Low Complexity for Scalable Video Coding Extension of H.264 based on the Complexity of Video

Mayada Khairy

Dept. of Computers and Systems,
Electronics Research ERI
Giza, Egypt

Alaa Hamdy

Dept. Telecomm. and Electronics
Faculty of Engineering, Helwan, Univ.
Cairo, Egypt

Amr Elsayed

Dept. Telecomm. and Electronics
Faculty of Engineering, Helwan, Univ.
Cairo, Egypt

Hesham Farouk Ali

Dept. of Computers and Systems,
Electronics Research ERI
Giza, Egypt

Abstract—**Scalable Video Coding (SVC) / H.264 is one type of video compression techniques. Which provided more reality in dealing with video compression to provide an efficient video coding based on H.264/AVC. This ensures higher performance through high compression ratio. SVC/H.264 is a complexity technique whereas the takes considerable time for computation the best mode of macroblock and motion estimation through using the exhaustive search techniques. This work reducing the processing time through matching between the complexity of the video and the method of selection macroblock and motion estimation. The goal of this approach is reducing the encoding time and improving the quality of video stream the efficiency of the proposed approach makes it suitable for are many applications as video conference application and security application.**

Keywords—Scalable video coding; motion estimation; SVC layers; quality scalability

I. INTRODUCTION

Scalable video coding (SVC) has been standardised to extend the capabilities of the H.264 advanced video coding (AVC) [1]. Whereas the objective of SVC is to enable the generation of the one-bit stream. In other words, it allows decoding partial streams depending on the specific rate, quality and resolution required by certain applications [2-3].

The SVC consists of a base layer (BL) and one or more enhancement layer (EL). The BL represents the main information of the video and should be transmitted with very high reliability [4]. On the other hand, the EL sent part of a bit stream according to the destination required [5]. SVC supports spatial, temporal and quality [6]. To encoded video using the SVC, divide frames to macroblocks (MBs). To encode this macroblock using two modes; first: intra-mode, which coded MB referring to the data in the same frame. The intra mode is utilised two types of MB there are (4×4 and 16×16) modes [7].
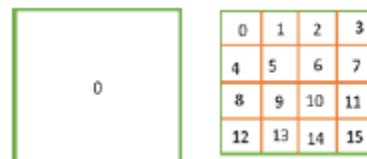

Fig. 1. Macroblock modes for intra prediction

Another one: Inter - mode; the MB coded referring data in the previously coded frames there are nine candidate mode of inter-modes (skip, 16×16, 16×8, 8×16, 8×8, 8×4, 4×8, and 4×4), Figure. 2 illustrates all the candidate inter-modes of H.264/SVC [8-9]. The process of the select best motion vector is called motion estimation as will be described in the next subsection.
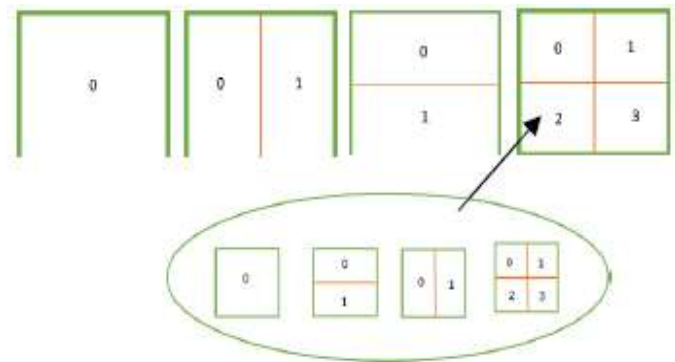

Fig. 2. Macroblock and sub-macroblock modes for inter prediction

A. Motion estimation in SVC

Motion estimation is one of the key elements of SVC technique. It is used to get benefit from the redundancy between frames in a video sequence, to investigate the high video compress rate of data.

The motion between frames estimates the prediction for the next frame. Once a motion estimate has been made, the algorithm only transmits the difference between two frames, which is contained in the motion information and the estimation error. The efficiency of the compression depends on the quality of motion estimation [10].

