High Quality Video Assessment Using Salient Features

K. Bhanu Rekha*[1], Ravi Kumar AV[2]
[1]Department of ECE, Sir M.V.I.T, Bangalore, India
[2]Department of ECE, SJBIT, Bangalore, India
*Corresponding author, e-mail:kbhanu.rekha@rediffmail.com

Abstract
An efficient modified video compression HEVC technique based on high quality assessment saliency features presented for the assessment of high quality videos. To create an efficient saliency map we extract global temporal alignment component and robust spatial components. To obtain high quality saliency here, we combine spatial saliency features and temporal saliency features together for different macroblocks in association with transformed residuals. In this way, our saliency model outperforms all the existing techniques. In this paper, we have generated high reconstruction quality video after compression considering SFU dataset. Our experimental result outperforms all the existing techniques in terms of saliency map detection, PSNR and high-resolution quality.

Keywords: saliency, quality assessment, HEVC, PSNR

1. Introduction

The growth in consumer demand for ultra-High Definitions (UHD) devices like Smart phones, iPad, MacBook Pro, LAPTOPS, HDTV (High-definition television), UHDTV (Ultra-high-definition television) has provide immense popularity to 2k/ 4k/8k videos in the entertainment world due to its high quality visibility and richer color. UHD videos becomes a common requirement in the field of entertainment, medical, photography, satellite imaging, HDTV, stereoscopic video processing, face recognition etc. However, there are few problems comes along with high quality UHD videos such as requirement of high storage capacity, limited battery power of high definition devices, long encoding time, high computational complexity. Therefore, in past decades, evolution of digital video compression methods has modernized the entire research industry to create, link and transfer visual data.

Video compression methods are the combination of spatial and temporal motion compression which decreases the data quantity to a huge amount [1]. Recent advanced Video compression methods consists of some essential features such as coding efficiency, channel robustness and application flexibility. A broadcasting HD video consist of compressed audio and video data, synchronized data, error detection and controlling signals. MPEG-2 [2], H.263 [3], AVC, SVC [4, 5], H.264/AVC [6] standards are some recent existing Video coding standards. However, it consists of few problems likes degradation in quality and coding efficiency, high bitrate, high computational complexity.

In [7], a distributed Map Reduce technique presented to make faster the encoding process based on scheduling and segmentation video encoding. In [8], to identify key-points and achieve efficient feature extraction a Key point Encoding technique presented at lower bit-rates. The reduction in bitrates is satisfactory for a single scene but can be complex for multiple scenes of a video. In [9], an adaptive scheduling adopted for heterogeneous devices to achieve real-time coding by using an arrangement of parallel CPU+GPU cores. However, heterogeneous devices consists of high computational complexity and optimization problem. In [10], a real-time H.264 encoding technique presented for HD (High Definition) videos to enhance the video quality after the efficient compression. However, it is difficult to select suitable features for embedded systems as it requires different features for different scenarios.

Higher bit-rate, optimization problem, high computational complexity, suitable feature selection, large coding time and degradation in coding efficiency are the problems often occurs in existing techniques. Therefore, there is a need of an efficient technique which can improve coding efficiency to a large extent without compromising the high quality of video. Therefore, in

this paper, we have presented an efficient modified video compression HEVC technique based on high quality assessment saliency features.

In this paper, we focus on extraction of high quality saliency features to get high compression efficiency. For the assessment of high quality videos. Here, to create an efficient saliency map we extract global temporal alignment component and robust spatial component. The small alterations between neighboring frames may be not sufficient to define salient regions. Hence, Motion estimation and detection is very critical phenomenon in our saliency model. We can determine saliency map precisely using HEVC (High Efficiency Video Encoding) in association with INTRA or INTER processing blocks with different sizes. To obtain high quality saliency here, we combine spatial saliency features and temporal saliency features together for different macroblocks in association with transformed residuals. Saliency features helps to get high quality compression. In this way, our saliency model outperforms all the existing techniques.

This paper is organize in following sections, which are as follows. In section 2, we describe about the video encoding issues and how they can eliminate by our proposed model. In section 3, we described our proposed methodology. In section 4, experimental results, evaluation shown, and section 5 concludes our paper.

2. Video Encoding Issues

Video compression is an essential need of a modern era due to enormous growth of many HD and UHD devices which requires an ultra-high definition (HD) quality. However, there are some issues which are associated with HD and UHD videos such as high storage capacity, limited battery power of high definition devices, long encoding time, optimization problem, degraded coding efficiency and high computational complexity. Therefore, in past decade many researchers have done some significant work to reduce these above mentioned issues. A brief of related work in the field of video compression presented in the following section.

