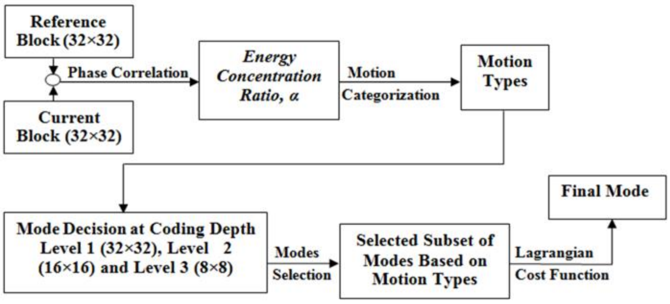
Graphical Abstract

A Novel Motion Classification Based Intermode Selection Strategy for HEVC Performance Improvement

Pallab Kanti Podder[1], Manoranjan Paul[1] and Manzur Murshed[2]
[1]*School of Computing and Mathematics, Charles Sturt University, Bathurst, NSW-2795, Australia.*
[2]*School of Information Technology, Federation University,VIC-3842, Australia.*



Neurocomputing

journal homepage: www.elsevier.com

A Novel Motion Classification Based Intermode Selection Strategy for HEVC Performance Improvement

Pallab Kanti Podder[1], Manoranjan Paul[1], and Manzur Murshed[2]

[1]*School of Computing and Mathematics, Charles Sturt University, Bathurst, NSW-2795, Australia*
[2]*School of Information Technology, Federation University, VIC-3842, Australia.*

ARTICLE INFO

Article history:
Received
Received in revised form
Accepted
Available online

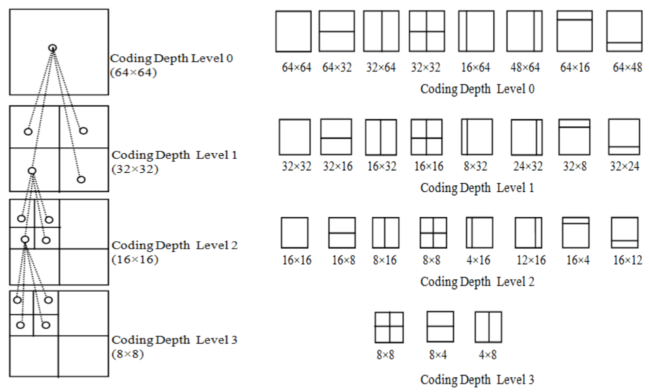
Keywords:
HEVC
Phase Correlation
Motion Identification
Motion Classification
Intermode Selection

ABSTRACT

High Efficiency Video Coding (HEVC) standard adopts several new approaches to achieve higher coding efficiency (approximately 50% bit-rate reduction) compared to its predecessor H.264/AVC with same perceptual image quality. Huge computational time has also increased due to the algorithmic complexity of HEVC compared to H.264/AVC. However, it is really a demanding task to reduce the encoding time while preserving the similar quality of the video sequences. In this paper, we propose a novel efficient intermode selection technique and incorporate into HEVC framework to predict motion estimation and motion compensation modes between current and reference blocks and perform faster inter mode selection based on three dissimilar motion types in divergent video sequences. Instead of exploring and traversing all the modes exhaustively, we merely select a subset of candidate modes and the final mode from the selected subset is determined based on their lowest Lagrangian cost function. The experimental results reveal that average encoding time can be downscaled by 40% with similar rate-distortion performance compared to the exhaustive mode selection strategy in HEVC.

1. Introduction

High Efficiency Video Coding (HEVC) is the state-of-the-art coding standard for video compression with some significant performance improvement [1]-[3]. Compared to its antecedent H.264/AVC, the aspects improved by HEVC are: (i) variable size *coding unit* (CU) such as 16×16, 32×32, and 64×64-pixels to accommodate different video resolutions; (ii) the symmetric and asymmetric adaptive block-partitioning phenomenon of HEVC results in distinct performance improvement especially in coding quality, and (iii) variable size *prediction unit* (PU) and *transform unit* (TU) to accommodate adaptive block-partitioning. In HEVC, the image is partitioned into different CUs and for the selection of *motion estimation* (ME) and *motion compensation* (MC) modes, a CU can be partitioned by 64×64 level down to 8×8-pixel level modes (block size with 64×64, 32×32, 16×16 and 8×8-pixels are denoted as depth level 0, 1, 2 and 3 respectively). Fig. 1 (a) illustrates the hierarchical structure of CU partitioning in HEVC standard and Fig. 1 (b) demonstrates this partitioning pattern at different coding depth levels using interprediction modes. Moreover, unlike H.264/AVC, HEVC provides



(a) Hierarchical Structure of HEVC block partitioning with different coding depth levels

(b) Symmetric and asymmetric block partitioning in HEVC at different coding depth levels for motion estimation using inter prediction modes

Fig. 1. Hierarchical structure and coding unit partitioning in HEVC at different coding depth levels.

asymmetric partitioning scheme such as 64×48, 64×16, 48×64, 16×64, 32×24, 32×8, 24×32, 8×32, 16×12, 16×4, 12×16 and 4×16-pixels as displayed in Fig. 1 (b). HEVC therefore achieves significantly improved compression efficiency at the cost of more than 4 times algorithmic complexity[4][5] compared to its predecessor H.264/AVC due to the extended number of coding depth levels, more complex CU partitioning ephemeron, residual

*Corresponding author
E-mail address: ppodder@csu.edu.au (P. Podder)

¥This work was supported in part by the Australian Research Council under Discovery Projects Grand DP130103670.

data compression using transformation and quantization and motion vectors transmission using entropy encoding. As a result, a number of electronic devices with limited processing capacity could not fully exploit HEVC encoding and decoding features.

In order to select a particular motion prediction mode, HEVC checks the Lagrangian cost function exhaustively using all modes in each coding depth level [6]. The Lagrangian cost function $€(m_k)$ for mode selection (m_k is the kth mode) is defined by-

$$€(m_k) = D(m_k) + \lambda \times R(m_k) \quad (1)$$

where λ be the Lagrangian multiplier, $D(m_k)$ be the distortion, and $R(m_k)$ be the resultant bit which are determined by a mode for each CU. ME and MC process in *HEVC model* (HM) is executed using all the possible depth levels and the best mode of any particular coding level is achieved by picking out the least cost ($€(m_k)$) using Lagrangian multiplier. To accomplish this task, HEVC necessitates at least 8 and at most 24 inter-prediction modes to select the best mode for any CU. Instead of traversing all the inter-prediction modes extensively, if we are able to fix a low computational criterion in order to determine the best subset of candidate modes we can significantly trim down the encoding time. In literature, to alleviate this complexity (in *rate distortion* (RD) optimized way), many researchers have contributed to several mode selection based fast approaches [7]-[12]. Shen *et al.* [13] introduce a TU size decision based early termination algorithm for HEVC encoders by using the Bayesian decision theory and the correlation between the variance of residual coefficients to reduce the number of candidate transform sizes for a given block. The experimental results confirm that their proposed algorithm is capable of saving 30–46% of transform processing complexity with some losses in coding efficiency.