In this technique the picture is divided into two parts; the motion vectors which estimate the motion in the image, and the residual, which is the error between the current frame and its estimate [11]. There are two broad categories of motion compensation is block-based motion compensation and pixel-recursive motion compensation [12].

The Block Matching Algorithms (BMA) is a preferable method with SVC more computationally realistic than other methods [13-14]. The BMA frames are divided into non-overlapping blocks as discussed in the previous section, and each block is compared with its counterpart in the previous frame[15-16]. To find an area that is similar. The same area in the reference frame is known as the best match.
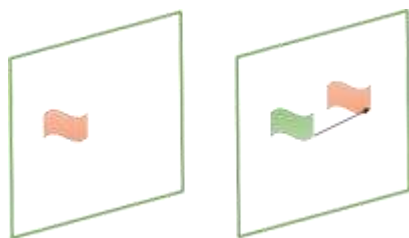

Fig. 3. Block matching algorithm

The relative difference in locations is the MV [17-19]. As illustrates in fig.3 the most common approach for block matching in ME is three steps search algorithm. It executes at three cascaded steps: first, is full-pixel ME here the ME start find the best match by the integer-pixel motion which performed for each square block of the frame to be encoded to the fine displacement vector(s) within a search range. To decide The best match, the Lagrangian cost function is used as follows [20]:

$$J = D + \lambda R \qquad (1)$$

Where λ is a Lagrangian multiplier, D is an error measure between the candidate MB taken from the reference frame(s) and the current MB, R stands for the number of bits required to encode the difference between the motion vector(s) and its prediction from the neighboring MBs (differential coding). The second step is half-pixel ME after the integer-pixel motion search finds the best match, the values at half-pixel positions around the best match are interpolate. At the last, the values of the quarter-pixel positions are generated by averaging pixels at integer and half-pixel positions. Figure 4 illustrates the interpolated fractional pixel positions. Upper-case letters indicate pixels on the full pixel grid, while numeric pixels indicate pixels at half-pixel positions and lowercase letters indicate pixels in between at quarter-pixel positions [21-22].

The SVC has many different MB modes, although the significantly improve the RD performance.

Several fast algorithms proceed to reduce the implementation complexity [23-25] most of them share the

same concept of using the correlation between macroblocks and it's neighbouring in different layers and also that MB in the bland its corresponding position in the EL. But these algorithms have a limitation when applied to the fast video sequence or with complexity background [26].
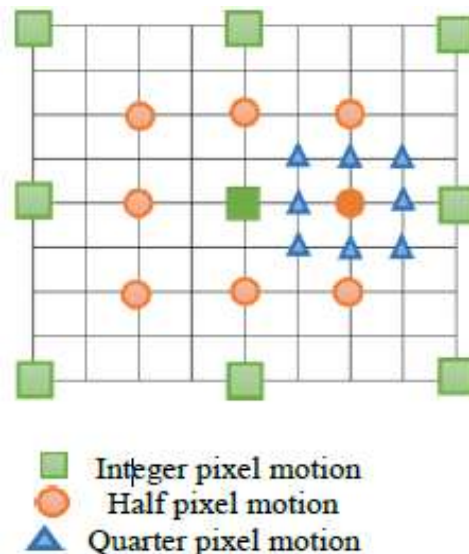

Fig. 4. Fractional pixel search positions

This paper focuses on decreasing the complexity of SVC encoding processing through reducing the time spent on the selection of both the mode of MB and motion estimation.