In [11], a no-reference based SQA (Subjective Quality Assessment) approach adopted to assess and improve quality of video encoding with the help of human eye traversal. In this approach, approximation of smooth eye traversal computed based on distance, angle and pupil-size feature using HEVC (High Efficiency Video Encoding).However, QMET(quality metric based on eye traversal) can be much more flexible if eye-tracking simulator employed in association with QMET. In [12], a 3D HEVC scheme introduced to reduce high computational complexity using online learning. In encoding process, online learning used to tune the two probabilistic models FMA (Fast Mode Assignment) which helps in reducing complexity. However, this scheme can marginally affect the video quality while reducing the computational complexity. In [13], a verified HEVC testing scheme presented to evaluate video quality. To get better efficiency and for bit-rate saving a MOS-based BD-rate measurement presented in association with HEVC approach. However, it increases computational complexity. In [14], asymmetric compressed stereoscopic technique adopted to evaluate high quality assessment and rate-distortion on 3D videos. In this approach, a combination of asymmetric transform coding and mixed resolution used to obtain asymmetric compression with better quality. However, it is highly complex process to implement in real-time. In [15], a quality assessment for streaming videos presented using estimated QOE (Quality of Experienced) Measurement. However, integration of QOE model with adaptive streaming decision making engine for optimal playback control is very challenging issue. In [16], a depth quality assessment approach based on no-reference edge misalignment error presented for texture plus depth T+D images. However, it is difficult to assess completely depth quality in no reference fashion. In [17], a weighted fixation density based approach presented to describe quality assessment using visual saliency map to obtain high quality compression. However, this approach marginally eliminates the central bias problem but not completely using shuffling method. In [18], a novel dynamic feature selection (DFS) model proposed to get high quality visual features to assess high video quality which can improve quality of visual saliency maps. However, measurement of background feature density and reconstruction error computed is very high. In [19], to predict gaze density and improve quality of visual saliency maps an emotion intensity incorporated along with emotional object detection. However, this model failed to define relationship between emotion and visual saliency.

In this section, many techniques described as a related work for high quality assessment for video saliency maps. However, every technique has its own drawback. The basic drawbacks are high storage capacity, high computational complexity, flexibility, coding efficiency, central bias problem and reconstruction error and low quality of visual saliency maps. Therefore, to overcome these issues we have proposed an efficient modified video compression HEVC technique based on high quality assessment saliency features for the assessment of high quality videos.

3.   Quality Saliency Features based Video Encoding

The popularity of High Definition Devices in recent time has changed visual real world due to its realistic visual power and true color. Therefore, availability of high end devices increased the demand of high resolution videos. However, high-definition videos takes large space and bandwidth spectrum. Therefore, to counter these drawbacks, we have presented a video encoding technique based on its quality saliency features using HEVC (High Efficiency Video Encoding) architecture to obtain high quality compression and reconstructed frames. This technique can used in field of medical, photography, satellite imaging, HDTV, stereoscopic video processing, face recognition and video compression to estimate saliency and compress high-definition videos. Our video compression provides fast computation for large training database such as SFU dataset [20].HEVC architecture in synchronization with saliency features can provide compression efficiency to a high extent. HEVC architecture can precisely work with large datasets, can helps to get lower bitrate and can reduce computational time.

There are very few techniques, which can effectively estimate saliency features and compress a high definition video and can provide a high definition visual quality. Therefore, to detect saliency precisely and to provide high quality video after compression, we have presented a video encoding technique based on its quality saliency features using HEVC (High Efficiency Video Encoding) architecture.

3.1. Saliency Aware Video Compression

Our proposed video compression technique based on saliency map consists of following principles.
a. In a saliency map while compression, low salient regions contains lower perceptual quality and higher salient regions contains higher quality. This demonstrate that quality is focused towards those viewers who probably wants to look.
b. Encoding should provide saliency features of different regions .The saliency of a low resolution video can be increased after compression if that area is highly salient initially. Therefore, this shows that region is of very high quality where large number of viewers can get attract.
c. Initially, if any region has low saliency in a video then its saliency can further increase after compression. Therefore, this shows that region is of very lower quality and probably less number of viewers can get attracted for such regions.

3.2. Detection of Saliency Regions

There are many factors, which can make video compression very effective such as prediction of motion and motion compensation, then transformation, quantization, estimated residuals and motion vectors entropy etc. Therefore, salient map can be an effective way to demonstrate precise quality compression. A region which attracts human eyes with highest probability concern can be called as salient region. A salient region can be colored, high resolution and motion. Here, to create a saliency map we extract global temporal alignment component and robust spatial component. The small alterations between neighboring frames may be not sufficient to define salient regions. Hence, Motion estimation and detection is very critical phenomenon in a perceptual coding. We can determine saliency map precisely using HEVC (High Efficiency Video Encoding) in association with INTRA or INTER processing blocks with different sizes to obtain high compression rate.

3.3. The IKN Saliency Model

There are number of conventional models exists in real world to define saliency model in video encoding field. Here, we have used Itti-Koch-Niebur (IKN) model [21] to define saliency

map estimation due to its wide popularity and low computational complexity. There are sufficient number of independent feature mediums are available to estimate saliency map by exploring the input images/frames. Every feature medium handle a low-level visual features inside an image. These low-level visual features are intensity, color and orientation contrast. There are total nine spatial scales are produced for image downsampling and gradually pass to low pass filter utilizing dyadic Gaussian pyramids to form a reduction factor of image size whose range is scale zero (1:1) to scale eight (1:1) [21].

