successively added to the decoded parts.

3.2.2. Full-Base Mode (FBM) Algorithm. As shown in Figure 10, this algorithm is used to improve full SVC when the application requires more shortening of the ET than the TT. This is done using the same steps as the FIP algorithm, but with replacing the exhaustive search of the encoding macroblocks with the mode-distribution correlation.

3.3. Comparison of the Four Proposed Algorithms. As illustrated in Figure 11, the four proposed algorithms have three pivotal parameters, which are considered in discriminating among them for a specific purpose. The required level of quality is the core parameter. Accordingly, if the quality required is higher than the THQ, a full-search SVC should be used, and the choice would be between the FIP and FBM algorithms. However, if the required quality is lower than the THQ, an interlayer residual concept should be applied, in which case the choice would be between the ILIP and ILBM algorithms.

All the proposed algorithms shorten the TT and ET, but to varying extents. Hence, the second parameter considered in choosing among the algorithms is the percentage reductions of the TT and ET. The third parameter is the type of fast-motion video, which is determined by zero-mean normalized cross correlation (ZNCC), as discussed in Section 2.3. When using full search, if the video motion is slow, the FBM algorithm should be directly applied; otherwise, a trade-off between the TT and ET should be used to choose between the FBM and FIP algorithms.

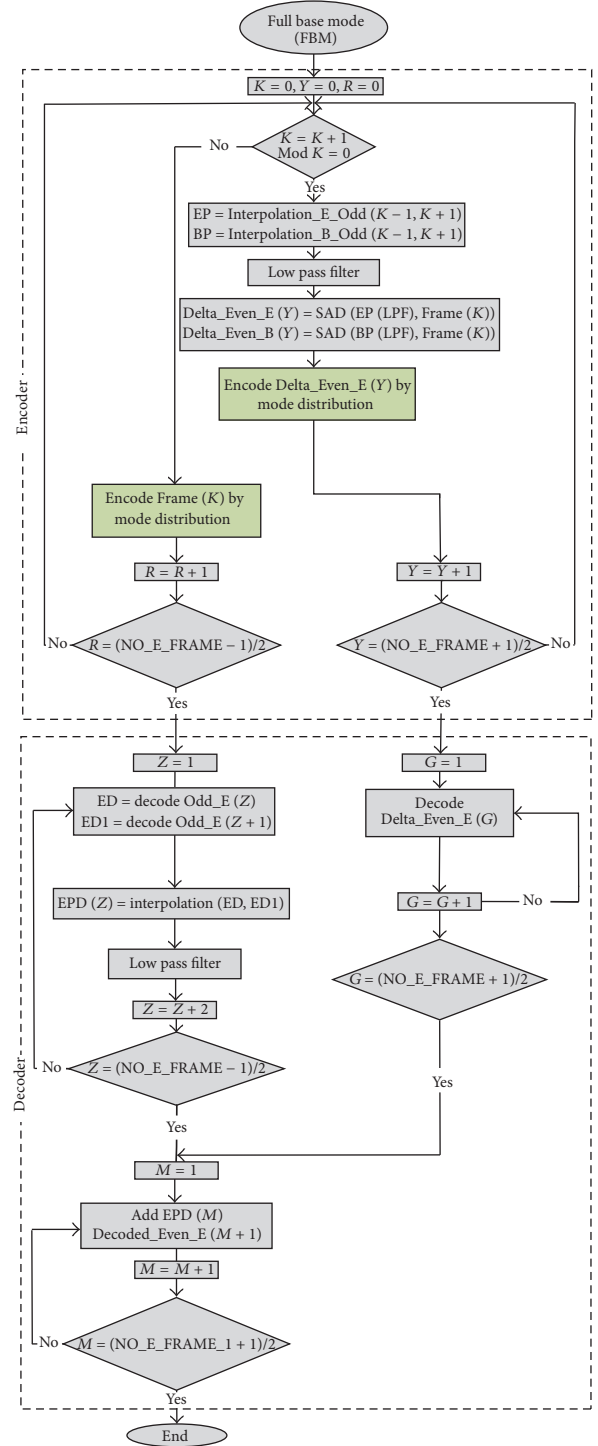


FIGURE 10: Full-base mode.