The proposed approach differs from the previous method in aspects: it linking the value of the difference between video sequences with the concept of correlation between MB in frames. To decide if we can utilise this concept or not and who it can be used to select the macroblocks mode and motion estimation.To solve the shortage of the previously proposed algorithms as will be discussed in the next section.
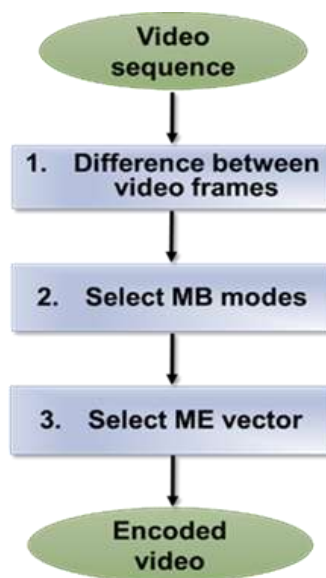
II. THE PROPOSED APPROACH

Fig. 5. The stages of proposed approach

This paper proposed a new approach to reducing encoding time for SVC. Through the minimizing time of; motion estimation (ME) and selection of macroblocks mode.

The proposed approach constants of three stages; first: Measuring the difference between video sequence test frames, second: select macroblocks modes and last stage: choose the method of motion estimation to calculate the motion vector. As outlined in fig. 5. Each of them will be discussed in details in the next subsections.

A. The first stage

The first stage in the proposed method as mention before is determination the difference between in the video sequence. This difference refers to the complexity and the motions of the video sequence.

There are many methods can detect the difference between frames. This work utilised sum of absolute differences (SAD) is used to determine the difference between frames. Seeing that the SAD has a higher quality precision and involves lower computational cost as. The SAD represent by the following equation:

$$SAD_n (P, Q) = \sum_{ij} | Pij - Qij| \qquad (2)$$

Where Pij is a pixel of the current frame, and Qij is the corresponding pixel next frame. After calculating SAD between every cascaded frame in a video sequence, determine the average (η) of SAD at the percentage of the difference between the frames according to the certain threshold as follow:

$$\eta = \frac{1}{n} * \sum_{i=1}^{n} SAD_i \qquad (3)$$

if η>th1 then the difference is high.
else if th1>η>th2 then the difference is medium.
else η<th2 then the difference is low

B. The second stage

Select the MB mode –outlined in figure 6- dependents on the value of η which calculated in the first stage as following:

For η>th1: meaning there is significant change between frames hence the selected MB mode should achieve high efficient. Hence, we cannot depend on the concept of correlation between macroblocks in different frames. In this case, the proposed approach reduce the complexity depends on two factors; out some of the modes from the competition from the start, and using the correlation equation to detriment the best MB instead of Lagrangian cost function.

The two concept applied as follows: According the smaller MB modes give highestCoding efficient, so the inter16*16 and intra 16*16 modes should be out from the competition.

- If the frame should be encoded is I frame selected intra 4*4 mode to encode it.

- Else if the frame is p frame test all candidate modes accept 16*16 inter mode.

The proposed approach select The MB modes, which give the best correlation between the current MB and candidate modes of MB using the following equation of correlation:

$$Correlation = \frac{\sum_{mn}(Wmn - W\acute{}\,)(Fmn - F\grave{})}{\sqrt{(\Sigma_m \Sigma_n(Wmn - W\grave{})^2)}\sqrt{(\Sigma_m \Sigma_n(Fmn - F\grave{})^2)}}, \qquad (4)$$

Where W, F are two MB, W`, F` are the average (μ) value of W, F

$$W\grave{} = \mu (w) = \frac{1}{K}\sum_{i=1}^{n} Wi \qquad (5)$$

$$F\grave{} = \mu (F) = \frac{1}{K}\sum_{i=1}^{n} Fi \qquad (6)$$

Where K is the total number of elements in the matrix and Wi are a number of elements in column.

For th2 > η < th1: Here there are medium changes between frames. The processes of select MB for frames in EL can be built on the tradeoff between the finer mode of corresponding MB in the BL and the mode of the previous encoded MB at the same frame.

For η<th2: Which meaning there is small change between frames. So, the MB mode depends on the tradeoff between the finer mode of the corresponding MB in the BL and the same mode of the previously encoded frame.

C. The third stage:

As discussed in Sec 1.1 the search of best motion vector execute at cascade three levels; full-pixel, half-pixel fraction and quarter pixel fraction

Which meaning the meager of encoding time consume in the motion estimation.