A 'center-surround' method adopted to compute contrast for every feature medium. This center surround method can be defined as the change between Coarse and fine scales. Here, 'scale' represents the pixels at scale $l \in \{2,3,4\}$ and 'surround' shows pixels at scale $d = l + s$ where $d \in \{3,4\}$. Interpolation and point- by point subtraction methods are used to find scale difference at finer scale. A normalization operator used to combine the extracted contrast features to develop a ''conspicuity map'' for every feature medium. The similar normalization operator can be used to combine all the conspicuity maps to create a "final saliency map" after resizing to level-4 by estimating pixel values.

A motion and flicker medium are introduced to IKN model to make it suitable for videos [22]. The flicker medium developed by a creating a Gaussian pyramid which is the difference between present and previous frames. Motion medium developed by an intensity pyramid which is difference between spatially-shifted present and previous frames [22]. The motion and flicker conspicuity maps created using the same center surround method which derived for intensity, orientation and color medium and combined together to form a final saliency map using spatial conspicuity maps.

3.4. Rate Distortion Optimization

H.264/AVC and HEVC video coding standards can support multiple types of block encoding modes like $INTER\ 16 \times 16$, $INTER\ 16 \times 8$, $INTER\ 8 \times 16$ and $INTRA\ 4 \times 4$, $INTRA\ 16 \times 16$, $INTRA\ 8 \times 8$ [23]. To reduce Lagrangian cost function of coding modes we have utilized RDO (Rate Distortion Optimization) method for each macroblock (MB) selection in association with HEVC coding [23-24].

$$K(\Psi|R, \lambda_J) = S_{mse}(\Psi|R) + \lambda_J J(\Psi|R), \tag{1}$$

Where, $S_{mse}(\Psi|R)$ represents MSE (Mean Squared Error) and $J(\Psi|R)$ shows bit rate of current macro block for the coding mode Ψ with R quantization step size. Here, λ_J quantifies the tradeoff comparison between distortion and rate and can be defined as Lagrange Multiplier [24]. λ_J is a specific value for which Lagrangian cost function [25] is minimized. Therefore, to accomplish optimum rate distortion model λ_J is very essential factor. In HEVC coding λ_J can be expressed as,

$$\lambda_J = 0.85 \cdot 2^{\frac{(\mathbb{QP})}{3}}, \tag{2}$$

Where, $\mathbb{QP}$ is the quantization factor.

3.5. Quality Saliency Estimation for a Video

Our model is divided into two components to evaluate saliency map such as SSC (Spatial Saliency Component) and TSC (Temporal Saliency Component). SSC is a convex approximation of IKN spatial saliency whereas TSC is used to predict saliency using global motion compensation by eliminating camera motion to obtain high compression. The prediction of the quality saliency map for a low-resolution video can be achieved in four sections which are as follows:

a. Convex Approximation to Spatial IKN Saliency

In Our IKN saliency model to create saliency map of a video frame we utilize the video frame content in the normalized frequency range $\left[\frac{\pi}{256}, \frac{\pi}{16}\right]$. Here, we presented a convex approximation method to find IKN saliency map. In convex approximation method, block DCT is utilized to predict saliency of that block Z. This can be accomplished by recapturing some section

of the video frame at position Z using normalized frequency range $\left[\frac{\pi}{256}, \frac{\pi}{16}\right]$. In our IKN model, windowing and spectral down-sampling methods used to extract a block from a video frame.

To extract a block from the video frame of desired size windowing method used and then its 2-D DCT evaluated. The resultant DCT can be expressed as $X_d(j, k)$. Assume that another resultant DCT can be denoted as $X_w(j, k)$ which covers $[0 - \pi/256) \cup (\pi - 16/\pi]$ frequency band. Weiner Filter DCT coefficients can be expressed as

$$T(k, i) = \frac{Xd^2(k,i)}{Xd^2(k,i) + Xw^2(k,i)}, \tag{4}$$

Where, $X_Z(k, i)$ is $(k, i)th$ 2-D DCT coefficient of Z and Weiner Filter DCT Coefficient denoted by $T(k, i)$. Here, for a known video resolution and block size, T can be pre-computed. The energy of all mediums can be added together whenever MB Z has multiple color mediums. Equation 4 can be applied for all macroblocks in a frame to evaluate the spatial salient features of that frame. To compute optimum saliency map the resultant map is normalized in the range $[0, 1]$. Therefore, to compute optimum saliency map the normalized block Z of spatial saliency can be obtained as:

$$D_d(Z) = \frac{1}{\sum_Z D_d^v(Z)} D_d^v(Z), \tag{5}$$

Here, for a known block and image dimensions, some coefficients of T may be considered as zero. Therefore, equation (4) represents un-normalized spatial saliency which is monotonically non-decreasing (either constant or increasing) for a total DCT energy of a block [26].

b. Global Motion-Compensated Temporal Saliency

The temporal saliency and spatial saliency both are different aspect of the video encoding field. Object motion is one of the most powerful and essential feature of video processing [27]. In many conventional techniques local motion contrast used to detect temporal saliency for visual attention [22].Consider an object with significant motion is measured as powerful and attention grabbing object to its surrounding in a visual system.

