Compression Efficiency improvement of HEVC Encoder by Introduction of First Pass Encoding

Shree Ganesha Sharma M S, Department of E & C, The National Institute of Engineering, Mysuru,
Rohini Nagapadma, Department of E & C, The National Institute of Engineering, Mysuru,
Prashanth NS, PathPartner Technology Pvt. Ltd, Bengaluru.

***Abstract*— High Efficiency Video Coding (HEVC/H.265) is the next generation video compression standard, the successor of Advanced Video Coding (AVC/H.264). HEVC is expected to improve the compression efficiency by 50% compared to its predecessor. The introduction of a first pass encoder results in dual pass encoder with the pre-encoder feeding the information about the video frame to the second pass. This article discusses an implementation of a pre-encoder module, which analyses the given video, performs motion estimation on a sub-sampled image and provides information about the video frame to the actual encoder. Approximate motion estimation information will help the actual encoder to perform accurate motion estimation, which improves the overall coding efficiency.**

***Index Terms*— HEVC, H.265, video compression, first pass encoder, motion estimation.**

I. Introduction

High Efficiency Video Coding, HEVC/H.265 is the newest video compression standard developed by the collaboration of the ISO/IEC Moving Pictures Experts Group (MPEG) and the ITU-T Video Coding Experts Group (VCEG) as the Joint Collaborative Team on Video Coding (JCT-VC). The major video coding standard directly preceding the HEVC project was H.264/MPEG-4 Advanced Video Coding (AVC), which is widely used for applications like the broadcast of high definition (HD) TV signals over satellite, cable, camcorders, Blu-ray discs, and real-time conversational applications. However, an increasing diversity of services, the growing popularity of HD video, and the emergence of beyond-HD formats (e.g. 4k×2k or 8k×4k resolution) are creating even stronger needs for coding efficiency superior to H.264/MPEG-4 AVC's capabilities [1].

HEVC is a block-based hybrid video codec and aims to cater to diverse applications like broadcast, internet streaming, medical imaging, remote video surveillance and storage media [2]. Its advantages include a 50% improvement in the compression efficiency compared to the AVC, accommodation of parallel architectures and higher resolutions like 8K. However, the computational complexity of the HEVC encoder and decoder is more than 10x and 1.5x compared to the AVC encoder and decoder respectively.

In a single-pass system, the encoder analyzes and encodes the video data "on the fly". Single-pass encoding is used when the encoding speed is crucial — e.g. for real-time encoding. Multi-pass encoding is used when the encoding quality is important. The most common multi-pass encoding scheme is the two-pass encoding. In the first-pass, the rate and distortion information of each frame are collected to model the frame scene complexity. In the second pass, the data collected from the first pass is used to achieve the best encoding quality. Multi-pass encoding takes much longer than single-pass encoding in terms of encoding time and is not suitable for real-time encoding requirements such as live broadcast or live streaming.

II. Preliminary Motion Estimation

The goal of the pre-encoder is to do a coarse motion estimation on a sub-sampled image to get approximate motion vector for every block. Approximate motion estimation will help the second pass, which is the actual encoder, in performing accurate motion estimation. The pre-encoder is hereafter referred to as PreME (stands for Preliminary Motion Estimation). We use Three Step Search for motion estimation for motion estimation and Sum of Squared Difference (SSD) as the measure of distortion. To reduce the coding complexity, the first pass does motion estimation only at depth 0. Fig. 1 depicts the primary blocks of the first pass encoder. The input to the pre-encoder is a 64 * 64 pixel block (at depth 0), also called a Largest Coding Unit (LCU). After the first pass, the control is passed to the actual HEVC encoder.

Initialization block executes once for every Coding Unit (CU) in the video and its function is to initialize variables used in the consequent blocks and to acquire information regarding the CU. Variables initialized are the initial search range for the step search algorithm and variable to store distortion values of reference CUs. The CU information acquired includes the address of starting point, stride, height and width of the current and reference pictures, and the pixel position (X, Y) of the present CU with reference to the frame.

The List Generation block acquires motion vectors (mvX, mvY) of the candidate CUs neighboring the current CU. Fig. 2 shows the six neighbors considered, namely Above Left CU, Above CU, Above Right CU, Left CU, Collocated in List 0, and Collocated in List 1. A default motion vector (0, 0) is also used as a test candidate. If any of the neighbors are yet to be encoded, their motion vector is set to (0, 0).

In some cases,

$$\textit{Target coordinates, } (X', Y') = (X + mvX, Y + mvY) \quad (1)$$

cross the picture boundaries. The third module, Motion Vector Clip block clips such target coordinates to frame boundaries.