4. Experiment and Results

The employed test platform was Intel (R), with 3.40 GHz CPU, 8 GB RAM, in the Windows XP Professional operating system. The experimental results of the proposed algorithms were variously compared with those of the JSVM reference software, Tae algorithm [11], and Seon-2 algorithm [13]. All

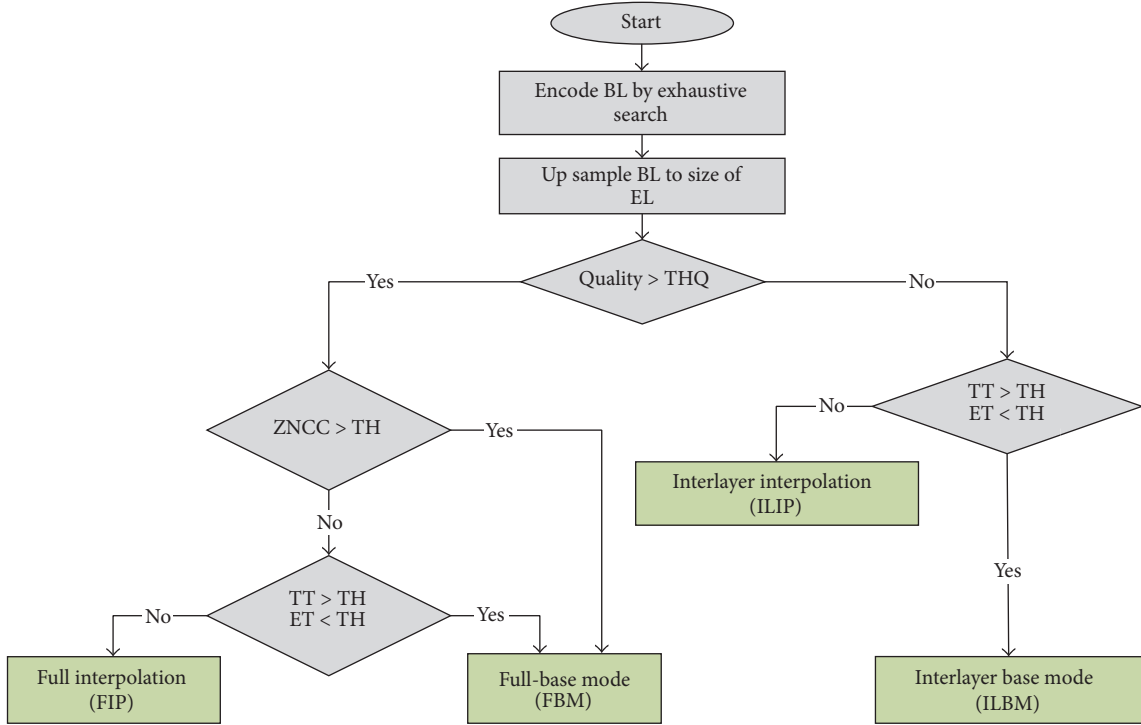


FIGURE 11: Main flowchart of algorithms.

TABLE 1: Experimental conditions.

	BL	EL
Resolution	SQCIF, QCIF	QCIF, CIF
Search type	Full search	Full search, interlayer residual search
GOP		16
Motion vector search range		32
Number of frames		99
Frame per second		30

the results were expressed as percentages relative to those of the reference software.

The performances of the proposed algorithms were evaluated using five standard test sequences, namely, “Forman” (slow motion), “City” and “Bus” (medium motion), and “Soccer” and “Football” (fast motion). Four parameters were considered in the evaluation, namely, saving in number of sending bits (Δ NSB), saving in ET (Δ ET), saving in TT (Δ TT), and degradation of peak signal-to-noise ratio (Y-PSNR) (see Table 1). The equations of the four parameters are presented as

$$\Delta\text{NSB} = \frac{\text{NSB (reference)} - \text{NSB (proposed)}}{\text{NSB (reference)}} * 100$$

$$\Delta\text{ET} = \frac{\text{ET (reference)} - \text{ET (proposed)}}{\text{ET (reference)}} * 100$$

$$\Delta\text{TT} = \frac{\text{TT (reference)} - \text{TT (proposed)}}{\text{TT (reference)}} * 100.$$

$$\text{Y-PSNR TT (reference) R}$$

$$= \text{PSNR (proposed)} - \text{PSNR (reference)}$$

(8)

Positive values of Δ ET, Δ TT, and Δ NSB indicate reductions of the encoding time, transmission time, and number of sending bits, respectively, while a positive value of Y-PSNR indicates an increase in quality.