The performance of IKN saliency model degrades whenever camera motion comes in picture due to apparent motion of background participate with object motion of foreground and can easily confuse any saliency model [22, 28]. Therefore, to overcome this drawback, we eliminate camera motion itself before evaluating the saliency map. An efficient compressed-domain global motion prediction technique [29] used to compute saliency map effectively by using previous frame motion field as a present frame approximation. Then global motion compensation used to subtract global motion from the motion field. From this method we can evaluate 4×4 block for motion compensated global motion vector $(MCG - MV)$. For every macro-block Z the magnitude of $(MCG - MV)$ is represented as motion saliency $D_n(Z)$.

To get spatial-temporal saliency features of Z, we combine both spatial saliency and motion saliency of Z together with the help of coherent normalization technique based on fusion scheme [30],

$$D(Z) = (1 - \alpha)D_d(Z) + \alpha D_n(Z) + \beta D_d(Z)D_n(Z), \tag{6}$$

c. Macroblock QP Selection

Consider that the quantization parameter of the present video frame is $\mathbb{QP}_f$ which is calculated using a suitable rate control technique for frames. Assume that $D(Z_j)$ is the saliency of jth macroblocks Z_j and $\bar{d}$ is the average saliency of present frame for all macroblocks. For jth macroblock of present video frame, $\mathbb{QP}$ can be defined as.

$$\mathbb{QP}_j = round\left(\frac{\mathbb{QP}_\mathbb{f}}{\sqrt{\omega_j}}\right),$$ (7)

Where, a sigmoid function is used to compute ω_j and it is defined as,

$$\omega_j = c + \frac{g}{1+\exp(-l(D(Z_j)-\bar{d})/\bar{d})}$$ (8)

Where, a, b and c are the constants. In our model, we have set $a = 0.7, b = 0.6$ and $c = 4$. Equation (7), provides quantization parameter of Z_j. In our model, the relationship between quantization parameter $\mathbb{QP}$ and size of quantization step Q can be defined as,

$$Q = 2^{\mathbb{QP}/6} \cdot w(\mathbb{QP} \bmod 6),$$ (9)

Where w for different modes can be defined as $w(0) = 0.625, w(1) = 0.6875$ $w(2) = 0.8125$, $w(3) = 0.875$, $w(4) = 1.0$ and $w(5) = 1.125$.

In order to obtain optimum block coding mode we present a saliency distortion phase as $S_{sal}(\Psi|R_j, Z_j)$ to fulfill requirements effectively. In our model cost function can be defined as,

$$K_j\left(\Psi|R_j, \lambda_{D_j}, \lambda_{J_j}, Z_j\right) = S_{mse}(\Psi|R_j, Z_j) + \lambda_{D_j}S_{sal}(\Psi|R_j, Z_j) + \lambda_{J_j}J(\Psi|R_j, Z_j),$$ (10)

Where, saliency distortion $S_{sal}(\Psi|R_j, Z_j)$ can be linked with Lagrangian multiplier λ_{D_j} which is absolute change between jth uncompressed macroblock saliency and jth coded macro block with the help of coding mode Ψ using size of quantization step Q.

$$S_{sal}(\Psi|R_j, Z_j) = \left|D(Z_j) - D\left(\tilde{Z}_j(\Psi|R_j)\right)\right|,$$ (11)

Where, $\tilde{Z}_j(\Psi|R_j)$ represents jth coded macroblock with the help of coding mode Ψ and Z_j is the jth macroblock in uncompressed form with size of quantization step Q. Consider that compression can change the magnitude or direction of motion of different regions only for extreme lower bitrates. Therefore, we can say the difference in motion saliency is negligible in contrast to spatial saliency. Therefore, with the help of equation (6) we can approximated $S_{sal}(\Psi|R_j, Z_j)$ as,

$$S_{sal}(\Psi|R_j, Z_j) = \mu_j \cdot \left|D_d(Z_j) - D_d\left(\tilde{Z}_j(\Psi|R_j)\right)\right|$$ (12)

where,

$$\mu_j = 1 - \alpha + \beta D_n(Z_i).$$ (13)

Here, weighted macroblock motion saliency is used to compute macroblock saliency distortion which actually refers to spatial saliency distortion from equation (12) and (13). Therefore, the salient regions where saliency is very high, there distortion also will be high. Equation 12 computes the saliency distortion so that $D_d\left(\tilde{Z}_j(\Psi|R_j)\right)$ can easily computed which requires to eliminate the chicken and egg problem arises after compressed block saliency of a frame. In section 3.1, some principle conditions are mentioned which are adopted in our model where in highly salient regions saliency increase after compression. These conditions are categorized by,

$$Condition\ 1 = \begin{cases} Z_j \in ROI, & and, \\ D_d(Z_j) < D_d\left(\tilde{Z}_j(\Psi|R_j)\right), & and, \\ S_{mse}(\Psi|R_j, Z_j) < \delta, \end{cases}$$ (14)

Where, user-defined threshold δ can be set as $\delta = 21$ in our model, which is same as $PSNR = 35$ and it is assumed as good quality. Similarly, condition 2 tell that saliency can be decreased after compression of lower salient regions.

$$Condition\ 2 = \begin{cases} Z_j \in non - ROI, \qquad and, \\ D_d(Z_j) > D_d\left(\tilde{Z}_j(\Psi|R_j)\right). \end{cases} \tag{15}$$

