Efficient HEVC Scheme Using Motion Type Categorization

Pallab Kanti Podder
School of Computing and Mathematics,
Charles Sturt University
Bathurst, NSW-2795, Australia
ppodder@csu.edu.au

Manoranjan Paul
School of Computing and Mathematics,
Charles Sturt University
Bathurst, NSW-2795, Australia
mpaul@csu.edu.au

Manzur Murshed
School of Information Technology,
Federation University
VIC-3841, Australia
manzur.murshed@federation.edu.au

ABSTRACT

High Efficiency Video Coding (HEVC) standard introduces a number of innovative tools which can reduce approximately 50% bit-rate compared to its predecessor H.264/AVC at the same perceptual video quality whereas the computational time has increased multiple times. To reduce the encoding time while preserving the expected video quality has become a real challenge today for video transmission and streaming especially using low-powered devices. *Motion estimation* (ME) and *motion compensation* (MC) using variable-size blocks (i.e., intermodes) require 60-80% of total computational time. In this paper we propose a new efficient intermode selection technique based on phase correlation and incorporate into HEVC framework to predict ME and MC modes and perform faster intermode selection based on three dissimilar motion types in different videos. Instead of exploring all the modes exhaustively we select a subset of modes using motion type and the final mode is selected based on the Lagrangian cost function. The experimental results show that compared to HEVC the average computational time can be downscaled by 34% while providing the similar *rate-distortion* (RD) performance.

Categories and Subject Descriptors

I.4 [Image Processing and Computer Vision]: I.4.2 Coding.

Keywords

HEVC; Phase Correlation; Motion Types; Intermode Selection

1. INTRODUCTION

HEVC has the better ability to compress video data by keeping the same perceptual image quality compared to its predecessor H.264 using a number of innovative tools for example, different sizes of *coding units* (CUs), *prediction units* (PUs), and *transformation units* (TUs) [1]. In order to select a particular PU mode, HEVC checks the Lagrangian cost function exhaustively using all modes in selected coding depth levels (level 0: 64×64, level 1: 32×32, level 2: 16×16, level 3: 8×8-pixel) and selects the final mode based on the minimum cost. The cost function of a CU is calculated using the distortion of the reconstructed CU and the weighted (using Langrangian multiplier) bits to encode the block.

In the mode selection process Shen *et al.*[2] propose an algorithm introducing an early termination method based on homogeneity, RD cost and skip mode checking. Gowen *et al.*[3] suggest all zeros in transformed residual for early termination of CU encoding. Podder *et al.*[4] propose a fast intermode selection technique in HEVC and based on the pattern of '0's and '1's in their predefined binary templates they execute mode selection process. As their process cannot fully exploit the complex motion properly it suffers from RD performance and they sacrifice on average 0.24dB PSNR compared to exhaustive mode decision process in HEVC. Other methods in the literature including [2][3] are based on statistical correlation, homogeneity and characteristics of residual among different modes. However, those relationships and analysis are based on the Lagrangian cost functions within HEVC framework. Thus, those methods could not reach the same RD performance with HEVC. On the other hand, the proposed scheme works in two phases where the first phase is based on motion type and category which is independent from Lagrangian cost function. Thus, the proposed scheme can provide similar or better RD performance compared to HEVC if the selection of motion type and category is appropriate. In this paper, we incorporate the phase correlation process in HEVC which approximates relative displacement of the current block against the reference block. Unlike mode selection process in the literature, we exploit three separate categories of motion (no motion, simple/single motion and multiple/complex motion) based on phase correlation to select a subset of ME and MC modes. The final mode from the selected subset of modes is determined by their lowest Lagrangian cost function. Thus, we can precisely categorize different motion types and exploit them for efficient mode selection which results in better RD performance and increased time savings and the proposed scheme saves on average 34% of encoding time with the similar image quality.

VideoNext'14, December 2, 2014, Sydney, Australia.
ACM-978-1-4503-3281-1/14/12.
http://dx.doi.org/10.1145/2676652.2683462.

2. PROPOSED TECHNIQUE

To calculate shifting information between two correlated images we use the phase correlation technique. Based on this technique we determine the *energy concentration ratio* (ECR- a good index of motion identification) of the low frequency component and the total energy of the transformed *phase matched error* (PME) block (details found in [5]). If this ratio is greater than *threshold1*and *threshold2* (Th1 & Th2), the motion types are tagged by '2' (multiple motions) and '1' (single motion) respectively, otherwise the motion type is tagged as '0' (i.e., no motion) where Th1>Th2. After categorizing motions the mode decision is taken at 32×32, 16×16 and 8×8 coding depth levels in order to select a subset of modes and the final mode from the selected subset is determined by their lowest Lagrangian cost function. Note that we use 32×32 and 8×8-pixel blocks as CU size and phase correlation calculation respectively in this paper. The whole process is shown as a block diagram in Figure 1.

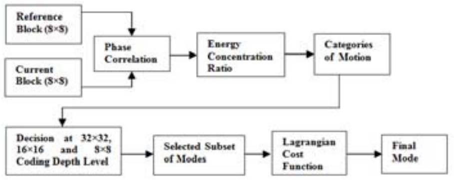
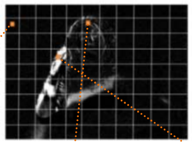


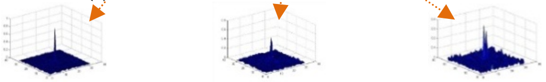
Figure 1.Block diagram of the proposed mode selection process

2.1 Motion Type Categorization

To exploit whether the current block (8×8) has any motion we apply FFT on both the current block and its co-located block in the reference frame. After calculating the phase difference of these two blocks we apply inverse FFT on the resultant phase difference. We finally calculate resultant two dimensional (2-D) array (see Figure 2) having a signal peak at coordinates corresponding to the shift between current block and its reference blocks. In PME, the energy will be concentrated on the upper left triangle of the transformed PME if there is no displacement between the current block and the co-located block. Thus, we calculate the ECR (in Figure 3) of the upper left triangle energy with respect to the entire area energy and finally determine the motion categories against predefined threshold. Figure 3 is identical to Figure 2, where reddish, bluish and other colored blocks represent complex motion, no motion, and simple motion respectively.