Tables 2–7 present the results of the proposed algorithms for spatial scalability using different video resolutions such as the Sub Quarter Common Intermediate Format (SQCIF) and Quarter Common Intermediate Format (QCIF) in the BL, and the QCIF and Common Intermediate Format (CIF) in the EL. All the results are presented as percentages relative to those of the JSVM. The ILIP and ILBM algorithms are also compared with the Seon-2 algorithm and the FIP and FBM algorithms with the Tae algorithm.

The results presented in Tables 2–4 are for the ILIP and ILBM algorithms using the interlayer residual concept of SVC. The “Forman,” “Bus,” and “Football” sequences were considered. Tables 5–7 present the results for the FIP and FBM algorithms using the full mode concept of JSVC. In this case, the “Football,” “Forman,” and “City” sequences were considered.

4.1. Results of Proposed ILBM and ILIP Algorithms for Inter-layer Residual Spatial SVC. Table 2 presents the results of

TABLE 2: Results for foremen with interlayer residual concept.

	Foreman					
	BL = SQCIF & EL = QCIF		BL = QCIF & EL = CIF			
	ILIP	ILBM	ILIP	ILBM	ILLP versus Seon-2	ILBM versus Seon-2
Δ NSB% (Δ TT%)	64.30	53.44	63.25	52.81	63.316	52.876
Δ ET%	69.27	83.61	67.68	83.07	-15.39	12.55
Y-PSNR	-0.0010	-0.0023	-0.0013	-0.0031	0.011	0.018

TABLE 3: Results for bus with interlayer residual concept.

	Bus					
	BL = SQCIF & EL = QCIF		BL = QCIF & EL = CIF			
	ILIP	ILBM	ILIP	ILBM	ILLP versus Seon-2	ILBM versus Seon-2
Δ NSB% (Δ TT%)	57.89	51.6	46.97	45.09	47.787	45.907
Δ ET%	60.40	78.75	57.93	74.74	-8.1	8.71
Y-PSNR	-0.0013	-0.0027	-0.0016	0.038	0.019	0.136

TABLE 4: Results for football with interlayer residual concept.

	Football					
	BL = SQCIF & EL = QCIF		BL = QCIF & EL = CIF			
	ILIP	ILBM	ILIP	ILBM	ILLP versus Seon-2	ILBM versus Seon-2
Δ NSB% (Δ TT%)	45.42	40.75	44.74	40.24	44.97	40.47
Δ ET%	51.36	74.01	50.41	72.32	9.89	31.8
Y-PSNR	-0.003	-0.0038	-0.0036	-0.004	0.0374	0.037

TABLE 5: Results for foreman with full mode concept.

	Foreman					
	BL = SQCIF & EL = QCIF		BL = QCIF & EL = CIF			
	FIP	FBM	FIP	FBM	FIP versus Tae	FBM versus Tae
Δ NSB% (Δ TT%)	55.33	49.45	54.74	48.36	55.36	48.98
Δ ET%	52.83	76.85	50.35	73.54	-3.03	20.16
Y-PSNR	-0.010	-0.013	-0.012	-0.020	-0.012	-0.020

TABLE 6: Results for city with full mode concept.

	City					
	BL = SQCIF & EL = QCIF		BL = QCIF & EL = CIF			
	FIP	FBM	FIP	FBM	FIP versus Tae	FBM versus Tae
Δ NSB% (Δ TT%)	49.02	47.36	47.98	46.83	48.31	47.16
Δ ET%	47.56	73.45	46.87	72.67	1.99	27.79
Y-PSNR	-0.013	-0.021	-0.023	-0.024	-0.024	-0.024

TABLE 7: Results for soccer with full mode concept.

	Soccer					
	BL = SQCIF & EL = QCIF		BL = QCIF & EL = CIF			
	FIP	FBM	FIP	FBM	FIP versus Tae	FBM versus Tae
Δ NSB% (Δ TT%)	43.45	—	42.00	—	42.29	—
Δ ET%	48.98	—	48.79	—	0.37	—
Y-PSNR	-0.021	—	-0.027	—	0.027	—

the “Forman” sequence test for an SQCIF BL resolution and QCIF EL resolution. The ILIP algorithm significantly reduced the NSB, and hence the TT, compared to the JSVM. It specifically decreased the TT, ET, and PSNR by 64.30%, 69.27%, and 0.001 dB (negligible), respectively. The corresponding decreases for the ILBM algorithm were 53.44%, 83.61%, and 0.0023 dB, respectively. The ILBM algorithm produced the highest saving in ET.