Note that, reduction of saliency distortion with the help of equation (12) can help to save the saliency of lower salient regions also.

d.   Statistical Modeling of Transformed Residuals

Here, a Laplace probability density function of zero mean used to compute peripheral density of transformed residuals H with parameter λ.

$$\mathbb{f}_H(m; \lambda) = \frac{\lambda}{2}e^{-\lambda|m|}, \tag{16}$$

Where, the relationship between standard deviation σ_H and λ is represented by,

$$\lambda = \frac{\sqrt{2}}{\sigma_H}, \tag{17}$$

In our model, Z_j is the Laplacian random variable which consists of a $(k, i)^{th}$ transform residual coefficient with parameter λ_j^{ki}

$$\lambda_j^{ki} = \frac{\sqrt{2}}{\sigma_{H_j(k,i)}}. \tag{18}$$

According to equation (12), weighted macroblock motion saliency is used to compute macroblock saliency distortion which actually refers to spatial saliency distortion. The energy of Wiener-filtered DCT macro block is the approximation for spatial macroblock saliency. To predict spatial saliency block distortion through quantization method using a similar quantization noise [31].Assume that Wiener-filtered DCT energy for a quantization noise is our spatial saliency distortion. Therefore, final saliency map $S_{sal}(\Psi|R_j, Z_j)$ using our proposed method, which helps to get high compression and visual quality, can be defined as

$$S_{sal}(\Psi|R_j, Z_j) = \mu_j \sum_{(k,i)} T(k,i)\chi_j^{ki}, \tag{19}$$

Where, χ_j^{ki} can be defined as,

$$S_{mse}(\Psi|R_j, Z_j) = \sum_{(k,i)} \chi_j^{ki} = \sum_{(k,i)} \frac{e^{\lambda_j^{ki}\mathbb{F}_j\left(2w_j^{ki}+\left(w_j^{ki}\right)^2-2\mathbb{K}_j^{ki}\right)+2-2\Psi_j^{ki}}}{-\Phi_j^{ki}\left(\lambda_j^{ki}\right)^2} \tag{20}$$

4. Performance Evaluation

We evaluate our outcomes with the same dataset as used in [20, 32] to compare the proficiency and performance of our saliency model to the conventional techniques described in the related work. Our efficient saliency model trained on various large dataset like SFU dataset [20] and HEVC video_database [32].Here, we have presented experimental outcomes only for SFU dataset [20]. Testing outcomes demonstrates that our proposed model outperforms most of the conventional approaches in terms of PSNR, feature extraction, compression rate and the quality prediction of saliency map. We have tested our proposed model for various coding standards HEVC and H.264/AVC. Our experimental results demonstrates accuracy, speed (bit rate) and Compression ratio enhancement to a large extent. Our proposed model requires very less amount of execution time to achieve efficient video compression. Our proposed model implemented on 64-bit windows 10 OS with 16 GB RAM which consists on INTEL (R) core (TM)

i5-4460 processor. It consists of 3.20 GHz CPU. We have compared our model with Itti [21], Surprise [33], Judd [34], PQFT [35], Rudoy [36], Fang [37] and OBDL [38] existing techniques.

4.1. Implementation Details

We have implemented our wide experiments on large video SFU dataset [20] and HEVC video_database [32]. UHD videos becomes a common requirement in the field of entertainment, medical, photography, satellite imaging, HDTV, stereoscopic video processing, face recognition etc. due to its high quality visibility and true color. Therefore, there is a massive demand of high-resolution videos in video processing field. However, there is a massive problem of limited storage capacity and bandwidth spectrum. Therefore, there is a need of an effective compression technique without any data loss and which can precisely maintain the high video quality.

In this paper, we focus on an efficient modified video compression HEVC technique based on high quality assessment saliency features for the assessment of high quality videos. Here, to create an efficient saliency map we extract global temporal alignment component and robust spatial component. The small alterations between neighboring frames may be not sufficient to define salient regions. Hence, Motion estimation and detection is very critical phenomenon in our saliency model. We can determine saliency map precisely using HEVC (High Efficiency Video Encoding) in association with INTRA or INTER processing blocks with different sizes to get high compression rate. To obtain high quality saliency here, we combine spatial saliency features and temporal saliency features together for different macroblocks in association with transformed residuals. In this way, our saliency model outperforms all the existing techniques. To compute the performance of proposed modified HEVC model, 12 raw videos from the testing SFU dataset [20] has been taken out of which all 12 videos are used in testing of PSNR computation. All the experiments undertaken on the MATLAB 2016b framework.

4.2. Comparative Study

In this paper, we have compared our experimental results with many state-of-the-art techniques such as $Itti$ [21], Surprise [33], $Judd$ [34], PQFT [35], $Rudoy$ [36], Fang [37] and OBDL [38] existing techniques. In this paper, all the raw videos sampled on YUV 4:2:0 sampling. All the videos are compressed to high quality (more than $30\ dB$). This videos includes contents such as sport events, video conferencing, surveillance, video games etc. All 12 raw videos in the SFU dataset are of 352×288 resolution with different frames.

