*International Journal of Engineering & Technology*, 7 (2.4) (2018) 93-96

**International Journal of Engineering & Technology**

*Website: www.sciencepubco.com/index.php/IJET*

***Research Paper***

# Review on using Region of interest for HEVC

**Parmeshwar Kokare [1] , Dr. MasoodhuBanu. N.M [2]**

*Research Scholar[1], Professor, ECE Dept [2]*
*Vel Tech university, Avadi, Chennai, India [1,2]*

**Abstract**

High efficiency video coding (HEVC) is the latest video compression standard. The coding efficiency of HEVC is 50% more than the preceding standard Advanced video coding (AVC). HEVC has gained this by introducing many advanced techniques such as adaptive block partitioning system known as quadtree, tiles for parallelization, improved entropy coding called Context-Adaptive Binary Arithmetic Coding (CABAC), 35 intra prediction modes (IPMs), etc. all these techniques have increased the complexity of encoding process due to which real time application of HEVC for video transfer is not yet convenient. The main objective of this paper is to provide a review of the recent developments in HEVC, particularly focusing on using region of interest (ROI) for reducing the encoding process time. Summaries of the different approaches to identify the ROI are discussed and a new method is explained.

***Keywords:*** *HEVC; Intra prediction modes; Quadtree partition.*

## 1. Introduction

In today's fast passing internet world, the use of 4K and Ultra high definition (UHD) video streaming has increase the need of the large bandwidth for transmission. But it's pointless to allocate large bandwidth for the sake of entertainment and communication. Here comes the need to develop efficient compression techniques/standards. Latest H.265/HEVC (High Efficiency Video Coding) standard has manage to encode video sequence in much lower bite rate as compared to previous H.264/AVC standard.

H.265/HEVC is the latest video coding standard which has the best compression up till now. The quality of encoding video is almost the same as that of H.264/AVCencoding video. But the advantage of H.265/HEVC is that it needs half the amount of bits required to encode a video using H.264/AVC [1]. In simple way if H.264/AVC standard requires 4MB data to compress 4K video sequence, H.265/HEVC can encode the same sequence into 2MB while maintaining its quality.

H.265/HEVC has introduced many improved features in video encoding techniques such as.

- Unlike H.264/AVC which allows largest block size as 16x16, HEVC can have a much larger block size 64x64 pixels.
- Flexible content dependent block partition called Quad tree partition which is unique.
- Separate block partition system for prediction blocks and transform blocks after quantization.
- HEVC supports 35 IPMs instead of only 9 IPMs supported by H.264/AVC, which ensures more accurate prediction.
- HEVC includes Adaptive Motion Vector Prediction, a new method to improve inter-prediction.
- An improved deblocking filter.
- Sample Adaptive Offset an additional filter that reduces artifacts at block edges.

The excellent performance of H.265/HEVC has a drawback of computational complexity. The level of complex algorithms has increased due to which real time encoding if video by H.265/HEVC is more time consuming. Various researchers have proposed algorithms to decrease the encoding time of H.265/HEVC. Beside real time applications H.265/HEVC compression technique is useful in storage of satellite video and images and medical images/video, which are full of crucial information.

Most of the time only a part of video frame or image contains desired information. This is common in video surveillance, medical video, and satellite videos. So, it will be more beneficial to encode desired region/ part of video / image with more accuracy and to encode the remaining part nominally. This will reduce the encoding time effectively while maintaining details of desired content. One way of recognizing desired part is by using ROI.

Today's highest challenge in capturing is to adapt the camera vision to human visual system (HVS). Normally HVS focuses

more on a certain part of any video/ image. Other information is not always crucial. Thus identifying those areas which gain the focus of human will be beneficial in reducing encoding time of video. The aim of future lossy video compression technique is to reduce entropy while maintaining the needed quality of the video and thus there is requirement of high end algorithms for selecting the region of interest. Section II gives the brief description of H.265/HEVC encoding process. Various techniques for analyzing the data in video for extracting or detecting meaningful content are explained in section III. Section IV gives the application where lossyH.265/HEVC encoding schemes can be used. Section V explains a proposed method for low bitrate transmission. Section VI concludes the research article.