Xiong *et al.* [14] propose a fast CU decision algorithm based on pyramid motion divergence (PMD) for HEVC inter prediction. They use a nearest neighboring algorithm to determine the splitting type of each CU by reducing the computational complexity of RDO. Their interprediction method speeds up the inter coding by sacrificing image quality. In order to terminate the exploring modes in lower level, Hou *et al*. [15] recommend a threshold based on the RD cost to explore mode in higher level. Their tested results affirm the time savings approximately 30% with 0.5% quality loss. Yang *et al.* [16] propose a fast intermode decision method for HEVC by checking whether the best prediction mode of their 2N×2N is skip mode or not by utilizing the correlated tendency of PU mode. This method reduces 39% computational time by sacrificing 0.8% bit-rate. Tan *et al.* [17] introduce an algorithm for HEVC standard where they investigate and compare a variety of algorithms for fast coding tree block and mode decision by introducing early partition decision and early CU termination approach. Experimental results evaluate the gain of 40% encoding time with the loss of 1.0% bit-rate. Correa *et al.* [18] prefer all possible modes for unconstrained frames and limited number of modes for constrained frames in order to control the computational time. From the experimental result their method sacrifices 5.7% bit rate while saving over 40% computing time compared to HEVC.

Apart from the above mentioned mode selection algorithms based on HEVC video coding standard (including [19]), there also exist other fast mode selection algorithms based on H.264 video coding standard [20]-[22]. Paul *et al*. [23] fully exploit the direct intermode selection process for H.264 video coding using phase correlation where they reduce the number of candidate modes based on motion prediction. This contribution would not be directly applied in HEVC because of extended number of coding depth levels, 3 times of modes and quadruple size of CU compared to H.264. Thus the encoding time reduction without degrading the RD performance in HEVC has become a real challenge today.

Podder *et al.* [24] used phase correlation to approximate the motion information between current and reference blocks by comparing with a number of predefined binary pattern templates. For each CU, they compare the generated binary matrix with their proposed predefined templates and then select the best-matched binary pattern template using a similarity metric to determine a subset of inter-modes. They introduce only two types of motion (tag '0' for no motion and '1' for both simple and complex motion) and based on the pattern of '0's and '1's combination in the templates they execute mode selection process in all coding depth levels (from 32×32 to 8×8 level). This motion categorization process suffers from RD performance for the high motion videos especially where the blocks have heterogeneous foreground and background areas. As their process cannot fully exploit the complex motion properly, it suffers from RD performance especially dealing with high motion videos and they sacrifice 0.24 dB PSNR on average for six videos. On the other hand, since the proposed method fully focuses on exploring single and multiple motions separately based on video contents, it can improve the RD performance (similar with HM as shown in Fig. **12** and Table V). Moreover, the process in [24], the generation of binary matrix, pattern matching with predefined templates and selection of best matched binary pattern template are projected to become more time consuming. However, the proposed method concentrates additionally on exploring video contents by finding both homogeneous and heterogeneous foreground and foreground/background regions and cares about categorizing single and multiple motions accordingly. Thus, the proposed method should improve the RD performance compared to [24] as well as HM with exhaustive mode selection scheme. Moreover, the proposed method should outperform the method in [24] in terms of computational time as it avoids a number of preprocessing steps such as binary matrix generation and pattern matching related overheads.

The modes selection algorithms in the literature for HEVC standard are based on the properties of residual, homogeneity and statistical correlation among different modes. Based on the abovementioned analysis and relationships the procedures in the existing literature merely depends on the Lagrangian cost function within HEVC framework. Therefore, those methods could not reach the similar RD performance with the implementations of HEVC. In contrast, the proposed scheme performs in two distinguishing phases where in the first phase we execute the motion categorization (consists with consecutive preprocessing stages- see Fig. 2) that is absolutely independent from Lagrangian cost function. Thus, based on the appropriate motion selection and categorization the proposed scheme provides the similar or improved rate-distortion performance compared to the exhaustive mode selection in HM. It is also expected that the types of selected modes in different bit rates are more stable compared to HM.

Therefore, our motivation is to trim down the computational time of HEVC by smart selection of appropriate ME and MC modes with the exploration of motion based on a number of features in the videos. For more appropriate decision of ME and MC modes, in this paper, we incorporate a new technique- HEVC-PC (HEVC with phase correlation) that approximates relative displacement information of the current block against the reference block [25][26] to predict the motion type of a CU. The motion based CUs are then encoded by the modes in the higher coding depth levels while for the CUs without motion are encoded by the modes in lower coding depth levels. In this paper, we explore three dissimilar categories of motion (no motion, simple/single motion and complex/multiple motions) based on the video contents and exploit them for the selection of a subset of candidate modes. The final mode from the selected subset is determined by calculating their lowest Lagrangian cost function. Since the proposed technique properly cares about the motion features based on different video contents, it is expected to achieve similar RD performance with HM8.0 [28]. Moreover, unlike the exhaustive mode selection approach in HM, as the proposed method execute motion categorization based mode selection with simple criteria, it can also significantly reduce the computational time. As a result, the proposed technique enables a number of electronic devices with limited processing and battery capacity to use HEVC encoding and decoding features and operate using low bandwidth oriented Internet.

The remainder of this paper is organized as follows. Section 2 explicitly describes the key steps of the proposed method; Experimental results and discussions are evaluated in section 3, while section 4 is the conclusions of the paper.

2. Proposed Technique

To calculate shifting information between two correlated images, we use the phase correlation technique and this technique accomplishes the task of shifting information calculation by *Fast Fourier Transformation* (FFT). We measure the respective change between current block and co-located blocks of different frames and by regulating this change we can produce the motion-compensated block in the reference frame [25]-[27]. The *Phase Matched Error* (PME) is obtained by subtracting the motion-compensated reference block from the current block. We then calculate the *energy concentration ratio* (ECR- the phase correlation extracted motion feature) of the low frequency component and the total energy of the transformed PME block. Based on the values of this ratio, the proposed algorithm more accurately executes motion categorization at different CUs. For example, if this ratio is greater than predefined threshold1 and threshold2, motion types stand for multiple motions and simple motion respectively, otherwise no existence of motion is assumed in that block and in all cases Th2<Th1. Based on these three different categories of motion, the proposed algorithm adopts the mode decision at 32×32, 16×16 and 8×8 coding depth levels in order to select a subset of candidate modes. Hence the proposed strategy is called the motion categorization based subset of mode selection. Now from the selected subset of modes the final mode (ultimate mode) is determined based on their lowest Lagrangian cost function. The whole process of the motion categorization based mode selection is shown as a block diagram in Fig. 2.

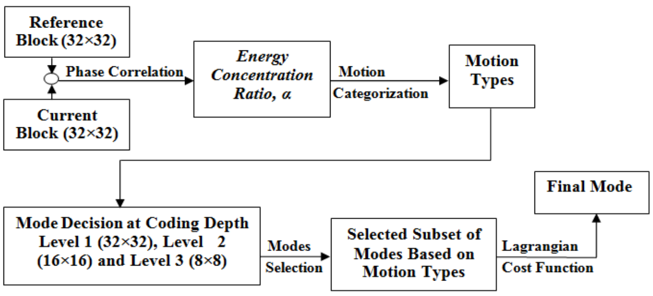


Fig. 2. Block diagram of the proposed motion categorization based mode selection strategy.