Our proposed approach reducing the time of motion estimation dependents on the two previous stage discussed in sec.2.1 and sec.2.3, as following:

For η>th1:   applied the smaller pixel fraction of the corresponding MB in the BL

For h2>η <th1: The ME dependents on which mode of the macroblock selected -as discussed in Sec. 2.2 as follows:

If the mode of macroblock selected depends on the corresponding MB in EL. Hence, select of motion estimation method depends on the same method of Pixel and half-pixel for ME in selected mode of MB in EL layer

For h2>η <th1: The ME dependents on which mode of the macroblock selected -as discussed 2.2 as follows:

If the mode of macroblock selected depends on the corresponding MB in EL. Hence, select of motion estimation method depends on

The tradeoff between the current pixel fraction and the smaller one meaning:

- If the ME method is full-pixel; then using tradeoff between full- Pixel and half-pixel for ME in selected mode of MB in EL layer

else If the ME in selected MB is half -pixel than using tradeoff between the half-pixel and quarter-pixel for ME in selected mode of MB in the enhancement layer

- Else selected using the using quarter pixel.

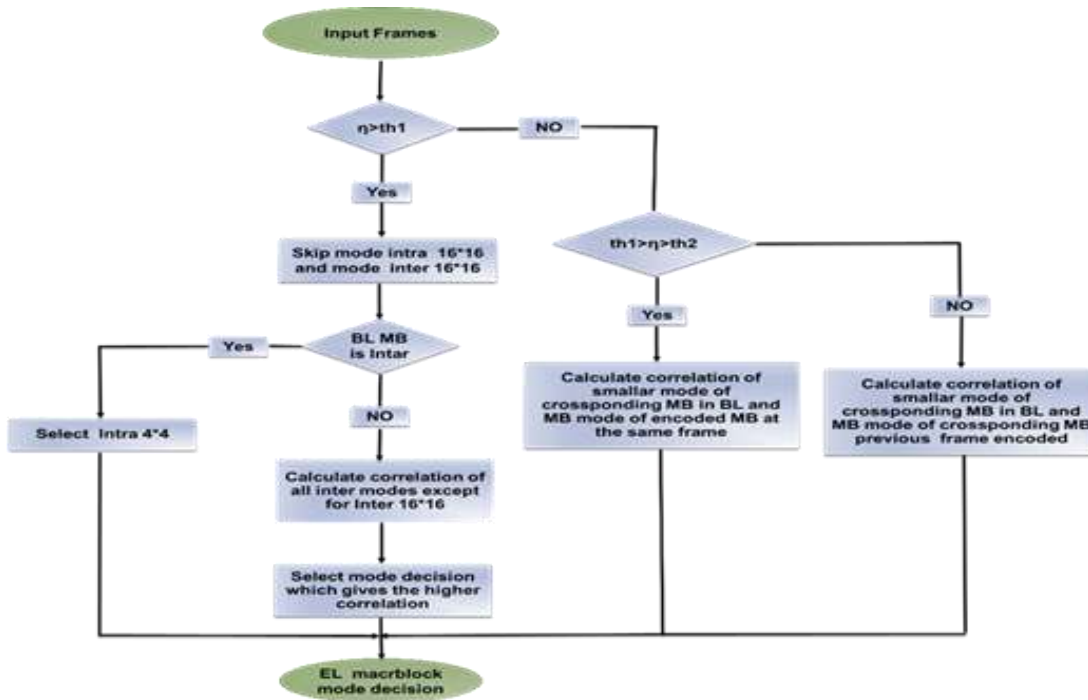For η<th2: meaning there is no significant change from frames to cascade so using the same
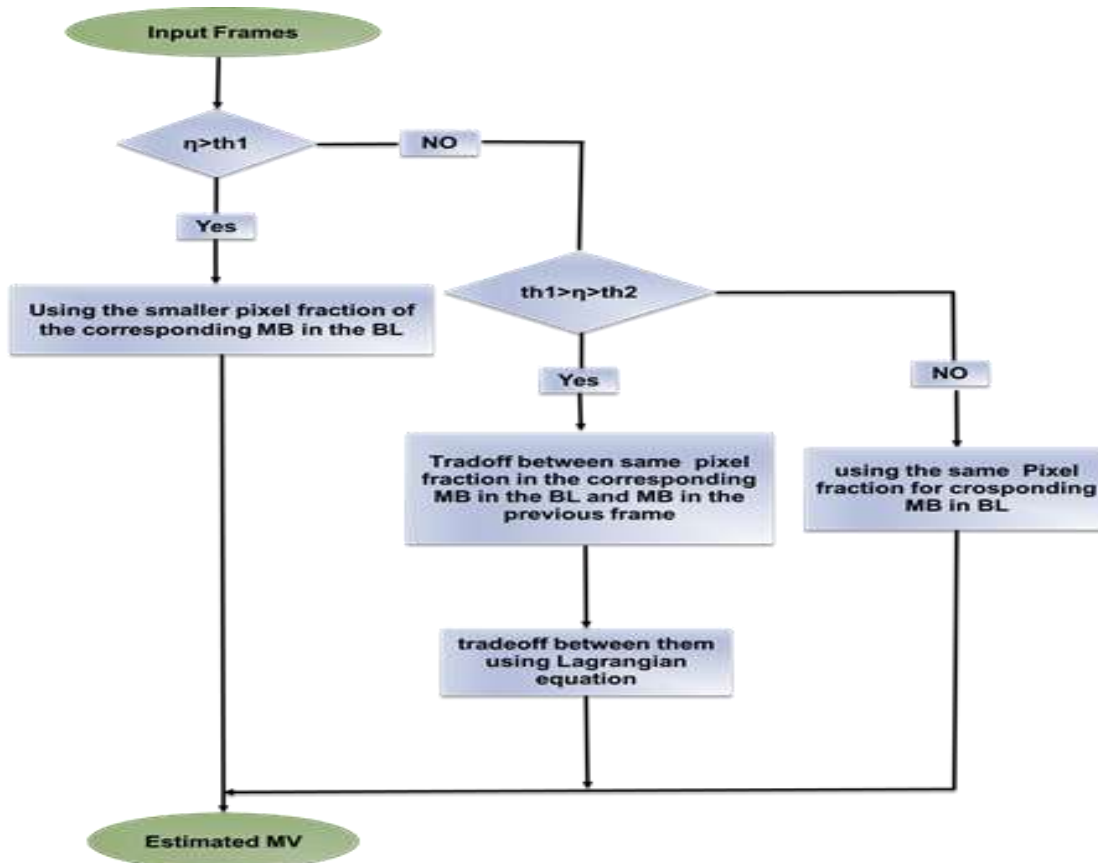
Fig. 6.   Select MB mode

