EFFICIENT VIDEO CODING TECHNIQUE BY EXPLOITING HUMAN VISUAL FEATURES

Author: Pallab Kanti Podder
Supervisor: Dr. Manoranjan Paul
School of Computing and Mathematics
Charles Sturt University, Australia

Abstract

High Efficiency Video Coding (HEVC) standard provides 50% bit-rate reduction compared to its predecessor H.264 with same perceptual quality. The computational time complexity of HEVC encoder has also increased 4-6 times to intensify this compression ratio and could not provide real time facilities to limited processing and battery powered electronic devices.

In this work, we determine human visual features that comprise with three motion features of phase correlation and the saliency feature captured by graph based visual saliency (GBVS) modeling. These area of interest (AOI) indicating visually sensitive features are innovatively combined through a fusion process for a subset of mode selection. The Lagrangian optimization criterion is used for final mode decision. The proposed technique achieves 42% computational time with similar rate distortion (RD) performance.

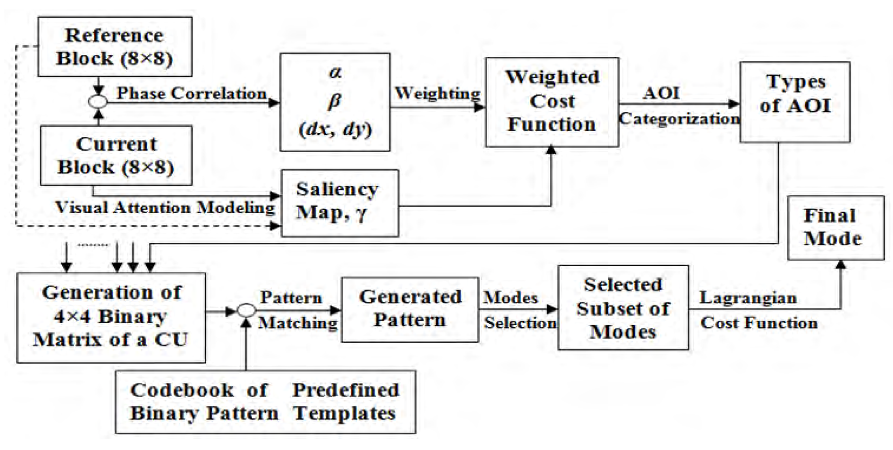


Fig 1: Block diagram of the proposed mode selection process

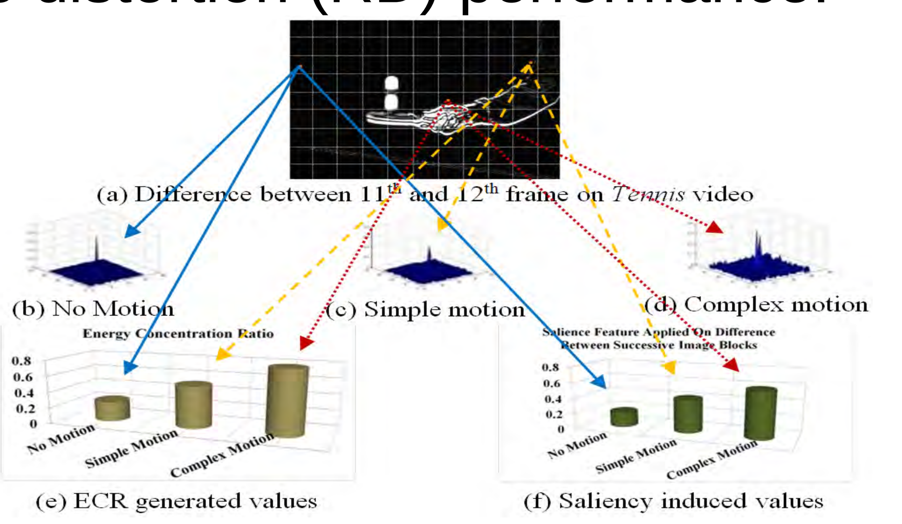


Fig 2: Illustration of motion features and saliency feature generated at different CUs of 12th frame on ***Tennis*** video

Motivation and Problem Statement

Fig 1 shows the process diagram of the proposed mode selection technique where AOI based motion saliency features are incorporated. Fig 2 illustrates different motion categories. The proposed method's effectiveness in terms of determining human visual feature is illustrated in Fig. 3.