The Distortion block calculates the qualitative difference between the current CU and the reference CU. The best match is the candidate CU that results in the minimum error. The SSD of two blocks, A and B, with n x n pixels is calculated as

$$SSD(A,B) = \sum_{i=1}^{n} \sum_{j=1}^{n} \left[A(i,j) - B(i,j)\right]^2 \qquad (2)$$

It is highly probable that a better motion vector can be found in the vicinity of the best motion vector found by the List Generation block. In Step Search block, nine positions around and including the center (x, y) are tested and the position that gives the least distortion becomes the center for next stage. For a step size s, the pixels (x - s, y - s), (x - s, y), (x - s, y + s), (x, y - s), (x, y), (x, y + s), (x + s, y - s), (x + s, y), and (x + s, y + s) are examined [3]. The initial step size is four and after each stage, the step size is halved until the step size is one. Fig. 3 shows an example of Three Step Search [4].

The Final module Interface integrates all other modules. It picks the best motion vector for each CU from the list generated by block 3 and calls necessary functions in the first pass at appropriate times.

III. Experimental Results

The first pass encoder was implemented on HEVC Test Module/HM code 14.0. In order to evaluate the performance of the pre-encoder, five YUV video sequences of different resolution and frame complexity were used. The GOP size is 1. The coding structure is IPPP… Each sequence was tested for 4 Quantization Parameters (QP). TABLE I summarizes the results with a BD rate analysis. TABLES II, III, and IV elaborate on this result. TABLE II shows the improvement in the bit rate, TABLE III shows the reduction in the PSNR-Y, and TABLE IV denotes the excess time taken due to PreME.

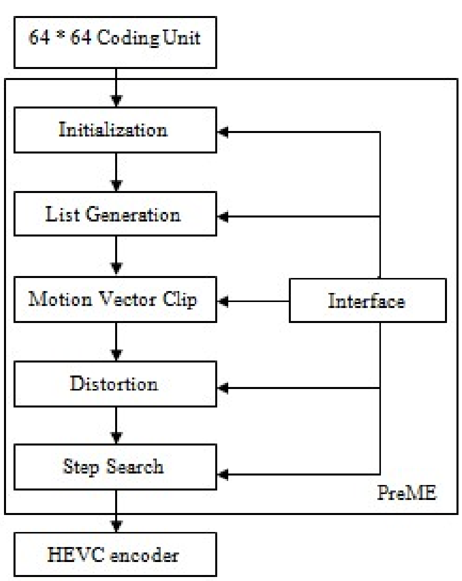

Fig. 1. The primary blocks of the PreME module.

The sequences Suzie, Waterfall, Bus, Bluesky, and Riverbed are in the increasing order of picture complexity, from Suzie with a negligible motion to Riverbed with constant and fast movements. The improvement in bit rate ranges from 1.209% to 4.987%. The PSNR-Y varies from a loss of 0.261% to a gain of 0.036%. The excess time required due to the first pass is up to 9.232% higher than the base / single pass system.

The BD rate analysis shows the fraction of bits required to represent the sequences while maintaining the same quality. For example, the table 4 shows that for the sequence Suzie, 3.032% fewer bits are required to represent the video with the introduction of the pre-encoder than with the single pass HEVC encoder.

IV. Conclusion

The introduction of the first pass encoder has resulted in an improvement in bitrate with very less trade off in PSNR and excess coding time. The reduction in the file size and hence the improvement in compression ratio is more significant compared to the excess encoding time and the loss in quality.

References

[1] Gary J. Sullivan, Jens-Rainer Ohm, Woo-Jin Han, Thomas Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard pre publication draft", IEEE TRANS. On Circuits and Systems for Video Technology, Dec 2012.
[2] ITU-T, "High Efficiency Video Coding", Recommendation H.265, Series H: Audiovisual and Multimedia Systems, Infrastructure of audiovisual services – coding of moving video, 10/2014.
[3] Zong Chen, "Efficient Block Matching Algorithm for Motion Estimation", International Journal of Computer, Electrical, Automation, Control and Information Engineering Vol:3, No:11, 2009.
[4] Aroh Barjatya, "Block Matching Algorithms For Motion Estimation", DIP 6620 Spring 2004 Final Project Paper.

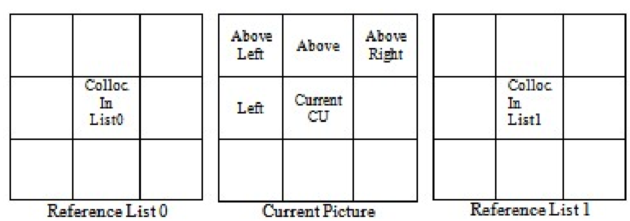

Fig. 2. The neighbors considered by the List Generation Block for motion estimation by the PreME module.