## 2. HEVC coding design

The encoding algorithm of generating H.265/HEV Cbitstream typically proceeds as follows. First each frame is split into fixed size block-shaped regions, which is also conveyed to the decoder as it is. Intra prediction is performed on the first frame of a video. In intra prediction only the spatial redundancies are exploited and it is not dependent on other frames. All other video frames blocks are mostly encoded by inter prediction. Motion data like motion vectors which are found by comparing sequential frames in video contribute in inter prediction encoding. The encoder and decoder generate identical interpicture prediction signals by applying motion compensation (MC) using the MV and mode decision data, which are transmitted as side information. The difference between the original frame and its predicted frame is called residual signal. The residual signal is then transformed by a linear spatial transform. The transform coefficients are then quantized, entropy coded, and transmitted together with the prediction information [2].

**Flexible CTU partitioning structure**

In HEVC quad tree block partitioning structure is used. It allows partitioning a video frame into fixed size blocks called Coding tree units (CTU) [3]. The size of all CTUs can be 64x64, 32x32 or 16x16. CTU is the basic and logical unit. It contains three blocks, one luma and two chroma along with syntax element. These blocks called Coding tree blocks can be split to four small Coding blocks (CBs) until the size becomes size 8x8 as shown in fig. 1. The size of CTU is chosen by the encoder according to the characteristics information present in the block. Rate distortion (RD) cost of each CTU including all CBs is found and then the decision is made of splitting or not splitting. The RD cost of constant information block is less and that of variation containing block is large. By assigning a threshold value the decision is made. Thus the flat portion of the frame has largest block size while the blocks having variations are further split into smaller size.

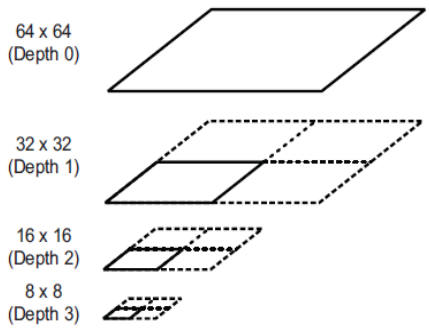


**Fig. 1** Quadtree structure following top- down order [3].

CB is the decision point where it is decided whether to perform inter or intra prediction. In inter prediction motion estimation is performed to get motion vectors. For this reason CB are further split into Prediction blocks (PBs) depending on temporal and spatial prediction. The intra prediction is done by comparing the neighboring blocks. For this there are 35 Intra prediction modes available in HEVC. The mode contains DC mode, planar mode and 33 angular modes as shown in Fig. 2 .The residual signal which is difference between the original and predicted blocks is the transformed and quantized. The quantized blocks are partitioned into square shape transfer blocks (TBs) to reduce bits required to transfer.

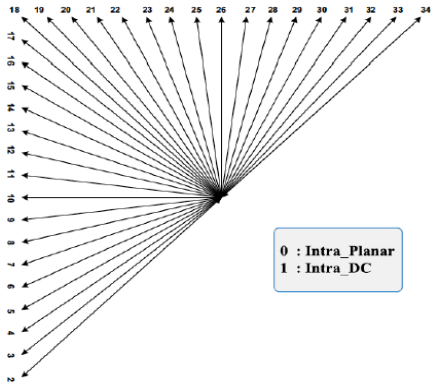


**Fig. 2** Intra Prediction Modes

Finally to get the most optimal intra prediction the encoder checks all possible combinations of CTU, PB, IPM and TB and choses the best combination. It is called Rate Distortion Optimization(RDO). This extra effort significantly increases the encoding time of HEVC.

## 3. ROC Based Video Coding

Region of interest (ROI) based video encoding has the advantage of assignment for two priorities i.e. high or low. Further we can encode the high priority region with lossless encoding technique and low priority region can be encoded by lossy methods. The definition for high and low priority assignment can differ from one application to other. ROI can be found by exploiting spatial information of the video frame and then by some temporal data from other frames. Further are some of the ways to identify the region of interest for H.265/HEVC video encoding.

**A. Saliency based**

Visual saliency is an idea that from the whole video sequence only some part is distinctive which creates immediate simulations in human visual system [5]. Saliency can be combinations of geometric objects, texture complexity, rarity and sometimes shape. A saliency aware fast intra coding algorithm for H.265/HEVC proposed in [6] consists of perceptual intra coding and fast intra prediction mode decision algorithm. The principle used in this paper is that human eyes generally focus on different areas at different levels. The subjective video quality is mainly concerned with the quality of focused area and the distortion of unfocused area barely influences the quality. Thus visual saliency value is high at areas that attract human attention and lower in the background.