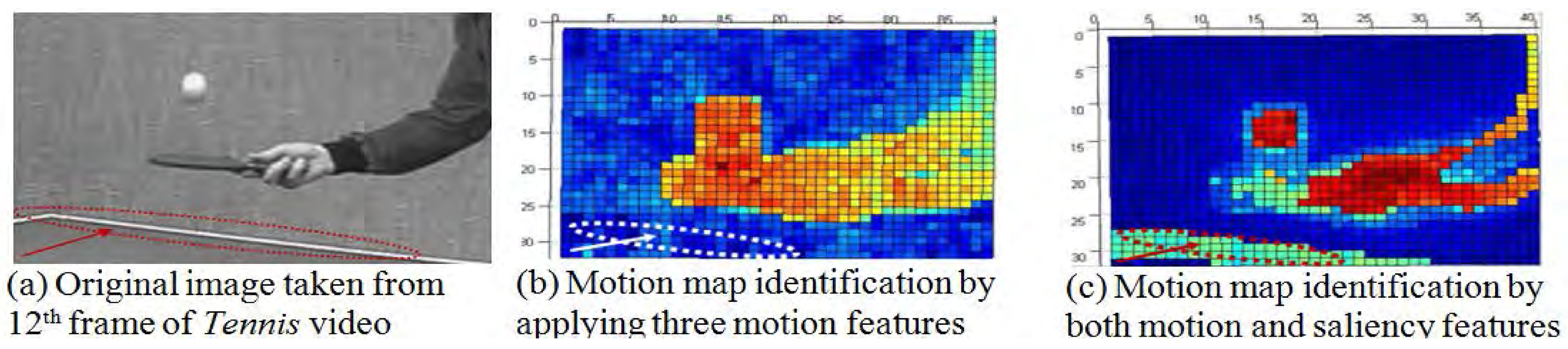


Fig. 3. Identification of motion and salient areas with and without saliency feature based cost function

Only the Lagrangian cost function (LCF) based mode decision would not provide the best RD performance due to different extended features of HEVC standard. Therefore, instead of solely depending on LCF, the proposed techniques works in two distinguishing phases that makes the mode decision process more efficient and less time consuming.

Proposed Intermode Selection

We use 8×8 pixel blocks for binary matrix generation in a CUs (32×32-pixels), thus, a CU has a matrix of 4×4 binary values which then compared to a codebook of predefined binary motion pattern templates (Fig. 4) and perform subset of mode selection at 32×32, 16×16 and 8×8 coding depth levels (Table I and Table II). From the selected subset, the final mode decision is taken from their lowest value of LCF.

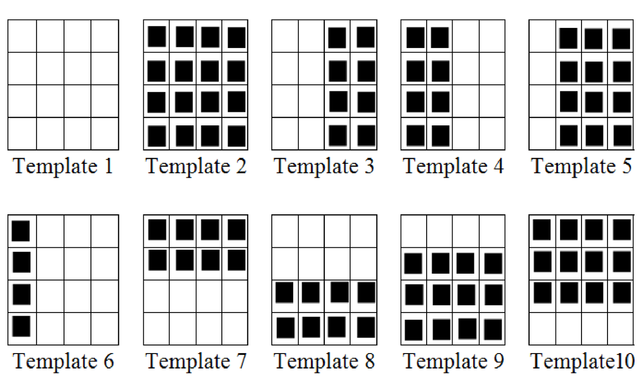


Fig 4: Codebook of the proposed binary pattern templates for a subset of inter-mode selection where template cells with black squares present AOI (i.e., binary 1) and the rest are non-AOI (i.e., binary 0).

Templates Based on AOI at 32×32 Block Level	Selection of Modes at 32×32 Block Level
Template 1	Skip or 32×32
Template 2	Intra 16×16 or Inter 16×16
Template 3 & 4	32×16 or Inter 16×16
Template 5	32×8 or Inter 16×16
Template 6	32×24 or Inter 16×16
Template 7 & 8	16×32 or Inter 16×16
Template 9	8×32 or Inter 16×16
Template 10	24×32 or Inter 16×16

Table 1: Selection technique of inter-modes at 32×32 coding depth level

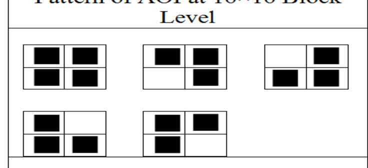
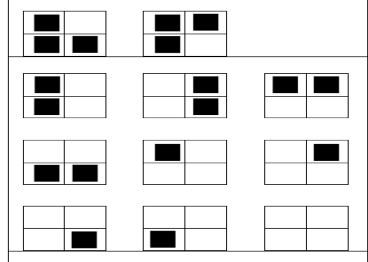
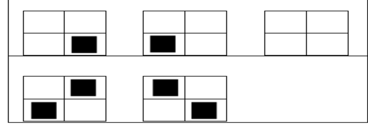
Pattern of AOI at 16×16 Block Level	Selected Subset of Inter-modes
	16×8, 8×16 or 8×8
	16×8 and 8×16
	16×12, 16×4, 12×16 or 4×16

Table 2: Selection technique of inter-modes at 16×16 and 8×8 coding depth level

Experimental Results

CU Partitioning: More appropriate block partitioning is demonstrated and justified by the proposed method (Fig 5 and Fig 2).