The following four successive key steps such as (i) Generation of Phase Matched Error (PME) (ii) Motion Extraction and Categorization (iii) Selection of Interprediction modes and (iv) Threshold selection for different bitrates explicitly describes the whole system.

2.1 Generation of Phase Matched Error

We set each CU as a maximum of 32×32 pixel block in the proposed method and also adopt this pixel block to categorize its motion type. In order to exploit whether any of the 32×32 pixel blocks encompasses with motion, we apply FFT both on the current block and its co-located block in the reference frame. We calculate the relative change between current block and its co-located block and by regulating this change we can produce the motion-compensated block in the reference frame. We then extract the phase of the current block and magnitude of the motion compensated block in the reference frame. We finally apply inverse FFT operation on both phase and magnitude in order to generate the matched block. The PME block which is a good index of motion identification is generated by the difference between original block and the matched block. The whole PME generation process is illustrated as a block diagram in Fig. 3. The rationality of generating PME block is to obtain exact motion information in a CU. In PME, if there is no displacement between current and co-located block, then the energy would be concentrated on the top-left triangle of the transformed PME, otherwise energy would be scattered through entire area. Thus, we calculate energy concentration ratio (i.e., α) of the top-left triangle with respect to the energy of the whole area and then finally predict the presence of motions against predefined threshold.

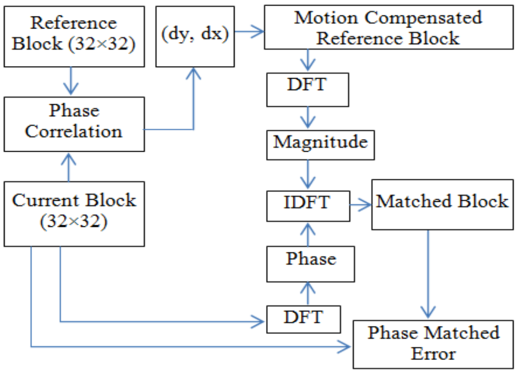


Fig. 3. Block diagram of phase matched error generation

2.2 Motion Extraction and Categorization

The phase correlation is calculated by applying the FFT and then inverse FFT of the current and reference blocks and finally applying the FFTSHIFT function as follows:

$$\Omega = fftshift\left|ifft\left(e^{j(\angle F_r - \angle F_c)}\right)\right| \tag{2}$$

where F_c and F_r be the Fast Fourier transformed blocks of the current C and reference R block respectively and $\angle$ is the phase of the corresponding transformed block. We evaluate the phase correlation peak (β) from the position of (dx+blocksize/2 + 1, dy+blocksize/2 + 1) as follows:

$$\beta = \Omega(dx + blocksize/2 + 1, dy + blocksize/2 + 1) \tag{3}$$

where the blocksize is 32 if 32×32-pixel block is used for phase correlation. Then the predicted motion vector (dx, dy) is determined by subtracting *blocksize*-1 from the (x, y) position of position of **Ω**, where we find the maximum value of **Ω.** In the matched block generation process, we use the phase of the current block and magnitude of the motion-compensated block in the reference frame and finally calculate the matched reference block (γ) for the current block by:

$$\gamma = \left|ifft\left(|F_r| e^{j(\angle F_c)}\right)\right| \tag{4}$$

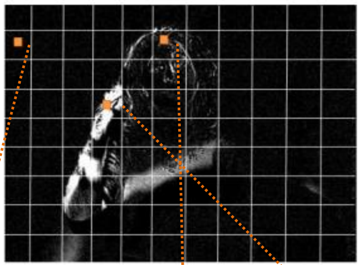
Now the displacement error (E) is enumerated by

$$E = C - \gamma \tag{5}$$

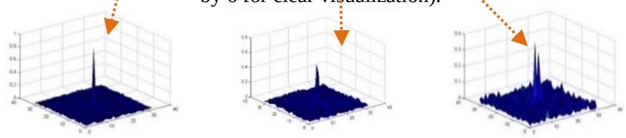
We then apply the *discrete cosine transform* (DCT) to error E in order to calculate the whole area energy of a particular block (D_{error_total}) as determined by:

$$D_{error\_total} = \Sigma(\Sigma(D_{error} \times D_{error})) \qquad (6)$$

Fig. 4 stands for the phase correlation peaks with proper motion at different blocks of the 13th frame on *Silent* video. The magnitudes of the motions illustrated in Fig. 4 (b) and Fig. 4 (c) correspond to no motion (motion type '0') and simple motion (motion type '1') respectively. These two types of motion are achieved from the CU at (2, 2) position and the CU at (2, 6) position respectively for the 13th frame on Silent video. Finally, from Fig. 4 (d), the phase correlation generated multiple (more than one) peaks that represent complex/ multiple motions (motion type '2') is obtained from the CU at (4, 4) position of the same frame on *Silent* video.



(a) Difference between 12th and 13th frame on *Silent* video (multiplied by 6 for clear visualization).



(b) No motion (c) Simple motion (d) Multiple motions

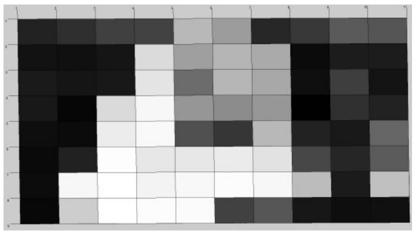
Fig. 4. Illustration of motion types generated at different blocks of 13th frame on *Silent* video; (b)-(d) are the phase shifted plots of no motion (0.4), simple motion (0.7) and multiple motions (0.8).

We then calculate the energy concentration ratio (α) of the low frequency component and the total energy of the transformed PME block, i.e., ratio from the top-left triangle energy with respect to the whole energy of a transformed CU as follows:

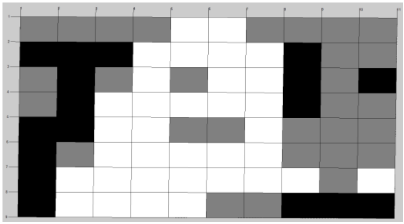
$$\alpha = (D_{error\_low} / D_{error\_total}) \qquad (7)$$

where D_{error_low} and D_{error_total} represent the top-left triangle energy and the whole area energy of a particular block. If this ratio is greater than the *Threshold*1 (Th1), the motion type is tagged by '2' (i.e., complex/ multiple motions) else if ratio is greater than *Threshold*2 (Th2), the motion type is tagged as '1' (i.e., simple/ single motion), otherwise the motion type is tagged as '0' (i.e., no motion) where in all the cases Th1>Th2.

The Energy concentration ratio between low-frequency coefficients (taken from top left triangle) and all coefficients of transformed PME blocks is shown in Fig. 5 (a). In addition, Fig. 5 (b) is generated by the proposed thresholding procedure to reflect different type motions. This figure (Fig. 5 (b)) is the absolute reflection of Fig. 5 (a) which indicates the motion representation map where dominating motion regions are marked with white colored blocks. If we compare Fig. 5 (a) and Fig. 4, we find the stability and uniformity in terms of the presence of respective motion categories (for example- for CUs at (2, 2), (2, 6) and (4, 4) positions in Fig. 4, we visualize black, ash, and while colored blocks respectively in Fig. 5 (a)). Thus, the proposed strategy properly categorize motions where multiple motions are denoted by white color, no motion regions are indicated with black and any other colored blocks corresponds to the simple motion in Fig. 5.