Table 1. PSNR Comparison Results for Y, U, and V Channel Using Our Proposed and Existing Method for HM 36.0 Considering SFU Dataset

VIDEO NAME	HEVC PSNR [19] (dB)					PROPOSED PSNR (dB)
	Y	U	V	Y	U	V
BUS	27.8354	37.7225	39.3719	29.768	37.793	39.048
CITY	29.0913	40.1978	41.5116	29.798	39.820	41.118
TEMPETE	31.0809	37.4756	35.5797	30.504	30.385	30.384
STEFAN	28.3679	35.1228	35.1513	30.135	34.901	34.729
SOCCER	30.4608	39.3976	41.2553	31.898	38.785	40.911
FOREMAN	30.7809	38.3184	39.5504	31.742	38.006	39.021
CREW	31.0809	37.4756	35.5797	32.752	37.394	36.073
HALL	28.176	34.575	36.775	33.147	37.453	39.543
HARBOUR	27.3567	38.5148	40.6429	28.583	38.610	40.039
GARDEN	26.5463	32.6592	35.0204	28.528	32.815	34.795
MOBILE	26.7781	32.7365	32.2783	27.729	31.524	31.241
MOTHER	33.8238	41.2334	42.2492	34.288	40.370	41.338
AVERAGE	29.28158	37.1191	37.9138	30.739	36.488	37.35

Table 2. Frame Reconstruction Quality Comparison Results between Our Proposed Method and Existing Method

Input Image	Image reconstruction from existing HEVC technique	Image reconstruction from our proposed technique

Here, in Table 2 we have presented reconstruction frames from the input frames using our proposed and existing HEVC technique for 4 videos such as Bus, Stefan, Soccer and Foreman out of 12 videos. Here, we have used different frames for all 4 videos such as frame 12 for Bus, frame 22 for Stefan, frame 7 for foreman and Soccer. It is clearly visible from table 2 that the reconstruction frames are of higher quality using our proposed modified HEVC method compare to existing HEVC method.

Table 1 and 2 shows that PSNR (Peak Signal to Noise Ratio) and video frame reconstruction quality comparison results using our proposed method for coding standard HM 36.0 respectively considering SFU dataset .The PSNR values are very high (more than $30\ dB$) using our proposed method (for all three Y, U and V channels) compare to existing techniques. Similarly, frame reconstruction quality is very high using our proposed method compare to existing methods for all three Y, U and V channel which shows high quality of our reconstructed videos. Overall PSNR and bitrate values are very effective to get high quality reconstructed compressed video. The average PSNR considering Y-channel is 30.739, U-channel is 36.488 and V-channel is 37.35 using our proposed method which are very higher compare to the existing techniques. Similarly, the frame reconstruction quality which is very high compare to the existing techniques and shows high quality of reconstructed video. However, quality for each video differ according to their frame rate and dimensions of video. This result demonstrates our proposed method dominance towards existing state-of-the-art techniques in terms of PSNR and reconstruction quality and provides a high quality saliency map to obtain high compression rate.

Experimental results demonstrate that our proposed techniques completely outperforms existing techniques in terms of PSNR and reconstruction quality. This results signifies our proposed method produces high quality reconstructed saliency map to get high quality compression. Here, table 3 presents the comparison of performance matrices with existing techniques for all 12 videos in SFU dataset with our Proposed System. Experimental results demonstrates that our proposed model also dominates the other existing state-of-the-art techniques in terms of AUC (Area under Curve) and NSS (Normalized Scan-path Saliency). The average AUC for all 12 videos using our proposed method is 0.846 and average NSS for all 12 videos using our proposed technique is 1.478. This results assures high quality reconstructed saliency map after efficient compression. Figure 1 shows the graphical comparison of our proposed method and existing techniques in terms of performance matrices AUC and NSS.

Table 3. Comparison of Performance Matrices with Existing Techniques for All Videos in SFU Dataset with our Proposed System

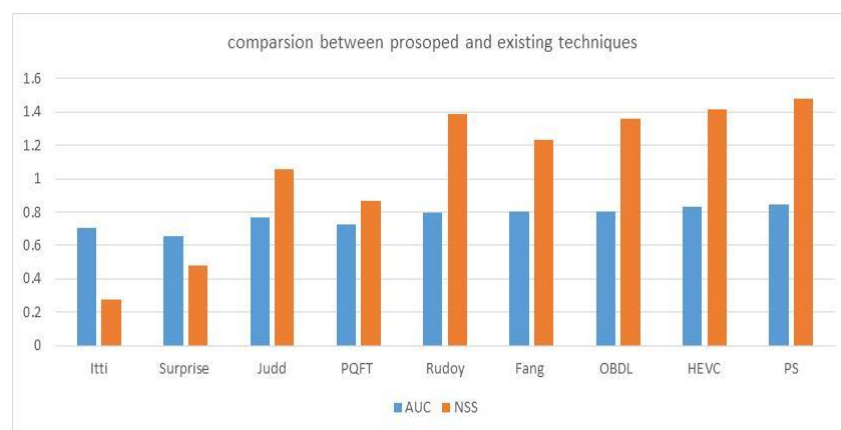| PARAMETER	$Itti$ [24]	Surprise [25]	$Judd$ [26]	PQFT [27]	$Rudoy$ [28]	Fang [29]	OBDL [30]	HEVC [19]	PS
AUC	0.705	0.658	0.770	0.729	0.799	0.80	0.802	0.832	0.846
NSS	0.278	0.479	1.058	0.867	1.388	1.23	1.361	1.415	1.478