For a QCIF BL resolution and CIF EL resolution, compared to the JSVM, the ILIP algorithm produced TT and ET savings of 63.25% and 67.68%, respectively, while decreasing the video quality by 0.0013. Compared to the Seon-2 algorithm, the ILIP algorithm shortened the TT and ET by 63.316% and 15.39%, respectively, while the video quality was increased by 0.011 dB. In the case of the ILBM algorithm, compared to the JSVM, the TT and ET were shortened by 52.81% and 83.07%, respectively, while the video quality was decreased by 0.0031. Compared to the Seon-2 algorithm, the TT and ET were shortened by 52.876% and 12.55%, respectively, while the PSNR was increased by 0.005 dB.

The results for the “BUS” sequence are presented in Table 3. For an SQCIF BL resolution and QCIF EL resolution. ILIP algorithm reduced the NSB, and hence the TT, compared to the JSVM. It specifically decreased the TT, ET, and PSNR by 57.89%, 60.40%, and 0.001 dB (negligible), respectively. The corresponding decreases for the ILBM algorithm were 51.6%, 78.75%, and 0.003 dB, respectively.

For a QCIF BL resolution and CIF EL resolution, compared to the JSVM, the ILIP algorithm shortened the TT and ET by 46.97% and 57.93%, respectively, while decreasing the video quality by 0.001 dB. Further, compared to the Seon-2 algorithm, the ILIP algorithm shortened the TT and ET by 47.78% and 8.1%, respectively, while increasing the video quality by 0.019 dB. In the case of the ILBM algorithm, relative to the JSVM, it shortened the TT and ET by 51.6% and 78.75%, respectively, while decreasing the video quality by 0.003 dB. Compared to the Seon-2 algorithm, the ILBM algorithm shortened the TT and ET by 45.907% and 8.71%, respectively, while increasing the PSNR by 0.136 dB.

The results for the “Football” sequence tests are presented in Table 4. For an SQCIF BL resolution and QCIF EL resolution, the ILIP algorithm reduced the NSB, and hence the TT, compared to the JSVM. It specifically decreased the TT, ET, and PSNR by 45.42%, 51.36%, and 0.003 dB (negligible), respectively. The corresponding decreases for the ILBM algorithm were 40.75%, 74.01%, and 0.004 dB (negligible), respectively.

In the case of a QCIF BL resolution and CIF EL resolution, compared to the JSVM, the ILIP algorithm shortened the TT and ET by 44.74% and 50.41%, respectively, while decreasing the video quality by 0.004 dB. Compared to the Seon-2 algorithm, the ILIP algorithm shortened the TT and ET by 44.79% and 9.8%, respectively, while increasing the video quality by 0.037 dB. In the case of the ILBM algorithm, compared to the JSVM, it shortened the TT and ET by 40.24% and 72.32%, respectively, while decreasing the video quality by 0.041 dB. Compared to the Seon-2 algorithm, the TT and ET were shortened by 40.47% and 31.8%, respectively, while the PSNR was increased by 0.037 dB.

From the above results for the proposed ILIP and ILBM algorithms, it can be seen that, for a slow-motion video, represented by the Foreman sequence, the ILIP algorithm is very efficient for reducing the NSB compared to the JSVM, implying that it shortens the TT. It also shortens the ET while negligibly decreasing the PSNR. The ILBM algorithm also reduces the NSB, shortens the TT, and produces the greatest decrease of the ET while negligibly decreasing the PSNR. The results of the ILIP and ILBM algorithms for a medium-motion video, represented by the Bus sequence, are similar to those for the Foreman sequence. However, the specific percentage changes produced by the two algorithms compared to the JSVM vary. The same applies to comparison with the Seon-2 algorithm.