(a) Energy concentration ratio between low-frequency coefficients and all coefficients of transformed PME blocks where black and white blocks indicate no motion and complex motion respectively while blocks with any other colour corresponds to simple motion.



(b) Motion representation map indicating different types motions where blocks with black and white indicate no motion and complex motion respectively while blocks with any other colour corresponds to simple motion according to the proposed technique.

Fig. 5. Motion type identification by energy concentration ratio and its justification through motion representation map generated between 12th and 13th frames of *Silent* video.

2.3 Selection of Interprediction modes

Once the motions are categorized, we fully make use of these motion types for the subset of mode selection process. Table I illustrates the intermode selection process from the generated motion types at 32×32, 16×16 and 8×8 coding depth levels. From the selected subset of ME and MC modes, the final mode is determined using their lowest Lagrangian cost function.

Table I. Proposed selection of modes at 32×32, 16×16 & 8×8 coding depth levels using motion types.

Motion Types	**Selected subset of Modes**
No Motion (Motion Type 0)	Skip or 32×32
Single Motion (Motion Type 1)	Intra 16×16, Inter 32×32, 32×16, 16×32, 32×8, 32×24, 24×32 and 8×32
Multiple Motions (Motion Type 2)	Inter 16×16, 16×8, 8×16, 12×16, 4×16, 16×12, 16×4 and 8×8.

From Table I, this is obvious that if there is no existence of motion in any CUs, the proposed algorithm selects a subset of modes- Skip or 32×32. Once there is the presence of single motion (simple motion) in a CU the subset of 8 modes (Intra 16×16, Inter 32×32, 32×16, 16×32, 32×8, 32×24, 24×32 and 8×32) are explored from depth level 1 (32×32 level). , From the explored subset of ME and MC modes at 32×32 level, the final mode is selected by estimating their lowest Lagrangian cost function. The equation for the final mode (ξ) selection is given by-

$$\xi = \arg \min_{\forall m_k} (\epsilon(m_k)) \qquad (8)$$

where $\epsilon(m_k)$ be the Lagrangian cost function for mode selection (m_k is the *kth* mode). Table I encapsulates all the selected subset of modes at 32×32, 16×16 & 8×8 (depth level 1, 2 and 3 respectively according to Fig. 1) level and also exemplifies that individual subset of modes are guided by individual motion types. Similarly for multiple motions (i.e., motion type '2') in a CU, we explore only the 16×16 pixel level ME and MC modes (depth level 3) based on different video contents. In particular, when more motion dominating CUs are explored, the proposed algorithm selects modes with higher coding depth levels such as 16×16 or 8×8 to serve the purpose of motion categorization in a finer level.

2.4 Determination of Threshold (Th) for Different Bitrates

As the proposed method exploits both single and multiple motions by considering homogeneous/heterogeneous background and/or foreground motion block, we derive threshold values against whole range of *Quantization Parameters* (QPs) used in HEVC while sample threshold values against six popularly used QPs are mentioned in Table II. The values of Th1 and Th2 are mainly used to extract multiple motions and single motion respectively. We investigate different values of Th1 and Th2 to cover whole range of QPs for the proposed method and see rate-distortion performance compared to the HM 8.0 using a wide range of videos. Then we approximate the Th1 and Th2 by linear functions where QPs are used as only independent variables. According to the linear function we can easily approximate Th1 and Th2 as Th1= 0.005×QP+0.45 and Th2 = 0.005×QP+0.2 respectively. The Fig. 6 shows two Thresholds.

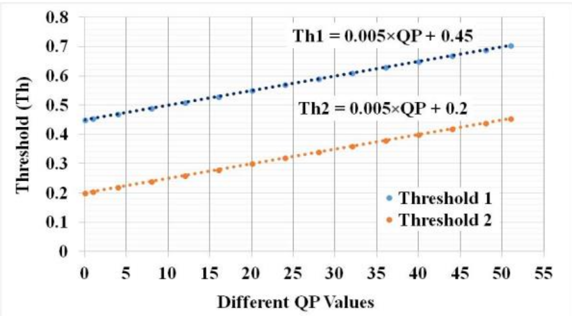


Fig 6. Proposed thresholds against different QPs for all sequences

Table II. Proposed thresholds for all videos in our experiment against distinct QPs.

QP	Th1	Th2
40	0.65	0.40
36	0.63	0.38
32	0.61	0.36
28	0.59	0.34
24	0.57	0.32
20	0.55	0.30

Paul *et al.* [23] propose the direct Intermode selection algorithm based on H.264 video coding standard where they use thresholds ranging from 0.27 to 0.91. Note that this range of thresholds could not perform well in our algorithm because of extended number of modes and complex CU partitioning paradigm in HEVC compared to H.264.We find the trend that when QP increases, both thresholds also increase. The reason is that, at high bit-rate, to serve the purpose of finer motion categorization, we use lower Th1 and Th2 values. Moreover, the number of motion blocks is inversely proportional to the threshold and at different bitrates different thresholds work best. From experiment, we observe that the proposed range of Th1 and Th2 values properly fit with all videos as they are exploited only for grasping and categorizing motions.

3. Experimental Results and Analysis

In this paper, to verify the performance of the proposed algorithm, the experimental results are presented with six *Standard Definition* (SD) videos- *Silent, Tennis, Paris, Bridgeclose, Tempete, waterfall,* four *High Definition* (HD) videos- *Bluesky, Pedestrian, Rushhour, and Parkrun,* and two *Multiview* (MV) videos- *Exit* and *Ballroom*. Each of the sequences are encoded with 25 frame rate and search length ±15 (for SD), ±31 (for HD and MV). Table III demonstrates the performance comparison results of seven recent and relevant fast mode selection algorithms where all the algorithms obtain significant computational time savings compared to different implementations of HEVC with exhaustive mode selection technique by increasing bit-rates and reducing PSNR. The Table confirms that the proposed technique improves the performance in terms of both reducing encoding time and improving rate-distortion performance compared to the existing state-of-the-art methods. The table also reveals that the proposed method outperforms HM8.0 by saving 40% computational time on average with insignificant bit rate increment (i.e., 0.08%) and negligible PSNR reduction (i.e., 0.01dB). The experimental results also confirm that, with the same experimental setup, the proposed method saves 52% computational time savings compared to HM8.0 that is 12% more compared to [24] for six videos used in [24].

Table III. Performance comparison of different fast mode selection algorithms compared to different implementations of HEVC.

Algorithms	Increased Bit-rate (%)	Reduced PSNR (dB)	Average Time Savings (%)	Remarks
Pan *et al.* [12], 2014	0.32	0.11	35	19 videos
Shen *et al.* [13], 2015	0.60	0.01	38	15 videos
Hou *et al.* [15], 2014	0.50	0.08	30	17 videos
Yang *et al.* [16], 2013	0.80	0.03	39	07 videos
Correa *et al.* [18], 2011	5.70	0.80	50	03 videos
Podder *et. al.* [24], 2014	2.56	0.24	40	06 videos
Proposed Method	0.08	0.01	40	12 videos including SD, HD and MV

3.1. Experimental Setup

In this paper, the experiments are conducted by a dedicated desktop machine (with Intel core i7 3770 CPU @ 3.4 GHz, 16 GB RAM and 1TB HDD) running 64 bit Windows operating system. We set each CU as a maximum of 32×32 pixel block in the proposed method and also adopt this pixel block to categorize its motion type. We use IPPP… format with the Group of Picture (GOP) 32 for both schemes and two reference frames. The sequences are encoded with 25 frame rate with search length ±15 and ±31. The proposed scheme and HEVC with exhaustive mode selection scheme are developed based on HEVC *test model* (HM) version 8.0 [28]. We compare the RD performance of HM and the proposed method considering both the symmetric and asymmetric partitioning block size of 32×32 to 8×8 levels for a wide range of bit-rates (i.e., using QP=20, 24, 28, 32, 36 and 40).