Figure 1. comparison of our proposed model with existing state-of-the art-techniques

5.   Conclusion

The significance and complexities in designing real time modified video compression HEVC technique is discussed. Drawbacks of current existing systems is presented. Here we have presented an efficient modified video compression HEVC technique based on high quality assessment saliency features for the assessment of high quality videos. To create an efficient

saliency map we extract global temporal alignment component and robust spatial component. To obtain high quality saliency here, we combine spatial saliency features and temporal saliency features together for different macroblocks in association with transformed residuals. In this way, our saliency model outperforms all the existing techniques. In this paper, we have generated high quality video reconstruction after compression considering SFU dataset. Here, we have shown average AUC and NSS comparison with other existing techniques for all 12 videos in table 3. The average AUC and NSS for all 12 videos using our proposed method is 0.846 and 1.478 for SFU dataset respectively. The average PSNR for Y, U and V channel using our proposed method is 30.739, 36.488 and 37.35 respectively presented in table 1. Similarly, the frame reconstruction quality using our proposed method for all 12 videos is very high compare to existing techniques and high resolution quality of 4 videos is presented in Table 2. These results verify that our model outperforms any other state-of-the-art-techniques. In future, this model can be used in the field of medical, photography, satellite imaging, HDTV, stereoscopic video processing, face recognition and video coding or encoding.

References

[1] IEG. Richardson. H. 264 and MPEG-4 Video Compression: Video Coding for Next Generation Multimedia. Wiley, UK, 2003.
[2] ISO/IEC 13818-2 and ITU-T Rec. H.262, Generic Coding of Moving Pictures and Associated Audio Information–Part 2: Video. (MPEG-2). November
[3] Khoo Zhi Yion, Ab Al-Hadi Ab Rahman. Exploring the Design Space of HEVC Inverse Transforms with Dataflow Programming. *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS).* April 2017; 6(1): 104-109. DOI: 10.11591/ijeecs.v6.i1.pp104-109.
[4] M Vijayalakshmi, L Kulkarni. Quality aware protocol to support multimedia data delivery in wireless network. 2016 3rd International Conference on Devices, Circuits and Systems (ICDCS), Coimbatore, 2016: 322-326.
[5] Jignesh Patel, Haresh Suthar, Jagrut Gadit. VHDL Implementation of H.264 Video Coding Standard. *International Journal of Reconfigurable and Embedded Systems (IJRES)*, November 2012; 1(3): 95-102. ISSN: 2089-4864
[6] D Indoonundon, TP Fowdur, KM S Soyjaudah. A Concealment Aware UEP Scheme for H.264 using RS Codes, *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS).* June 2017; 6(3): 671-681. DOI: 10.11591/ijeecs.v6.i3.pp671-681.
[7] M Jeon, N Kim, BD Lee. MapReduce-Based Distributed Video Encoding Using Content-Aware Video Segmentation and Scheduling. in IEEE Access. 2016; 4: 6802-6815,. doi: 10.1109/ACCESS.2016.2616540
[8] J Chao, E Steinbach. Keypoint Encoding for Improved Feature Extraction from Compressed Video at Low Bitrates. in IEEE Transactions on Multimedia. Jan. 2016; 18(1): 25-39. doi: 10.1109/TMM.2015.2502552
[9] A Ilic, S Momcilovic, N Roma, L. Sousa. Adaptive Scheduling Framework for Real-Time Video Encoding on Heterogeneous Systems. in IEEE Transactions on Circuits and Systems for Video Technology. March 2016; 26(3): 597-611. doi:10.1109/TCSVT.2015.2402893
[10] Hao Fusheng, Huang Jijiang, Liu Wei, Wang Yanan, Yang Hongtao, Cao Jianzhong. *Real time H.264 High Definition Videos Encoding Based on TMS320DM368 and a Video Quality Evaluation Framework.* 2016 IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), Xi'an, 2016: 128-132.
[11] P. K. Podder, M. Paul and M. Murshed, "QMET: A new quality assessment metric for no-reference video coding by using human eye traversal," 2016 International Conference on Image and Vision Computing New Zealand (IVCNZ), Palmerston North, 2016, pp. 1-6.
[12] H. R. Tohidypour, M. T. Pourazad and P. Nasiopoulos, "Online-Learning-Based Complexity Reduction Scheme for 3D-HEVC," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 26, no. 10, pp. 1870-1883, Oct. 2016.
[13] T. K. Tan et al., "Video Quality Evaluation Methodology and Verification Testing of HEVC Compression Performance," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 26, no. 1, pp. 76-90, Jan. 2016.
[14] J. Wang, S. Wang and Z. Wang, "Asymmetrically Compressed Stereoscopic 3D Videos: Quality Assessment and Rate-Distortion Performance Evaluation," in IEEE Transactions on Image Processing, vol. 26, no. 3, pp. 1330-1343, March 2017.
[15] Z. Duanmu, K. Zeng, K. Ma, A. Rehman and Z. Wang, "A Quality-of-Experience Index for Streaming Video," in IEEE Journal of Selected Topics in Signal Processing, vol. 11, no. 1, pp. 154-166, Feb. 2017.