Hence a smaller CU size is chosen for those areas with high attention, and a larger CU size for those areas with low attention. The relationship between the visual saliency value and block partition depth is adjusted properly to achieve good coding performance shown as in Table I.

It is seen that visual saliency and depth level are directly proportional to each other. Higher the level of partition higher will be the complexity. So prediction of depth level in advance will minimize the complexity of RD calculation. Experimental results show that 45.39% encoding time can be reduced by the proposed saliency aware fast intra coding algorithm and 2.18% bit-rate reduction with negligible perceptual quality loss.

**Table I.** The relationship between the visual saliency value and block partition depth

| VS value | 0-0.20 | 0.20-0.35 | 0.35-0.56 | 0.56-1 |
|---|---|---|---|---|
| Depth | 0 | 1 | 2 | 3 |

### B. Rate Control using tiles

The H.265/HEVC reference software contains two rate control (RC) algorithms, first is Rate distortion model with mean absolute difference (MAD) and the second is R- $\lambda$ model. The HEVC test Model (HM.13) software has more improvement is done based on characteristics of video content. All these features provides for an effective Rate Control model. In [7] authors have proposed a similar way for RC. They achieved low bitrates and an improved ROI quality but at the cost of low non-ROI quality. The scheme in [8] provided improved visual quality for non-ROI region by appointing independent rate between different tile regions. Encoding and transmission of the tile regions is also done independently.

### C. Clustering method

For Identifying ROI grouping of some meaningful events is done in blocks which can be done by breaking down higher level structure. In case of video grouping can be done on the bases of spatial and time regions. By exploring regions in spatial-time approach certain homogeneity can be detected. Homogeneity can be same color or a certain group of motion vectors. Such homogeneous groups can be used as region of interest. Now it is difficult to formulate a clustering algorithm for a video having different features and meaning. Clustering is giving common name to a set of data points which belong to same circle. Few traditional clustering approaches are k-means, normalized graph-cut, mean-shift and Gaussian-mixture modeling. Unsupervised clustering method proposed in [9] separates clustering of groups having different features. This method of clustering helps to compose a meaningful object.Generally objects in most videos are complex in shape, texture and motion. Natural objects in most videos are generally complex in shape and color content. However by using grouping technique individual part can be clustered to encode the whole object.

### D. Skin color detection

Color of the skin special of face is most important cue for human localization. It's a popular concept in detecting human face in a video or image frame. There are many skin classifiers defining boundaries of the human face. But its very challenging to achieve high accuracy in case of many faces of different shades in a single frame. In [10] a human face detection technique is integrated in H.265/HEVC encoder. Intra and inter frames are exploited to get skin color information without searching every pixel in the frame. This has reduced the time required for the skin detector to locate the human face. This information is then appended to the video encoder output. So by using high Quantization Parameter (QP) for the human face area and using low QP for rest of the background much encoding time can be saved.

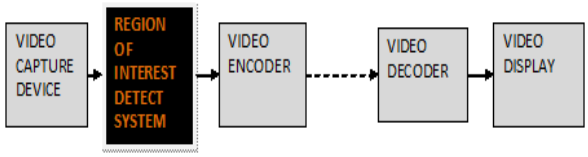


*Fig. 3 General block diagram of ROI in Video encoding*

### E. Texture based

In medical images and videos textural material is in focus than any other information. To identify the textural info from video frequency-domain based image enhancement is used in [11] for transform coding in prediction in H.265/HEVC. A matrix of transform coefficient equilibrium is formed on the bases of transform coefficient of prediction residual. A parameter $T_D$ is introduced which defines textural density i.e. degree of variation in textural information. Value of $T_D$is assigned according previously extracted ROI and intra prediction modes in H.265/HEVC. The directional variation in textural information is indicated by angular intra prediction modes. For every angular intra prediction modes in PU,$T_D = 2$. For homogeneous texture in PU, DC or planar intra prediction modes are used. These are denoted by $T_D = 1$. For a non-ROI region $T_D$is always assigned as 0. This tactic improves the textural significance by adjustment of transform coefficients in frequency domain. The accuracy and consistency of medical video for diagnostic purpose is maintained while reducing the encoding time.