3.2. Mode Analysis of Coding Depth Levels

Table IV provides us a clear viewpoint illustrating the percentage of individual modes selected based on divergent QP values ranging from 20 to 40. Evidently at high bit-rates (QP=20), the number of motion dominating blocks selected are more than the number of motion dominating blocks at low bit-rates (QP=40). It is also obvious that both at high and low bit rates, compared to the exhaustive mode selection strategy in HM the proposed technique selects the consistent percentage of both higher and lower level modes at 32×32 or 16×16 coding depth level.

Table IV. Comparative study on HM and the proposed method comprising the selected percentage of individual modes based on QPs.

	Selected percentage (%) of modes																
QP	Skip	32×32	32×16	16×32	32×24	32×8	24×32	8×32	16×16 intra	16×16	16×8	8×16	12×16	4×16	16×12	16×4	8×8 or Smaller
40	1.67	51.52	6.57	5.56	5.84	6.36	5.67	4.23	3.83	0.94	2.90	0.62	0.76	0.46	1.18	0.93	0.91
36	2.21	43.94	7.28	6.23	6.83	7.40	6.50	4.88	4.11	0.68	3.79	0.86	1.07	0.54	1.49	0.86	1.41
32	3.28	34.11	7.38	6.64	8.10	9.10	7.70	5.60	4.63	0.49	4.87	1.19	1.52	0.69	2.06	0.42	2.34
28	3.53	28.70	7.08	6.61	8.60	9.68	7.99	5.86	4.70	0.60	5.01	1.65	2.16	0.80	2.83	0.31	3.94
24	3.87	25.46	6.48	6.16	8.49	10.0	8.01	5.61	6.35	0.81	4.83	1.86	2.62	0.78	3.36	0.14	5.20
20	4.62	26.25	6.12	5.71	7.48	8.82	7.10	4.84	8.39	1.14	3.82	1.94	2.91	0.81	3.70	0.06	6.28

(a) Percentage of individual modes selected by HM at different QPs.

	Selected percentage (%) of modes																
QP	Skip	32×32	32×16	16×32	32×24	32×8	24×32	8×32	16×16 intra	16×16	16×8	8×16	12×16	4×16	16×12	16×4	8×8 or Smaller
40	0.00	34.75	2.50	2.59	2.36	2.65	2.67	1.62	0.28	5.16	24.71	2.63	2.23	3.17	5.08	5.41	1.71
36	0.00	34.86	2.91	3.00	2.77	3.04	2.94	1.81	0.25	4.25	21.19	3.03	2.94	3.51	5.58	4.91	2.52
32	0.00	32.22	3.05	3.30	3.34	3.83	3.45	2.13	0.29	3.67	18.07	3.92	3.91	3.69	6.41	4.07	4.21
28	0.06	30.44	2.91	3.16	3.51	4.27	3.41	2.17	0.22	3.26	14.43	4.98	5.43	3.66	7.58	2.74	7.33
24	0.00	27.52	2.84	2.87	3.42	4.36	3.21	2.04	0.28	4.28	11.85	5.57	6.76	3.72	8.86	1.69	10.4
20	0.04	25.77	2.53	2.71	3.05	3.88	2.78	1.71	0.31	4.22	12.01	6.05	8.05	3.84	9.72	1.06	11.8

(b) Percentage of individual modes selected by the Proposed method at different QPs.

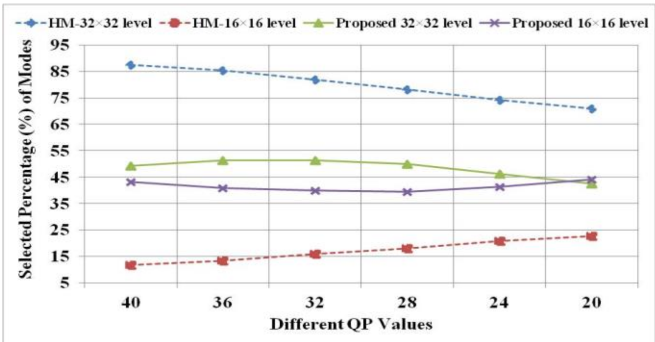
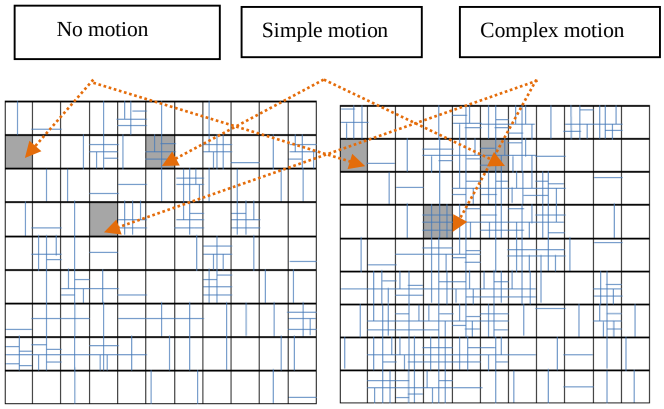


Fig. 7. Overall selected percentage of modes at 32×32 and 16×16 coding depth level by HM and the proposed method at different QPs.

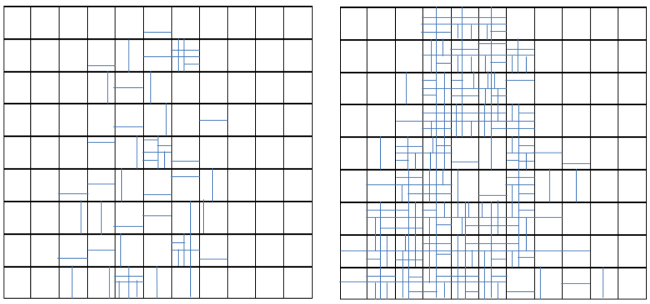
Fig. 7 is an identical reflection of Table IV which at a glance represents the overall selected percentage of modes at 32×32 and 16×16 coding depth levels by HM and the proposed method at different QPs. From Fig. 7, it is clear that at QP=40, HEVC selects around 87% and 13% of the 32×32 and 16×16 depth level modes respectively. Compared to HM at QP=40, the proposed method selects about 47% and 44% of depth level 32×32 and 16×16 modes. These values seem to be more stable compared to the exhaustive mode selection strategy in HM and the same trend is found down to QP=20. Thus, the overall selected percentage of modes by the proposed method at 32×32, 16×16 and 8×8 coding depth level for each QPs (from 40 to 20) seem to be more consistent. This strategy influences the concept of modes transcending [29]. Furthermore, the selection of coding depth level 2 or higher level modes (as stated earlier) is assumed to be an indication of high motion dominating region. Based on this strategy, if we compare the percentage of depth level 2 modes of the proposed method with HM, the proposed algorithm selects higher percentage of 16×16 depth level modes (as shown in Fig. 7) and selects the motion blocks more accurately in a finer level. From Fig. 7, at QP=32, HEVC selects 82% 32×32 level modes, which means, it almost cannot grasp any motion although motions are obvious in the videos. However, at the same QP, the proposed method selects 52% 32×32 level and 40% 16×16 level modes by being more sensitive to video contents, especially the motion features.