Computational Time: Compared to the exhaustive mode selection approach in HEVC encoder (HM12.1), the proposed method reduces 42% computational time (Fig 6 and Fig 7) on average.

Rate-distortion Performance: Proposed technique achieves an almost similar RD performance (small average reduction of 0.01dB PSNR) with a negligible bit-rate increment of 0.13% (Fig 8). It also shows similar subjective quality (Fig 9).

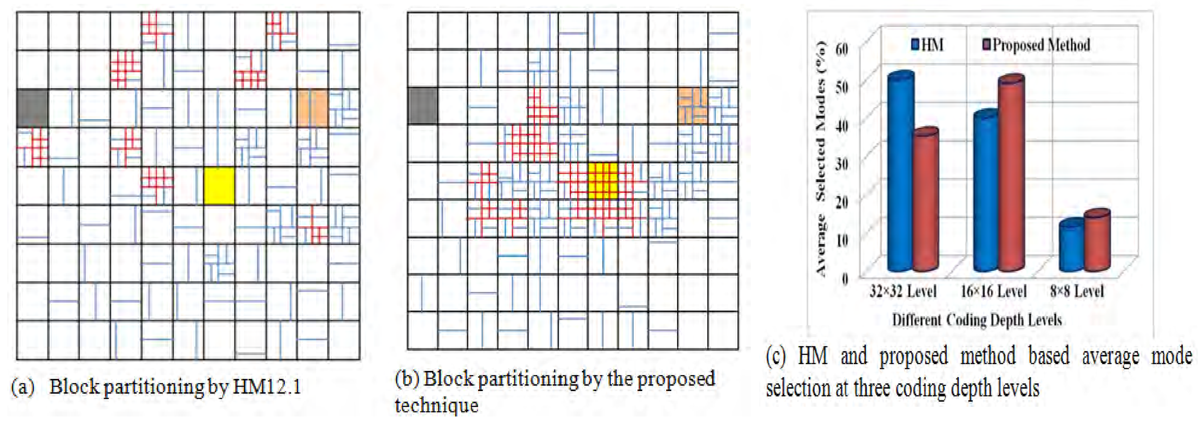


Fig 5: Block partitioning for the 12th frame of the *Tennis* video at QP=24 with HM and the Proposed method and modes at three depth levels.