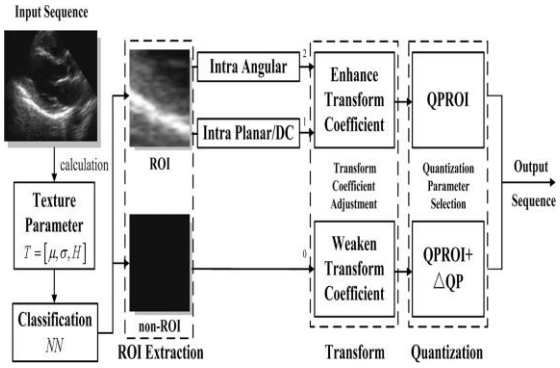


**Fig. 4** Hierarchical coding framework

## 4. Analysis of Methods and their Applications

The first method described is visual saliency, which is very popular for its applications in computer vision. It can be used in almost every field like satellite communication, entertainment, HD TV, etc. Rate control using tiles in HEVC is the simple yet high time achieving method. This method will be beneficial in channel broadcasting, to encode movies, and online streaming of videos. Clustering is a complex method with moderate results which can be used for traffic surveillance. For news channel broadcasting skin color detection method will be most suitable. Texture based

method will reduce the size of medical images and encoding time of medical video coding.

## 5. Proposed Approach

We propose a new arrangement of video based on ratio control which will require lower bit rates. This method will be especially for video conferencing. First the video will be filtered to remove the noise content, which will remove the high frequency flickers. Then the face region will be detected separately from the video which will further reduce encoding time. The background and face area will be processed using different algorithms. New adaptive quantization parameter method will be used for face region and single quantization parameter for background, which will help to reduce bit rate and encoding time.

## 6. Conclusion

In this paper the latest algorithms to reduce the encoding time of HEVC by using region of interest technique are discussed fundamentally. By finding the particular region of interest we can cut down unnecessary processing time of algorithms effectively and hence improves the encoding time of HEVC. Such approach is proposed that will help accelerate encoding speed with reduced bit rate.

## References

[1] Ohm J R, Sullivan G J, Schwarz H, Tan T K, Wiegand T, "Comparison of the coding efficiency of video coding standards—including high efficiency video coding (HEVC)." Circuits and Systems for Video Technology, IEEE Transactions on 22.12 (2012): 1669-1684.

[2] Gary J. Sullivan, Jens-Rainer Ohm, Woo-Jin Han and Thomas Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard", IEEE transactions on circuits and systems for video technology, vol. 22, no. 12, December 2012.

[3] II-Koo Kim, Junghye Min, Tammy Lee, Woo-Jin Han, and Jeong Hoon Park, "Block Partitioning Structure in the HEVC Standard", IEEE transactions on circuits and systems for video technology, vol. 22, no. 12, december 2012.

[4] Thomas E. Slowe and Ivan Marsic, "Saliency-based Visual Representation for Compression", In Proc. IEEE Int'l. Conf on Image Processing (ICIP), 1997, vol.2, pp. 554 – 557.

[5] Yuan Gao, Pengyu Liu, Yueying Wu and Kebin Jia, "Quadtree Degeneration for HEVC," IEEE Transactions on Multimedia 2016, Vol. 18, Issue: 12.

[6] LiyuanXiong, Wei Zhou, Xin Zhou, Guanwen Zhang and Ai Qing, "Saliency Aware Fast Intra Coding Algorithm for H.265/HEVC", 2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA).

[7] M. Meddeb, M. Cagnazzo, and B. Pesquet-Popescu, "Region-of-Interest Based Rate Control Scheme for High Efficiency Video Coding," in Proceed. of IEEE Intern. Conf. Acoust., Speech and Sign.Proc., Florence, Italy, 2014.

[8] M. Meddeb, M. Cagnazzo, and B. Pesquet-Popescu, "Roi-based rate control using tiles for an H.265/HEVC encoded video stream over a lossy network," in 2015 IEEE International Conference on Image Processing (ICIP), Sept 2015, pp. 1389–1393.

[9] SubarnaTripathi, "ROI based Parametric Video Coding Scheme", A Thesis submitted for the award for the degree of Master of Science (Research), IIT Delhi.

[10] PiyaliGoswami, PalankiVenkataSrikanth, JasminRahiman, "Low Complexity In-Loop Skin Tone Detection for ROI Coding in the HEVC Encoder", IEEE 2015.

[11] Wu Y, Liu P, Gao Y, Jia K (2016) "Medical Ultrasound Video Coding with H.265/H.265/HEVC Based on ROI Extraction", PLoS ONE 11(11): e0165698. doi:10.1371/journal.pone.0165698