Moreover, for accurate and clearer visualization of mode selection if we compare Fig. 8 (a) and Fig. 8 (b), it is apparent that the proposed method selects more motion blocks compared to HM. However, although CU at (2, 2) position and CU at (2, 6) position are designated as no motion and simple motion blocks by both HM and the proposed method but the evidence of CU at (4, 4) position is different. HM selects the CU at (4, 4) position as no motion region whereas, the proposed method picks out the block as a high motion region and partition the CU more ideally for more accurate motion estimation and categorization. For further justification, if we compare this CU at (4, 4) position with Fig. 5 (a) and Fig. 5 (b), obviously that block is explicitly a motion dominating region and denote as white marked. The same trend is exhibited also at QP=36 shown in Fig. 9.



(a) Block partitioning by HM. (b) Block partitioning by the proposed method

Fig. 8. Block partitioning framework for the 13th frame of the *Silent* video at QP=24 with HM and the Proposed method.



(a) Block partitioning by HM. (b) Block partitioning by proposed method.

Fig. 9. Block partitioning framework for the 13th frame of the *Silent* video at QP=36 with HM and the proposed method.

The block partitioning phenomenon in the proposed method is therefore distinguishable and unlike HM with exhaustive mode selection strategy, the proposed technique does not fix any of the motion dominating blocks un-partitioned and this effect is influential for the improved RD performance for a wide range of bit-rates. As a consequence, compared to HM, we not only get the similar RD performance but also at some QPs we achieve better

RD performance (for Bluesky sequence of Fig. **12**) at QP=20 and also in other sequences of Table V at different bit-rates).

3.3. Comparison of RD Performance

For the performance justification of the proposed method, we first compare the RD performance against HM using six sequences (two SD, one MV and three HD) for a wide range of bit-rates as demonstrated in Fig. **12**. The experimental results exhibit that the proposed scheme retains the similar image quality (as HM) for a wide range of bit-rates in all video sequences especially by appropriate selection of motion dominated CUs and partitioning them with higher depth level modes. Table V reveals six additional divergent sequences (four SD, one MV and one HD) where PSNR performance of both techniques is compared at specific bit-rates [30].To produce the results in Table V, we first generate the RD performance curve (for instance the RD curves in Fig. **12** using different QPs) and from the curve we just determine a number of bit-rates and their corresponding PSNRs for both HM8.0 and the proposed method. Obviously for all video sequences, the PSNR values of the proposed method is very closely compared to the values with HEVC. Thus, from Fig. **12** and Table V, it is clear that the proposed mode selection strategy provides the similar RD performance with HM.

Table V. Additional results of HM and the proposed technique on six other different video sequences.

Name of the Sequences	Bit Rates (Kbps)	PSNR (dB)		Name of the Sequences	Bit Rates (Kbps)	PSNR (dB)	
		HM	Proposed			HM	Proposed
Paris	500	32.01	31.98	*Waterfall*	500	30.52	30.50
	1200	38.12	38.10		1500	36.01	35.99
	1900	41.20	41.20		2500	39.55	39.55
	2700	43.52	43.53		3500	43.02	43.04
Bridgeclose	300	30.51	30.52	*Pedestrian*	1000	33.65	33.50
	1200	37.12	37.10		3000	39.76	39.61
	2000	39.26	39.25		5000	42.72	42.52
	3000	41.75	41.73		7000	45.25	45.16
Tempete	2000	34.13	34.10	*Ballroom*	600	31.52	31.40
	3000	38.00	37.98		2000	36.83	36.68
	4000	40.80	40.80		4000	39.74	39.66
	4900	43.10	43.12		7000	42.95	43.02

3.4. Comparison of Computational Time

If any method exhaustively checks all the options in a level to select a particular option, theoretically it should necessitate more computational time. This complexity increases multiple times if any technique has to explore all modes in more depth levels to select a particular mode. Therefore, for all video types, we calculate the average number of modes selected per CU of both techniques and the results of Table VI show that HM checks more options in all cases and requires more encoding time. From Table VI, the overall average percentage of time savings by the proposed method is 44.05 and the reason behind this acquisition is the efficient subset of intermode selection with simple criteria (see Table I). However, we cannot ignore the pre-processing stages of the proposed method and by calculation we find that over twelve sequences on average 3.1% extra encoding time is required to execute phase correlation related overheads. Thus, theoretically we anticipate to acquire 41% computational time on average. Note that we only explicitly analyze encoding time but not the decoding time separately. However, the analysis of encoding time has also included decoding time as we need to perform decoding in the encoder. As the proposed method selects more smaller-block modes compared to HM8.0, we anticipate that decoding time might need extra time compared to HM8.0 decoding time.

Table VI. Percentage of time savings by the proposed method (against HM) for each individual sequence based on average no. of inter-modes selected per CU- a theoretical analysis.

Sequence types	Average no. of inter-modes selected per CU by HM	Average no. of inter-modes selected per CU by proposed method	Average percentage (%) of time savings
SD	16.89	8.07	52.22
HD	17.93	10.52	41.32
MV	19.58	12.02	38.61
Average percentage (%) of time savings			44.05

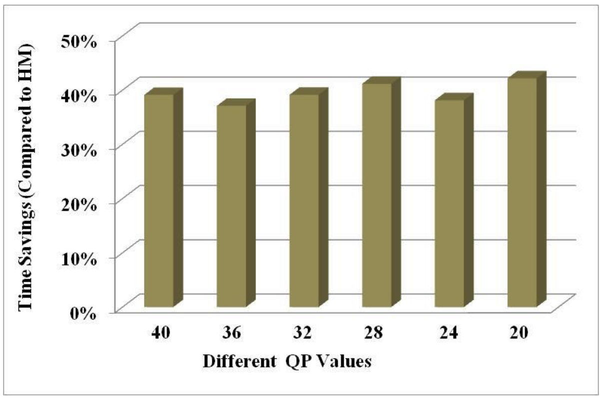


Fig. 10. Comparative Study on HM and the proposed method based on average percentage of time savings.

Fig. 10 provides the comparative Study on HM and the proposed method based on average percentage of time savings for a wide range of bit-rates. The theoretical computational time savings (41%) is therefore consistent with the experimental computational time savings (40%) for the proposed method against HM with similar RD performance. For comprehensive performance test, we execute the computational time analysis of both techniques based on video categories and find that the proposed method saves on average 39% encoding time compared to the exhaustive mode selection scheme in HM. This scenario is represented in Fig. 11. The figure also reveals that the proposed method saves more time for SD videos than MV or HD videos. It can be concluded that the proposed technique achieves significant computational time savings in terms of both QP and video category basis and also demonstrates similar rate-distortion performance with HM.