Fig. 7.   Select MV

III. RESULTS AND EXPERIMENT

The proposed algorithm was implemented and evaluated using the "JSVM 9.18" software and Lu and Martin's method [27]. Six standard video test sequences Bus, City, Crew, Football, Foreman, Harbour.

Table 1 shows the experiment conditions. Were selected based on the commonly used values in similar research work

TABLE I.   EXPERIMENTAL CONDITIONS

	BL	EL
Quantization parameters	30	25
Resolution	QCIF	QCIF
GOP	16	
MV Search range|	32	
Number of frames	150	
Frame per second	30	

The metrics used for evaluation The video test are time saving (TS), (ΔTT), the number of sending bits (ΔNSB), and degradation of peak signal-to-noise ratio (Y-PSNR) representing as following equation:

$$\Delta TS = \frac{TS(proposed) - TS(reference)}{TS(reference)} * 100 \quad (7)$$

$$R = \frac{BR(proposed) - BR(reference)}{BR(reference)} * 100 \quad (8)$$

$$Y - PSNR = PSNR\ (proposed) - PSNR\ (reference) \quad (9)$$

All the results were expressed as percentages about reference software. The experimental results of the proposed approach were variously compared with those of the JSVM reference software, and with Lu and Martin's method.

The results demonstrate that the proposed approach achieved significant saving in time and improvement video quality with neglecting increasing in BR comparing with JSVM reference software, and with Lu and Martin's method. From the results in table 2 posted the following:

First for the time-saving parameter: Overall the proposed approach achieve good enhancement in decreasing the encoding time for all video sequence test especially with video represent the low complexity video as Foreman.

Second for the quality parameter: improvement the quality comparing with the Lu and Martin's method, however, there is small degradation in the quality relative to the original software JSVC 9.81.

The last for the Bit Rate: there are negligible increasing in BR comparing with JSVM reference software, and with Lu and Martin's method.

IV. CONCLUSION

This paper proposes an effective and efficient video compression approach that is suitable for multi user required the video with different quality. This approach is summarised as follows. Firstly, this measure difference between video sequence frames, second depends on the output from the first step to select macroblocks modes, last minimising the search method to choose the motion vector depending on the complexity and motion for video.

The efficiency of our approach compares it with reference software JSVC and another similar approach. It is shown that our approach has a high saving time up to (85.9 %) with improving quality and smaller decreasing in bit rate

TABLE II.   RESULTS OF THE PROPOSED APPROACH

Video Sequence Test	Proposed algorithm Vs. JSVM 9.18			Proposed algorithm Vs Lu and Martin's		
	ΔTS	ΔPSNR	ΔBR	ΔTS	ΔPSNR	ΔBR
Bus	70.45%	-0.05	1.81	8.20%	0.07	0.3
City	85.93%	-0.02	2.00	18.09%	0.02	0.24
Crew	72.18%	-0.01	2.01	11.6%	0.02	0.05
Football	61.07%	-0.03	1.41	3.28%	0.02	0.49
Foreman	81.47%	-0.08	1.94	18.47%	0.66	0.08
Harbour	75.77%	-0.04	1.01	7.58%	0.12	0.29

REFERENCE

[1] T. Wiegand, L. Noblet, and F. Rovati, "Scalable video coding for IPTV services," IEEE Transactions on Broadcasting, vol. 55, no. 2, pp. 527–538, 2009.

[2] C. Bouras, S. Charalambides, K. Stamos, S. Stroumpis, and G. Zaoudis, "Power management for SVC video over wireless networks," in Broadband and Wireless Computing, Communication and Applications (BWCCA), 2011 International Conference on, 2011, pp. 270–276.

[3] W.-H. Peng et al., "Rate-distortion optimised SVC bitstream extraction for heterogeneous devices: A preliminary investigation," in Multimedia Workshops, 2007. ISMW'07. Ninth IEEE International Symposium on, 2007, pp. 407–412.

[4] M. M. Hassan and U. Farooq, "Adaptive and ubiquitous video streaming over Wireless Mesh Networks," Journal of King Saud University-Computer and Information Sciences, 2016.

[5] T. Schierl, C. Hellge, S. Mirta, K. Gruneberg, and T. Wiegand, "Using H. 264/AVC-based scalable video coding (SVC) for real time streaming in wireless IP networks," IEEE International Symposium on Circuits and Systems, pp. 3455–3458,2007.

[6] T. Varisetty and P. Edara, "Systematic Overview of Savings versus Quality for H. 264/SVC," 2012.

[7] R. Ruolin, H. Ruimin, and C. Hao, "An adaptive selection algorithm of macroblock coding candidate modes," in Computer Science and Education (ICCSE), 2010 5th International Conference on, 2010, pp. 1499–1502.

[8] A. Kessentini, A. Samet, M. A. B. Ayed, and N. Masmoudi, "Performance analysis of inter-layer prediction module for H. 264/SVC," AEU-International Journal of Electronics and Communications, vol. 69, no. 1, pp. 344–350, 2015.

[9] X. Li, P. Amon, A. Hutter, and A. Kaup, "Performance analysis of inter-layer prediction in scalable video coding extension of H. 264/AVC," IEEE Transactions on broadcasting, vol. 57, no. 1, pp. 66–74, 2011.