For a fast-motion video, represented by the Football sequence, the performances of the ILIP and ILBM algorithms relative to the JSVM are similar to those for the Foreman and Bus sequences, although the percentage changes differ. However, compared to the Seon-2 algorithm, the ILIP and ILBM algorithms are superior with respect to all the test parameters. Overall, the proposed ILIP and ILBM algorithms afford good videos when very small details are not important, as well as the shortest TTs and ETs for broadcast applications.

4.2. Results of Proposed FIP and FBM Algorithms for Interlayer Residual Spatial SVC. The results for the Foreman sequence tests are presented in Table 5. For an SQCIF BL resolution and QCIF EL resolution, the FIP algorithm reduced the NSB, and hence the TT, compared to the JSVM. It specifically decreased the TT, ET, and PSNR by 55.33%, 52.83%, and 0.010 dB (negligible), respectively. The corresponding decreases for the FBM algorithm were 49.45%, 50.35%, and 0.013 dB, respectively.

For a QCIF BL resolution and CIF EL resolution, compared to the JSVM, the FIP algorithm decreased the TT and ET by 54.74% and 50.35%, respectively, while decreasing the video quality by 0.012 dB. Moreover, compared to the Tae algorithm, the FIP algorithm decreased the TT, ET, and video quality by 55.36%, 3.03%, and 0.012 dB, respectively. Compared to the JSVM, the FBM algorithm decreased the TT, ET, and video quality by 48.36%, 73.54%, and 0.020 dB, respectively. Compared to the Tae algorithm, it decreased the TT, ET, and PSNR by 48.98%, 20.61%, and 0.02 dB, respectively.

The results for the City sequence are presented in Table 6. For an SQCIF BL resolution and QCIF EL resolution, compared to the JSVM, the FIP algorithm decreased the NSB (TT), ET, and PSNR by 49.02%, 47.56%, and 0.013 dB (negligible), respectively. The corresponding decreases for the FBM algorithm were 47.36%, 73.45%, and 0.021 dB (negligible), respectively.

For a QCIF BL resolution and CIF EL resolution, compared to the JSVM, the FIP algorithm decreased the TT, ET, and video quality by 47.98%, 46.87%, and 0.023 dB, respectively. Compared to the Tae algorithm, the corresponding decreases were 48.31%, 1.99%, and 0.023 dB, respectively. In the case of the FBM algorithm, compared to the JSVM, it decreased the TT, ET, and video quality by 46.83%, 72.67%, and 0.024 dB, respectively. Compared to the Tae algorithm,

the FBM algorithm decreased the TT, ET, and PSNR by 47.16%, 27.79%, and 0.024 dB, respectively.

The results for the Soccer sequence tests are presented in Table 7. For an SQCIF BL resolution and QCIF EL resolution, the FIP algorithm decreased the NSB (TT), ET, and PSNR by 43.45%, 48.98%, and 0.021 dB (negligible), respectively. The FBM was not applicable to the Soccer sequence. In the case of a QCIF BL resolution and CIF EL resolution, the FIP algorithm decreased the TT, ET, and video quality by 42.00%, 48.79%, and 0.027 dB, respectively. Compared to the Tae algorithm, the corresponding decreases were 42.29%, 0.37%, and 0.027 dB, respectively.

The foregoing results of the proposed FIP and FBM algorithms indicate that their performances compared to the JSVM for the City sequence using different resolutions are similar to those for the Foreman sequence, although the percentage changes differ. Compared to the Tae algorithm, the FIP and FBM algorithms significantly shorten the TT while decreasing the video quality.

However, only the FIP algorithm, and not the FBM algorithm, is applicable to the Soccer sequence. Owing to its utilization of the mode-distribution correlation between the BL and EL, the FBM algorithm produces bad results when applied to a fast-motion video. Overall, the FIP and FBM algorithms are the best choice for applications in which small video details are important, such as in the medical field.

From the results for FIP and FBM proposed algorithms illustrated before the following is observed: For “city” stream compared with JSVM at different resolution, the same as explained in “Foreman” case will be valid and results with different percentage for two proposed algorithms FIP and FBM are obtained, where in the compression relative to Tae, the FIP and FBM algorithms save large transmission time with video quality decrement relative.

With “Soccer” video sequence, FIP algorithm is the only one applicable since FBM algorithm uses the concept of mode-distribution correlation between BL and EL that gives bad results with fast motion video. The FIP and FBM proposed algorithms are the best choice in applications that require very small details to appear in the video as medical applications.