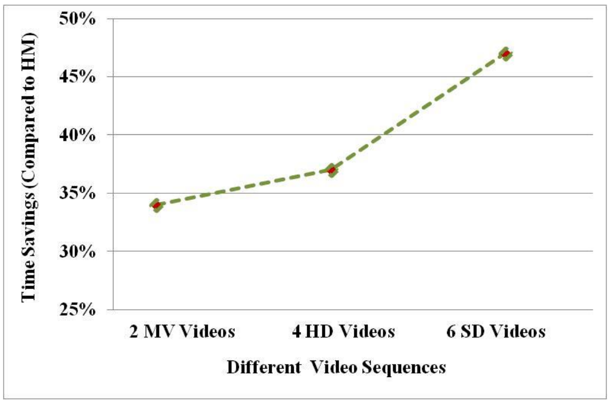


Fig. 11. Average time savings by the proposed method (against HM) based on different video categories.

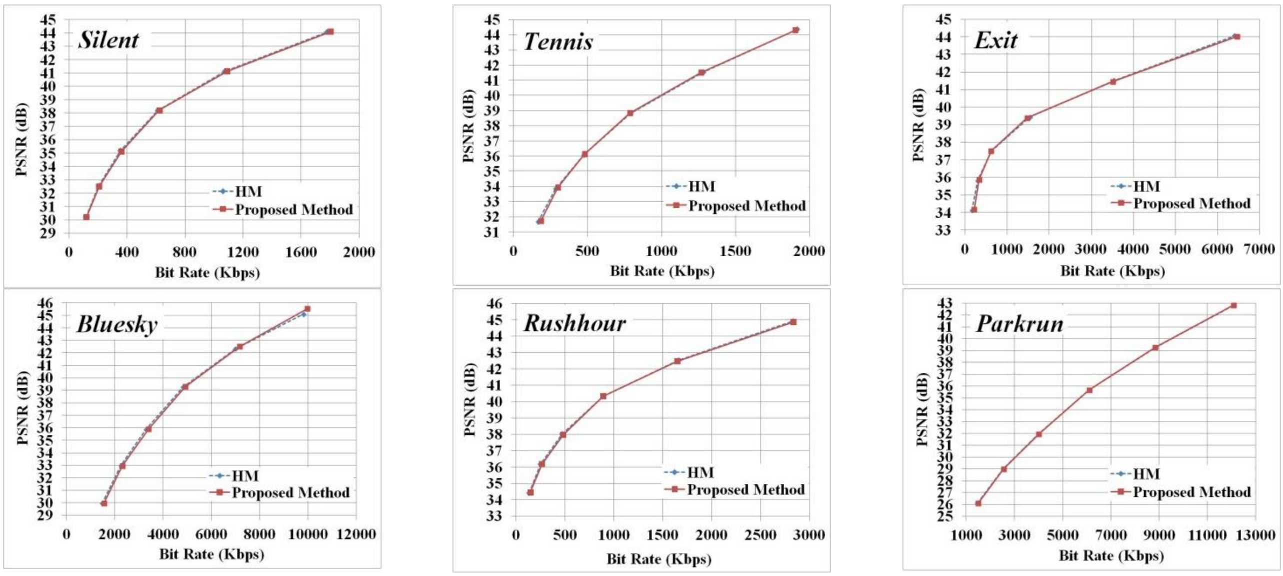


Fig. 12. Rate-distortion performance relationship using HM and the proposed method on six divergent video sequences.

4. Conclusion

In this paper, we incorporate a novel, fast and efficient intermode selection technique for HEVC video coding standard by categorizing different motion types based on phase correlation. The motion estimation and motion compensation modes are selected by the proposed method in a faster manner compared to HM by exploiting different categories of motion. The Lagrangian optimization criterion is set among the selected subset of modes to fix the final mode. Compared to the exhaustive mode selection approach in HM, the proposed coding strategy significantly reduces the computational time (on average 40%) while providing the similar perceptual rate distortion performance over a wide range of bitrates which is expected to become more suitable for all real time video coding applications especially for the electronic devices with limited processing and battery capacity.

References

[1] B. Bross, W. J. Han, G. J. Sullivan, J. R. Ohm and T. Wiegand, “High efficiency video coding text specification” Draft 9, document JCTVC-K1003, ITU-T/ISO/IEC, JCT-VC, October 2012.

[2] High Efficiency Video Coding, document ITU-T Rec. H.265 and ISO/IEC 23008-2 (HEVC), ITU-T and ISO/IEC, April 2013.

[3] G. J. Sullivan, J. R. Ohm, W.J. Han and T. Wiegand “Overview of the High Efficiency Video Coding (HEVC) standard ,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649-1668, December 2012.

[4] Y. Lu “Real-Time CPU based H.265/HEVC Encoding Solution with Intel Platform Technology,” Intel Corporation, Shanghai, PRC, December 2013.

[5] F. Bossen, B. Bross, K. Suhring and D. Flynn “HEVC Complexity and Implementation Analysis,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, pp. 1684-1695, December 2012.

[6] M. B. Cassa, M. Naccari and F. Pereira, “Fast Rate Distortion Optimization for the emerging HEVC standard,” *Picture Coding Symposium*, pp. 493-496, May 2012.

[7] J. Vanne, M. Viitanen and T. D. Hamalainen, “Efficient Mode Decision Schemes for HEVC Inter Prediction,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 9, pp. 1579-1593, September 2014.

[8] L. Shen, Z. Liu, X. Zhang, W. Zhao and Z. Zhang, “An Effective CU Size Decision Method for HEVC Encoders”, *IEEE Transactions on Multimedia*, vol. 15, pp. 465-470, February 2013.

[9] J. Leng, L. Sun, T. Ikenkga, and S. Sakaida, “Content Based Fast Coding Unit Decision Algorithm for HEVC,” *IEEE International Conference on Multimedia and Signal Processing*, pp. 56-59, 2011.

[10] A. Lee, D. S. J. Jun, J. Kim and J. Seok, “An Efficient Interprediction Mode Decision Method for Fast Motion Estimation in HEVC” *International Conference on ICT Convergence* (*ICTC*), pp. 502-505, October 2013.

[11] R.H. Gweon and Y. L. Lee, “Early Termination of CU Encoding to Reduce HEVC Complexity,” JCTVC-F045 July 2011.

[12] Z. Pan, S. Kwong, M. T. Sun and J. Lei, “Early MERGE Mode Decision Based on Motion Estimation and Hierarchical Depth Correlation for HEVC” *IEEE Transactions on Broadcasting*, vol. 60, issue 2, pp. 405-412, June 2014.

[13] L. Shen, Z. Zhang, X. Zhang, P. An, and Z. Liu, “Fast TU size decision algorithm for HEVC encoders using Bayesian theorem detection” *ELSEVIER journal on Signal Processing: Image Communication*, volume 32, pp. 121-128, March 2015.

[14] J. Xiong, H. Li, Q.Wu, and F. Meng, “A Fast HEVC Inter CU Selection Method Based on Pyramid Motion Divergence” *IEEE Transactions on Multimedia*, vol. 16, no. 2, pp. 559-564, February 2014.

[15] X. Hou and Y. Xue, “Fast Coding Unit Partitioning Algorithm for HEVC,” *IEEE International Conference on Consumer Electronics* (*ICCE*), pp. 7-10, 2014.