(a) Difference between 12th and 13th frame on *Silent* video



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

Figure 2. Motion types at different blocks of 13th frame on *Silent* video; (b)-(d) are the phase shifted plots of different motion types and corresponding energy concentration ratios 0.2, 0.4, and 0.8.

(a) Energy concentration ratio (b) Motion representation map

Figure 3. Motion type identification (a) and its justification (b).

2.2 Intermode Selection

We fully exploit the generated motion types for the subset of inter mode selection process as illustrated in Table 1. It also encapsulates all the selected subset of modes at each coding depth levels and also exemplifies that individual subset of modes are guided by individual motion type. The final mode from the selected subset of modes is specified using their lowest Lagrangian cost function.

Table 1. Selection of modes at 32×32, 16×16 & 8×8 coding depth levels using motion types.

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

2.3 Threshold Selection

In the proposed scheme, we explore different threshold combinations, however, the combination indicated in Table 2 provides improved results for five video sequences we have used from wide range of resolutions and different applications (*Multiview*-MV, *High Definition*-HD or *Standard Definition*-SD sequences).

Table 2. Proposed Threshold at different bit-rates

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

3. SIMULATION RESULTS

In this paper to verify the performance of the proposed scheme experimental results are presented using three SD (*Silent*, *Paris* & *Bridgeclose*), one HD (*Bluesky*) and one MV (*Exit*) videos. Each of the sequences is encoded with search length ±15 (for SD), ±31 (for HD and MV) and frame rate 25 per second. The proposed strategy and HEVC exhaustive mode selection strategy are developed based on HEVC *test model* 8.0 (HM-8.0). Figure 4 (a-e) reveals that in all types of sequences the proposed method retains the similar RD performance with HEVC. Moreover, for a wide range of bit-rates and five divergent video sequences the proposed scheme reduces around 34% encoding time (Figure 4-f) with the loss of only 0.05 dB PSNR (BD-PSNR).

4. CONCLUSIONS

In this paper we incorporate a fast and efficient video coding technique by using motion categorization based on phase correlation. The proposed strategy does not explore all the modes exhaustively but preserves the similar quality of the video sequences with HEVC. The technique outperforms the exhaustive mode selection strategy of HEVC in terms of computational time over a wide range of bitrates and reduces on average 34% of encoding time. This phenomenon influences the video streaming and a new compression algorithm can be combined with the proposed method for the measurement of real time video transmission (especially using Smartphone) over the internet.

5. REFERENCES

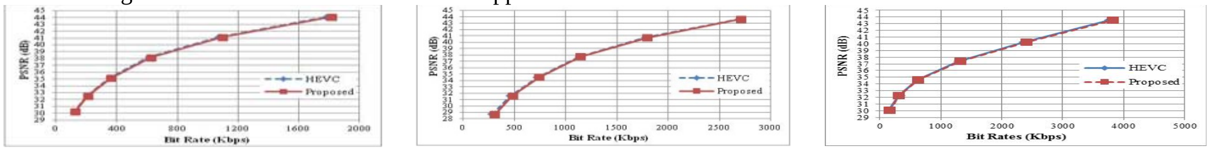
[1] Bross, B., Han, W.J., Sullivan G.J., and Ohm, J.R., “High Efficiency Video Coding Text specification”, draft 9, JTCVC-K1003, October-2012.

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

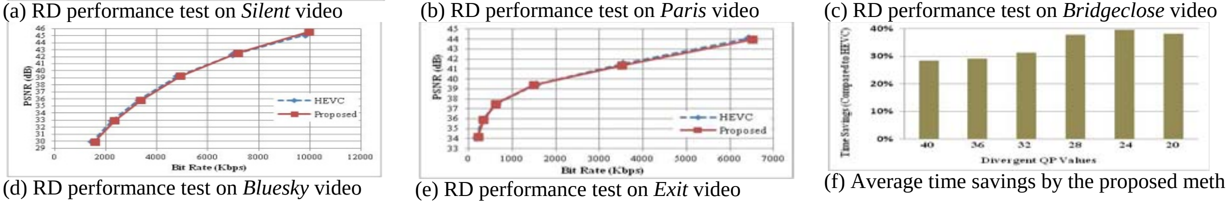
[3] Gweon R., Lee Y., “Early termination of CU encoding to reduce HEVC complexity”, JCT-VC document, JCTVC-F045, 2011.

[4] Podder, P., Paul, M., Murshed, M., and Chakrabarty, S., “Fast Intermode Selection for HEVC Video Coding using Phase Correlation”, IEEE DICTA, in press, 2014.

[5] Paul, M., Lin W., Lau, C. T., and Lee, B.S., "Direct Inter-Mode Selection for H.264 Video Coding using Phase Correlation" *IEEE Transactions on Image Processing*, 20(2), 461-473, 2011.



(a) RD performance test on *Silent* video (b) RD performance test on *Paris* video (c) RD performance test on *Bridgeclose* video



(d) RD performance test on *Bluesky* video (e) RD performance test on *Exit* video (f) Average time savings by the proposed method.

Figure 4. Rate-distortion performance comparison and average percentage of time savings using HEVC and the proposed method.