International Journal of Scientific & Engineering Research, Volume 7, Issue 12, December-2016
ISSN 2229-5518

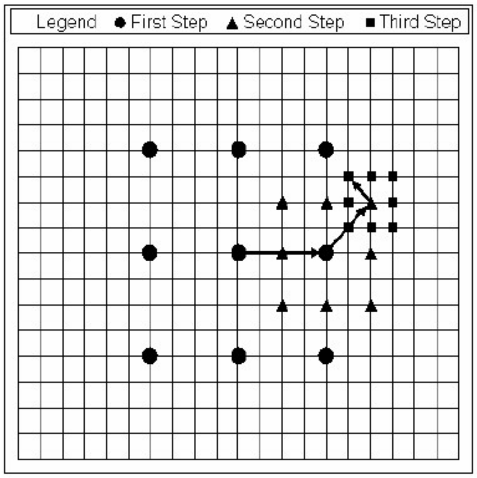

Fig. 3. Three Step Search algorithm for motion estimation.

TABLE I
DETAILS OF TEST SEQUENCES AND BD RATE ANALYSIS

Sl. no	Sequence	Resolution	FPS	Frames Encoded	BD Rate
1	Suzie	176 * 144	25	50	-3.032
2	Waterfall	352 * 288	25	50	-1.406
3	Bus	640 * 480	30	50	-0.540
4	Bluesky	1920 * 1080	25	50	-0.775
5	Riverbed	1920 * 1080	25	30	-0.414

TABLE II
IMPROVEMENT IN THE BIT RATE WITH THE INTRODUCTION OF THE FIRST PASS ENCODER

Sequence	QP	Bit Rate (kbps)		
		Base	With PreME	Gain (%)
Suzie	28	50.487	48.473	-3.988
	32	23.127	21.973	-4.987
	38	10.153	9.807	-3.414
	41	7.473	7.367	-1.426
Waterfall	28	289.740	280.507	-3.187
	32	140.187	137.460	-1.945
	38	49.867	48.413	-2.915
	41	30.293	29.720	-1.892
Bus	28	2420.698	2373.845	-1.936
	32	1354.243	1329.293	-1.842
	38	541.478	530.837	-1.965
	41	337.963	329.784	-2.420
Bluesky	28	2794.108	2730.964	-2.260
	32	1535.420	1501.688	-2.197
	38	682.052	666.756	-2.243
	41	450.820	440.972	-2.184
Riverbed	28	16594.753	16394.113	-1.209
	32	10450.173	10315.920	-1.285
	38	5037.940	4955.193	-1.642
	41	3364.747	3313.147	-1.534

TABLE III
LOSS IN PSNR-Y WITH THE INTRODUCTION OF THE FIRST PASS ENCODER

Sequence	QP	PSNR-Y		
		Base	With PreME	Loss (%)
Suzie	28	37.915	37.856	-0.157
	32	35.453	35.384	-0.194
	38	32.183	32.195	0.036
	41	30.759	30.691	-0.223
Waterfall	28	36.402	36.314	-0.243
	32	33.398	33.342	-0.167
	38	29.352	29.339	-0.046
	41	27.741	27.736	-0.018
Bus	28	39.224	39.180	-0.114
	32	36.597	36.533	-0.175
	38	32.746	32.691	-0.168
	41	30.876	30.796	-0.261
Bluesky	28	41.654	41.608	-0.111
	32	39.572	39.514	-0.146
	38	35.928	35.859	-0.193
	41	33.954	33.876	-0.229
Riverbed	28	39.058	39.021	-0.094
	32	37.026	36.985	-0.111
	38	33.802	33.752	-0.147
	41	32.194	32.153	-0.126

TABLE IV
EXCESS TIME TAKEN BY THE FIRST PASS ENCODER

Sequence	QP	Time (s)		
		Base	With PreME	Excess (%)
Suzie	28	11.7	12.5	6.783
	32	9.1	9.7	6.784
	38	7.8	8.1	3.723
	41	7.6	7.5	-1.965
Waterfall	28	55.0	56.2	2.240
	32	40.6	42.8	5.572
	38	32.5	34.9	7.381
	41	29.9	32.5	8.731
Bus	28	425.2	435.8	2.492
	32	344.6	355.1	3.042
	38	255.6	268.6	5.065
	41	226.2	239.5	5.851
Bluesky	28	1614.2	1698.2	5.206
	32	1425.9	1482.0	3.932
	38	1171.8	1271.5	8.502
	41	1107.3	1209.6	9.232
Riverbed	28	2303.3	2316.3	0.566
	32	1980.3	2043.5	3.195
	38	1664.1	1727.5	3.814
	41	1507.3	1581.0	4.884