[16] S. Yang, H. Lee, H. J. Shim and B. Jeon, “Fast Intermode Decision Process for HEVC Encoder,” *IEEE International Workshop on* Image, Video, and Multidimensional Signal Processing (IVMSP), pp. 1-4, June 2013.

[17] H. L. Tan, F. Liu, Y.H. Tan, and C. Yeo, “On Fast Coding Tree Block and Mode Decision for High Efficiency Video Coding” *International Conference on Acoustic and Speech Signal Processing* (ICASSP), pp. 825 – 828, March 2012.

[18] G. Correa, P. Assuncao, L. Agostini, and L.A.D.S. Cruz, “Complexity Control of High Efficiency Video Encoders for Power- Constrained Devices”, *IEEE Trans. on Consumer Electronics* 57, November 2011.

[19] P. Podder, M. Paul and M. Murshed, “Efficient HEVC scheme using motion type categorization,” *ACM VideoNext: Design, Quality and Deployment of Adaptive Video Streaming in 10th International Conference on emerging Networking Experiments and Technologies* (*CoNEXT*), pp. 41-42, 2014.

[20] H. Huang, D. Hu, “An Efficient Fast Intermode Decision Algorithm for H.264/AVC,” *International conference on Communications and Mobile Computing*, pp. 161-165, 2009.

[21] H. Zeng, C. Cai and K. K. Ma, “Fast Mode Decision for H.264/AVC Based on Macroblock Motion Activity,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 4, pp. 491-499, April 2009.

[22] D. Kim, and J. Jeong, “A Fast Mode Selection Algorithm in H.264 Video Coding”, Proc. *IEEE Int. Conf. Multimedia Expo*, pp. 1709-1712 (2006).

[23] M. Paul, W. Lin, C. T. Lau, and B. S. Lee, “Direct Inter-Mode Selection for H.264 Video Coding using Phase Correlation”, *IEEE Transactions on Image Processing*, vol. 20, no. 2, pp. 461–473, 2011.

[24] P. Podder, M. Paul, M. Murshed and S. Chakrabarty, "Fast Intermode Selection for HEVC Video Coding using Phase Correlation," *IEEE International Conference on Digital Image Computing: Techniques and Applications* (*DICTA*)- 2014.

[25] M. Paul, M. Frater, and J. Arnold, "An Efficient Mode Selection Prior to the Actual Encoding for H.264/AVC Encoder," *IEEE Transactions on Multimedia*, vol. 11, no. 4, pp. 581-588, April 2009.

[26] V. Argyriou, and T. Vlachos, "A Study of Sub-pixel Motion Estimation using Phase correlation", *International Conference on Brit. Mach. Vis. Assoc.*, pp. 387–396, 2006.

[27] G. A. Thomas, "Television Motion Measurement for DATV and other Applications," *BBC research department*, U.K., November 1987.

[28] Joint Collaborative Team on Video Coding (JCT-VC), HM Software Manual, CVS server at:(http://hevc.kw.bbc.co.uk/svn/jctvc-hm/), 2013.

[29] E. Peixoto, B. Macchiavello, and E. M. Hung, "Fast H.264/AVC to HEVC Transcoding Based on machine Learning" *IEEE International Telecommunications Symposium* (*ITS*), PP. 1-4, August 2014.

[30] G. Bjontegaard, "Calculation of average PSNR differences between RD curves," ITU-T SC16/Q6, VCEG-M33, Austin, USA, April 2001.

Bibliography

Pallab Podder has completed his B.Sc (Hons) and Masters (M.Sc.) Degree from the Department of Information & Communication Engineering of Islamic University, Kushtia, Bangladesh, in the year 2008 and 2010 respectively. After accomplishing his M.Sc. degree he joined as a lecturer in the Computer Science & Engineering Department of Bangladesh University, Dhaka, Bangladesh. Then he joined as a lecturer in the Department of Information & Communication Engineering of Pabna Science & Technology University, Pabna, one of the public universities in Bangladesh. Currently he is working as a PhD student and teaching staff in Charles Sturt University, Australia. Pallab has received "Faculty of Business Compact Scholarship". He has published several journal and conference papers in the area of image processing and video coding.

Manoranjan Paul received B.Sc.Eng. (hons.) Degree in Computer Science and Engineering from Bangladesh University of Engineering and Technology (BUET), Bangladesh, in 1997 and PhD Degree from Monash University, Australia in 2005. He was an Assistant Professor in Ahsanullah University of Science and Technology. He was a Post-Doctoral Research Fellow in the University of New South Wales, in 2005–2006, Monash University, in 2006–2009, and Nanyang Technological University, in 2009–2011. He has joined in the School of Computing and Mathematics, Charles Sturt University (CSU) at 2011. Currently he is a Senior Lecturer and Associate Director of the Centre for Research in Complex Systems (CRiCS) in CSU.

Dr Paul is a Senior Member of the IEEE and ACS. He is in editorial board of three international journals including EURASIP Journal on Advances in Signal Processing (JASP). Dr. Paul has served as a guest editor of Journal of Multimedia and Journal of Computers for five special issues. He has been in Technical Program Committees of more than 30 international conferences. Dr Paul obtained more than A$1M competitive grant money including Australian Research Council (ARC) Discovery Project. He has published more than 80 refereed publications including journals and conferences. He organized a special session on "New video coding technologies" in IEEE ISCAS 2010. He was a keynote speaker in the IEEE ICCIT 2010 and Video Everywhere workshop in IEEE WoWMoM 2014, and tutorial speaker on "Multiview Video Coding" in DICTA 2013. His major research interests are in the fields of video coding, computer vision, and EEG Signals analysis. Dr Paul received Vice Chancellor Research Excellence Award in Faculty level at 2013.

Manzur Murshed received the B.Sc.Engg. (Hons) degree in computer science and engineering from Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh, in 1994 and the PhD degree in computer science from the Australian National University (ANU), Canberra, Australia, in 1999. He also completed his Postgraduate Certificate in Graduate Teaching from ANU in 1997. He is Federation University Australia's Robert HT Smith Professor and Personal Chair (formerly Monash University, Gippsland Campus) and was one of the founding directors of the Centre for Multimedia Computing, Communications, and Artificial Intelligence Research (MCCAIR). His major research interests are in the fields of video technology, information theory, wireless communications, distributed computing, and security & privacy.

He has so far published 175 refereed research papers and received more than $1 million nationally competitive research funding, including three Australian Research Council Discovery Projects grants in 2006, 2010, and 2013 on video coding and communications, and a large industry grant in 2011 on secured video conferencing. To date he has successfully supervised 19 PhD students. He is an Editor of International Journal of Digital Multimedia Broadcasting and has served as an Associate Editor of IEEE Transactions on Circuits and Systems for Video Technology in 2012 and as a guest editor of special issues of Journal of Multimedia in 2009-2012. He received a University Gold Medal from BUET in 1994, the inaugural Early Career Research Excellence award from the Faculty of Information Technology, Monash University in 2006, and the Vice-Chancellor's Knowledge Transfer Award (commendation) from the University of Melbourne in 2007. He is a Senior Member of IEEE as well as the Emeritus professor in the School of Engineering and Information Technology under the Faculty of Science and Technology in the Federation University Australia.