[16] S. Xiang, L. Yu and C. W. Chen, "No-Reference Depth Assessment Based on Edge Misalignment Errors for T + D Images," in IEEE Transactions on Image Processing, vol. 25, no. 3, pp. 1479-1494, March 2016.

[17] M. S. Gide and L. J. Karam, "A Locally Weighted Fixation Density-Based Metric for Assessing the Quality of Visual Saliency Predictions," in IEEE Transactions on Image Processing, vol. 25, no. 8, pp. 3852-3861, Aug. 2016.

[18] S. S. Naqvi, W. N. Browne and C. Hollitt, "Feature Quality-Based Dynamic Feature Selection for Improving Salient Object Detection," in IEEE Transactions on Image Processing, vol. 25, no. 9, pp. 4298-4313, Sept. 2016.

[19] H. Liu, M. Xu, J. Wang, T. Rao and I. Burnett, "Improving Visual Saliency Computing With Emotion Intensity," in IEEE Transactions on Neural Networks and Learning Systems, vol. 27, no. 6, pp. 1201-1213, June 2016.

[20] Hadizadeh H, Enriquez MJ, Baji´c IV (2012) Eye-tracking database for a set of standard video sequences. IEEE Trans Image Process 21(2):898–903

[21] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," IEEE Trans. Pattern Anal. Mach. Intell., vol. 20, no. 11, pp. 1254–1259, Nov. 1998.

[22] L. Itti, "Automatic foveation for video compression using a neurobiological model of visual attention," IEEE Trans. Image Process., vol. 13, no. 10, pp. 1304–1318, Oct. 2004.

[23] I. E. Richardson, The H.264 Advanced Video Compression Standard New York, NY, USA: Wiley, 2010.

[24] T. Wiegand and B. Girod, "Lagrange multiplier selection in hybrid video coder control," in Proc. IEEE Int. Conf. Image Process., vol. 3.Oct. 2001, pp. 542–545.

[25] M. A. Robertson and R. L. Stevenson, "DCT quantization noise in compressed images," IEEE Trans. Circuits Syst. Video Technol., vol. 15,no. 1, pp. 27–38, Jan. 2005.

[26] H. Hadizadeh, "Visual saliency in video compression and transmission,"Ph.D. dissertation, School Eng. Sci., Simon Fraser Univ., Apr. 2013.

[27] L. Itti, "Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes," Vis. Cognit., vol. 12, no. 6, pp. 1093–1123, 2005.

[28] V. A. Mateescu, H. Hadizadeh, and I. V. Baji´c, "Evaluation of several visual saliency models in terms of gaze prediction accuracy on video,"in Proc. IEEE Globecom Workshops, Dec. 2012, pp. 1304–1308.

[29] Y.-M. Chen and I. V. Baji´c, "Motion vector outlier rejection cascade for global motion estimation," IEEE Signal Process. Lett. vol. 17, no. 2, pp. 197–200, Feb. 2010.

[30] Y. Fang, Z. Chen, W. Lin, and C.-W. Lin, "Saliency detection in the compressed domain for adaptive image retargeting," IEEE Trans. Image Process., vol. 9, no. 21, pp. 3888–3901, Sep. 2012.

[31] B. Widrow and I. Kollar, Quantization Noise: Roundoff Error in Digital Computation, Signal Processing, Control, and Communications.Cambridge, U.K.: Cambridge Univ. Press, 2008.

[32] M. Xu, L. Jiang, X. Sun, Z. Ye and Z. Wang, "Learning to Detect Video Saliency With HEVC Features," in IEEE Transactions on Image Processing, vol. 26, no. 1, pp. 369-385, Jan. 2017. doi: 10.1109/TIP.2016.2628583

[33] L. Itti and P. Baldi, "Bayesian surprise attracts human attention," Vis. Res., vol. 49, no. 10, pp. 1295–1306, Jun. 2009.

[34] T. Judd, K. Ehinger, F. Durand, and A. Torralba, "Learning to predict where humans look," in Proc. ICCV, Sep./Oct. 2009, pp. 2106–2113.

[35] C. Guo and L. Zhang, "A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression," IEEE Trans. Image Process., vol. 19, no. 1, pp. 185–198, Jan. 2010.

[36] D. Rudoy, D. B. Goldman, E. Shechtman, and L. Zelnik-Manor, "Learning video saliency from human gaze using candidate selection," in Proc. CVPR, Jun. 2013, pp. 1147–1154.

[37] Y. Fang, W. Lin, Z. Chen, C.-M. Tsai, and C.-W. Lin, "A video saliency detection model in compressed domain," IEEE Trans. Circuits Syst.Video Technol., vol. 24, no. 1, pp. 27–38, Jan. 2014.

[38] S. Hossein Khatoonabadi, N. Vasconcelos, I. V. Bajic, and Y. Shan, "How many bits does it take for a stimulus to be salient?" in Proc.CVPR, Jun. 2015, pp. 5501–5510.