5. Conclusion

New SVC algorithms with spatial scalability were developed with the objectives of shortening the TT and ET reducing the computational complexity of the SVC encoding, thereby shortening the ET. Three pivotal factors were considered in assessing the four developed algorithms, namely, the video quality, the correlation between the EL frames, and compromise between the TT and ET. The developed algorithms are of two categories. The first category includes the ILIP and ILBM algorithms, which are based on the concept of interlayer residual and are suitable for applications in which the required video quality is below the THQ. During tests, they shortened the ET and TT by up to 64.30% and 83.61%, respectively, compared to the JSVM, with negligible decrease in the PSNR. The ILIP algorithm enabled more TT saving than the LIBM algorithm, but less ET saving. The second

category of the proposed SVC algorithms includes the FIP and FBM algorithms, which are based on the concept of full search and are suitable for applications in which the required video quality is above the THQ. These algorithms shortened the ET and TT by up to 55.33% and 76.85%, respectively, compared to the JSVM, with negligible decrease in the PSNR. The FIP enabled more TT saving than the FBM, but less ET saving. The experimental observations confirm the effectiveness of the proposed algorithms for enhancing SVC by reducing the NSB, as indicated by the shortening of the TT.

Competing Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors greatly appreciate the support of Engineer Gamel Saleh, Chairman of Systems Design Company.

References

- [1] L. C. Wolf, C. Griwodz, and R. Steinmetz, “Multimedia Communication,” *Proceedings of the IEEE*, vol. 85, no. 12, pp. 1915–1933, 1997.
- [2] W. Kinsner, “Compression and its metrics for multimedia,” in *Proceedings of the 1st IEEE International Conference on Cognitive Informatics (ICCI '02)*, pp. 107–121, Calgary, Canada, August 2002.
- [3] J. Reichel, H. Schwarz, and M. Wien, “Joint scalable video model 2.0 reference encoding algorithm description,” ISO/IEC JTC1/SC29/WG11 N 7556, 2005.
- [4] L. Shen, Z. Liu, P. An, R. Ma, and Z. Zhang, “Fast mode decision for scalable video coding utilizing spatial and interlayer correlation,” *Journal of Electronic Imaging*, vol. 19, no. 3, Article ID 033010, 2010.
- [5] I. E. Richardson, *The H. 264 Advanced Video Compression Standard*, John Wiley & Sons, New York, NY, USA, 2011.
- [6] J.-R. Ohm, “Advances in scalable video coding,” *Proceedings of the IEEE*, vol. 93, no. 1, pp. 42–56, 2005.
- [7] K. De Wolf, D. De Schrijver, W. De Neve, S. De Zutter, P. Lambert, and R. Van de Walle, “Analysis of prediction mode decision in spatial enhancement layers in H. 264/AVC SVC,” in *Computer Analysis of Images and Patterns: 12th International Conference, CAIP 2007, Vienna, Austria, August 27–29, 2007. Proceedings*, vol. 4673 of *Lecture Notes in Computer Science*, pp. 848–855, Springer, Berlin, Germany, 2007.
- [8] T. Wiegand, G. Sullivan, J. Reichel, H. Schwarz, and M. Wien, “Joint draft 11 of SVC amendment,” Tech. Rep. JVT-X201, Joint Video Team, 2007.
- [9] T. Schierl, K. Grüneberg, and T. Wiegand, “Scalable video coding over RTP and MPEG-2 transport stream in broadcast and IPTV channels,” *IEEE Wireless Communications*, vol. 16, no. 5, pp. 64–71, 2009.
- [10] H. Schwarz and M. Wien, “The scalable video coding extension of the H.264/AVC standard [standards in a nutshell],” *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 135–141, 2008.
- [11] T.-J. Kim, J.-J. Yoo, J.-W. Hong, and J.-W. Suh, “Fast mode decision algorithm for scalable video coding based on luminance