[10] M. Usha, "Motion detection in compressed video using macroblock classification," Advanced Computing, vol. 5, no. 2/3, p. 1, 2014.

[11] J. Ratnottar, R. Joshi, and M. Shrivastav, "Comparative study of motion estimation & motion compensation for video compression," International Journal of Emerging Trends & Technology in Computer Science (IJETTCS) Vol, vol. 1, 2012.

[12] S. Mattoccia, F. Tombari, L. Di Stefano, and M. Pignoloni, "Efficient and optimal block matching for motion estimation," in Image Analysis and Processing, 2007. ICIAP 2007. 14th International Conference on, 2007, pp. 705–710.

[13] C. Kas and H. Nicolas, "H. 264/SVC scene motion analysis," in 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, 2009, pp. 957–960.

[14] S. Kamp, D. Heyden, and J.-R. Ohm, "Inter-temporal vector prediction for motion estimation in scalable video coding," in Intelligent Signal Processing and Communication Systems, 2007. ISPACS 2007. International Symposium on, 2007, pp. 586–589.

[15] S. Mattoccia, F. Tombari, L. Di Stefano, and M. Pignoloni, "Efficient and optimal block matching for motion estimation," in Image Analysis and Processing, 2007. ICIAP 2007. 14th International Conference on, 2007, pp. 705–710.

[16] X. Lu and G. R. Martin, "Fast H. 264/SVC inter-frame and inter-layer mode decisions based on motion activity," Electronics letters, vol. 48, no. 2, pp. 84–86, 2012.

[17] L. Kulkarni, T. M. Manu, and B. S. Anami, "A Two-Step Methodology for Minimization of Computational Overhead on Full Search Block Motion Estimation.," International Journal of u-and e-Service, Science and Technology, vol. 7, no. 4, pp. 339–348, 2014.

[18] X. Yi and N. Ling, "Scalable complexity-distortion model for fast motion estimation," in Visual Communications and Image Processing, p. 59603Y–59603Y, 2005.

[19] E. Cuevas, D. Zaldívar, M. Pérez-Cisneros, and D. Oliva, "Block-matching algorithm based on differential evolution for motion estimation," Engineering Applications of Artificial Intelligence, vol. 26, no. 1, pp. 488–498, 2013.

[20] K. R. Rao, D. N. Kim, and J. J. Hwang, "Video coding standards," The Netherlands: Springer, 2014.

[21] A. Luthra and P. N. Topiwala, "Overview of the H. 264/AVC video coding standard," in Optical Science and Technology, SPIE's 48th Annual Meeting, pp. 417–431,2003.

[22] T. Wedi and H. G. Musmann, "Motion-and aliasing-compensated prediction for hybrid video coding," IEEE transactions on circuits and systems for video technology, vol. 13, no. 7, pp. 577–586, 2003.

[23] H. Li, Z. G. Li, and C. Wen, "Fast mode decision algorithm for inter-frame coding in fully scalable video coding," IEEE Transactions on Circuits and Systems for Video Technology, vol. 16, no. 7, pp. 889–895, 2006.

[24] J. Ren and N. Kehtarnavaz, "Fast adaptive early termination for mode selection in H.264 scalable video coding," in 2008 15th IEEE International Conference on Image Processing, pp. 2464–2467,2008.

[25] G. Goh, J. Kang, M. Cho, and K. Chung, "Fast mode decision for scalable video coding based on neighboring macroblock analysis," in Proceedings of the 2009 ACM symposium on Applied Computing, pp. 1845–1846,2009.

[26] A. Abdelazim, S. Mein, M. Varley, and D. Ait-Boudaoud, "Low complexity hierarchical prediction algorithm for H. 264/SVC," in Image and Video Technology (PSIVT), 2010 Fourth Pacific-Rim Symposium on, pp. 252–25,72010.

[27] X. Lu and G. R. Martin, "Fast mode decision algorithm for the H. 264/AVC scalable video coding extension," IEEE Transactions on Circuits and Systems for Video Technology, vol. 23, no. 5, pp. 846–855, 2013.