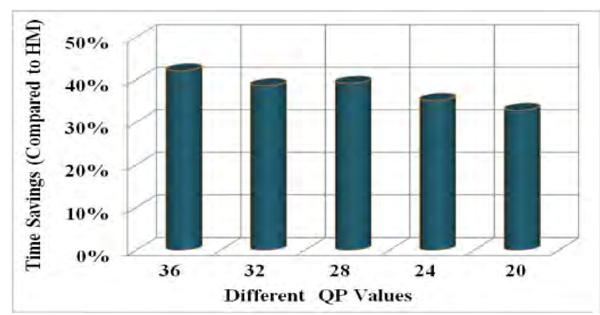


Fig 6: Computational time savings based at different bit-rates

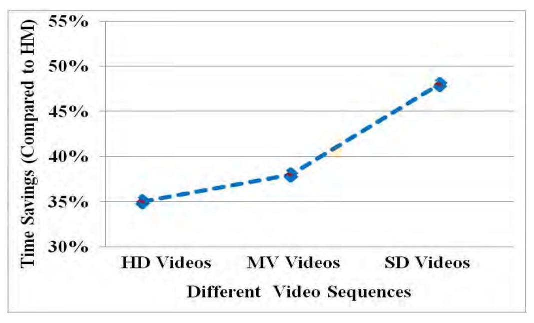


Fig 7: Computational time savings based on video categories

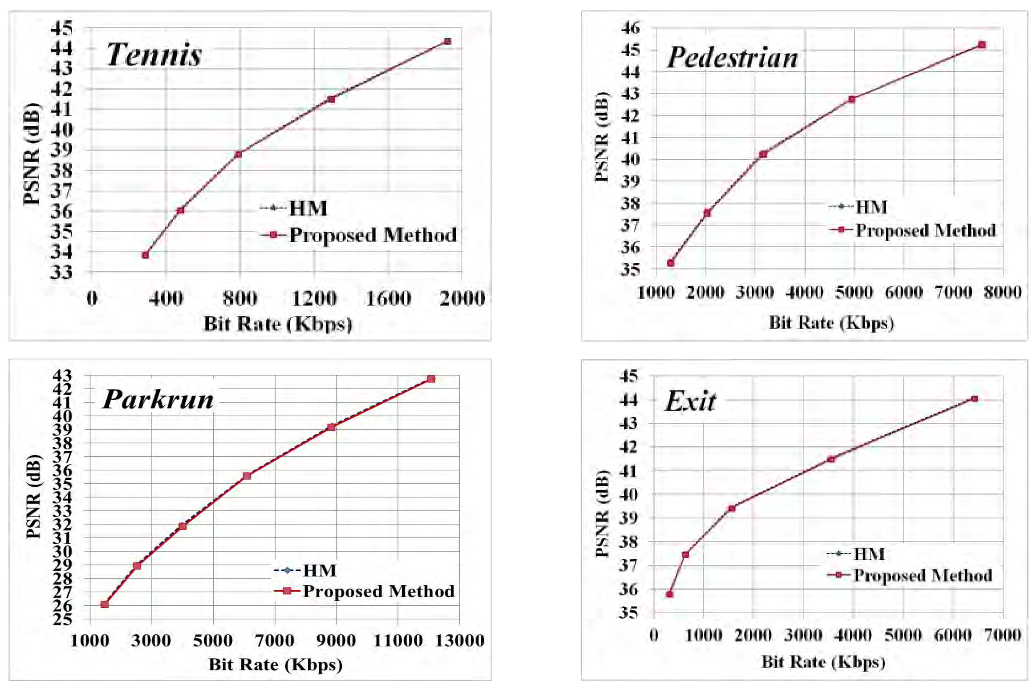


Fig 8: Rate-distortion performance comparison of Proposed method and HM

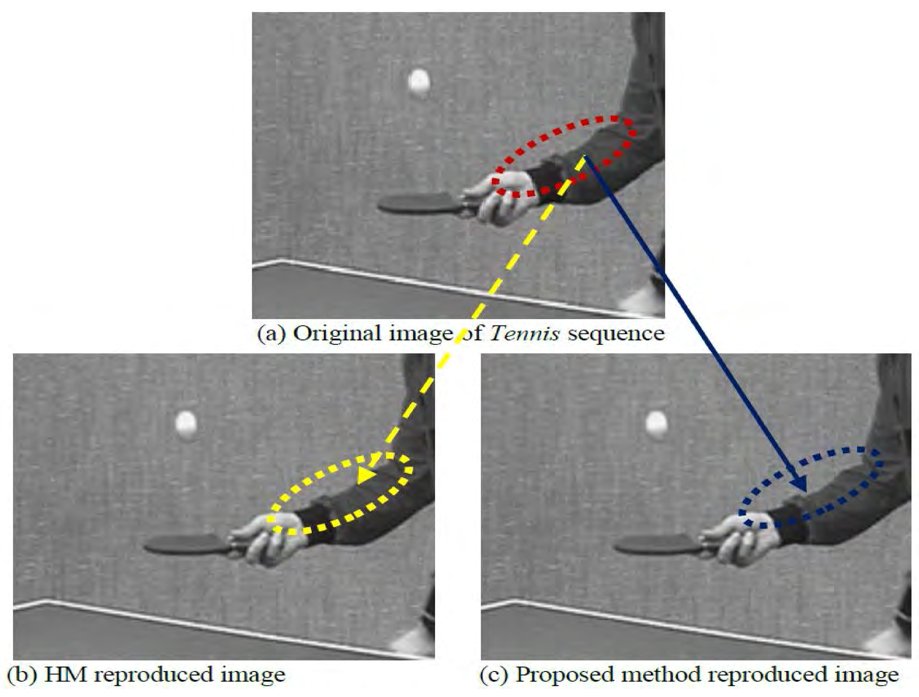


Fig 9: Subjective quality assessment for HM12.1 and the proposed method for *Tennis video sequence. The figures are achieved from the 20th frame of the Tennis video at the same bit-rate.*

Contact details:

Manoranjan Paul

Phone: +61 2 6338 4260

Email: mpaul@csu.edu.au