- coded block pattern,” *Optical Engineering*, vol. 52, no. 1, pp. 17401–17401, 2013.
- [12] H. Li, Z. G. Li, C. Wen, and L.-P. Chau, “Fast mode decision for spatial scalable video coding,” in *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS '06)*, pp. 3005–3008, Island of Kos, Greece, May 2006.
- [13] S.-T. Kim, K. R. Konda, C.-S. Park, C.-S. Cho, and S.-J. Ko, “Fast mode decision algorithm for inter-layer coding in scalable video coding,” *IEEE Transactions on Consumer Electronics*, vol. 55, no. 3, pp. 1572–1580, 2009.
- [14] P.-C. Wang, G.-L. Li, S.-F. Huang, M.-J. Chen, and S.-C. Lin, “Efficient mode decision algorithm based on spatial, temporal, and inter-layer rate-distortion correlation coefficients for scalable video coding,” *ETRI Journal*, vol. 32, no. 4, pp. 577–587, 2010.
- [15] T.-J. Kim, B.-G. Kim, C.-S. Park, and K.-S. Jang, “Efficient block mode determination algorithm using adaptive search direction information for scalable video coding (SVC),” *Journal of Convergence*, vol. 5, no. 1, pp. 14–19, 2014.
- [16] Y. Dawei, B. Hou, and R. Sotudeh, “A fast mode decision algorithm in spatial and temporal scalable video coding,” in *Proceedings of the 3rd International Symposium on Intelligent Information Technology Application (IITA '09)*, vol. 1, pp. 716–719, 2009.
- [17] H. Schwarz, D. Marpe, and T. Wiegand, “Hierarchical B pictures,” Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG JVTP014, 2005.
- [18] A. Abdelazim, S. Mein, M. Varley, and D. Ait-Boudaoud, “Low complexity hierarchical prediction algorithm for H.264/SVC,” in *Proceedings of the 4th Pacific-Rim Symposium on Image and Video Technology (PSIVT '10)*, pp. 252–257, November 2010.
- [19] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, “Overview of the H.264/AVC video coding standard,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, 2003.
- [20] A. Abdelazim, M. Yang, and C. Grecos, “Fast subpixel motion estimation based on the interpolation effect on different block sizes for H264/AVC,” *Optical Engineering*, vol. 48, no. 3, Article ID 030504, 2009.
- [21] K. Briechele and U. D. Hanebeck, “Template matching using fast normalized cross correlation,” in *Proceedings of the SPIE's 1996 International Symposium on AeroSense: Aerospace/Defense Sensing, Simulations and Controls*, pp. 95–102, Orlando, Fla, USA, 2001.
- [22] V. Thirumalai and P. Frossard, “Joint reconstruction of multi-view compressed images,” *IEEE Transactions on Image Processing*, vol. 22, no. 5, pp. 1969–1981, 2013.
- [23] D. Sun, S. Roth, and M. J. Black, “Secrets of optical flow estimation and their principles,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '10)*, pp. 2432–2439, San Francisco, Calif, USA, June 2010.
- [24] P. K. Didyk, P. Sitthi-Amorn, W. Matusik, F. Durand, and W. T. Freeman, “Joint view expansion and filtering for automulti-scopic 3D displays,” Google Patents, 2014.
- [25] A. Davis, M. Rubinstein, N. Wadhwa, G. J. Mysore, F. Durand, and W. T. Freeman, “The visual microphone: passive recovery of sound from video,” *ACM Transactions on Graphics*, vol. 33, no. 4, article 79, 2014.
- [26] S. Meyer, O. Wang, H. Zimmer, M. Grosse, and A. Sorkine-Hornung, “Phase-based frame interpolation for video,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '15)*, pp. 1410–1418, IEEE, Boston, Mass, USA, June 2015.
- [27] J. Portilla and E. P. Simoncelli, “A parametric texture model based on joint statistics of complex wavelet coefficients,” *International Journal of Computer Vision*, vol. 40, no. 1, pp. 49–70, 2000.
- [28] H. Schwarz, D. Marpe, and T. Wiegand, “Analysis of hierarchical B pictures and MCTF,” in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '06)*, pp. 1929–1932, Ontario, Canada, July 2006.
- [29] T. Zhao, S. Kwong, H. Wang, and C.-C. J. Kuo, “H.264/SVC mode decision based on optimal stopping theory,” *IEEE Transactions on Image Processing*, vol. 21, no. 5, pp. 2607–2618, 2012.
- [30] L. Balaji and K. K. Thyagarajan, “H.264/SVC mode decision based on mode correlation and desired mode list,” *International Journal of Automation and Computing*, vol. 11, no. 5, pp. 510–516, 2014.



Hindawi

Submit your manuscripts at
<https://www.hindawi.com>

