

High Definition Video Compression Usin Saliency Features

K. Bhanu Rekha^{*1}, Ravi Kumar AV²

¹Department of ECE, Sir M.V.I.T, Bangalore, India

²Department of ECE, SJBIT, Bangalore, India

*Corresponding author, e-mail: kbhanu.rekha@rediffmail.com

Abstract

High Definition (HD) devices requires HD-videos for the effective uses of HD devices. However, it consists of some issues such as high storage capacity, limited battery power of high definition devices, long encoding time, and high computational complexity when it comes to the transmission, broadcasting and internet traffic. Many existing techniques consists these above-mentioned issues. Therefore, there is a need of an efficient technique, which reduces unnecessary amount of space, provides high compression rate and requires low bandwidth spectrum. Therefore, in the paper we have introduced an efficient video compression technique as modified HEVC coding based on saliency features to counter these existing drawbacks. We highlight first, on extracting features on the raw data and then compressed it largely. This technique makes our model powerful and provides effective performance in terms of compression. Our experiment results proves that our model provide better efficiency in terms of average PSNR, MSE and bitrate. Our experimental results outperforms all the existing techniques in terms of saliency map detection, AUC, NSS, KLD and JSD. The average AUC, NSS and KLD value by our proposed method are 0.846, 1.702 and 0.532 respectively which is very high compare to other existing technique.

Keywords: HEVC, AUC, NSS, Encoding

Copyright © 2017 Institute of Advanced Engineering and Science. All rights reserved.

1. Introduction

21st century has provided the practical world enormous innovative electronic inventions, which takes entire electronic research industry to a completely different level. Out of these outstanding achievements by the researchers, one quite popular and trending topic in today's market is High definition (HD) devices such as Action cameras, Ultra Edge high definition cameras and Ultra Edge TV etc. These high-definition (HD) devices becomes the mandatory practical requirement in the field of medical imaging, satellite imaging, face recognition systems, military, entertainment(for advertisements and movies),animation 3D view and images motion detection etc. For the vulnerable use of these high definition devices, there is need of high-resolution pictures and videos.

In recent years, the demand of high resolution videos drastically increased due to immense popularity of high definition devices such as Smart phones, iPad, MacBook Pro, LAPTOPS, HDTV (High-definition television), UHD TV (Ultra-high-definition television) and 3-D TV systems. Therefore market is highly dominated by the requirement of high-resolution videos in the field of medical, photography, satellite imaging, HDTV, stereoscopic video processing, face recognition and video coding or encoding in recent few years.

However, it comes with some eye-catching issues such as high storage capacity, limited battery power of high definition devices, long encoding time, large bitrate degradation, and high computational complexity. This led to high bandwidth demand and need of large data storage in internet traffic [1]. Therefore, to minimize the memory and large bandwidth spectrum issue video encoding is a vital phenomenon. There are many algorithms presented in recent few years in the field of video encoding, as it is a hot topic from the prospect of industry and fast growing electronic market.

The encoding efficiency of high-resolution videos is quite low using earlier techniques. Therefore, to counter the problems the latest video encoding standard techniques H.264/AVC [2] and HEVC/H.265 [3] developed to reduce the size of high-resolution videos largely by producing large compression rates while maintaining the quality of videos. However, in real time

it is very difficult to achieve high compression rates due to drastic increase in computational complexities. Another issue related to limited battery lifetime as to capture high-resolution videos, high definition battery-powered cameras required in the field of sports, travel, military, rescue, surveillance, entertainment, daily life and broadcasting high-definition videos. This issue can further led to the encoding bit-rate and distortion trade-off issue [4-12].

Therefore, to counter these issues in global interest many researchers has carried out some great amount of work in video encoding field. A very popular and fast growing technique in recent time for video encoding is H.264/AVC (Advanced Video Encoding) standards. This standard approach used to increase the compression rate for high-resolution videos. However, the computational complexity is very high for this standard, which causes its degradation in performance [2]. Another biggest disappointment with this technique is its slowness (takes more computational time) [13]. Map Reduce based distributed algorithm [13] used to speed-up the encoding process. However, it requires large amount of bandwidth to encode high definition videos. A one more technique, Adaptive Scheduling Framework for Real-Time Video Encoding presented in [14] to reduce the computational complexity. However, it requires parallelization at inter-loop level, which is very hard to achieve.

There is a huge demand of High Definition videos in the market due to availability of HD devices. However, there are many issues occur while transmitting, storing and live streaming of HD videos. Therefore, there is a need of efficient video encoding without degradation in video quality. However, it is very complex process to achieve efficient video encoding as it consists of many issues such as limited battery power of high definition devices, long encoding time, large bitrate degradation, limited bandwidth and high computational complexity problems. Although, a healthy research has been done to counter these drawbacks and generates a high quality video encoding. However, still it lacks an effective application to remove these drawbacks. Therefore, this motivates us to present an efficient technique, which can handle most of these issues effectively. Therefore, we have proposed a video compression technique modified HEVC coding based on a saliency features.

In many popular existing techniques, they first compress the video dataset and then extract their features. Therefore, the compression rate become very low as the data is already compressed which also reduces the performance of encoding process. Therefore, to eliminate this drawback and make the encoding performance better we first extract HEVC features and then compress the raw data to very low size. This phenomenon provides better compression, required bit-rate and faster compare to other algorithms.

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

However, very few techniques emerges as a practical solutions to above mentioned, limited battery power of high definition devices, long encoding time, large bitrate degradation, limited bandwidth and high computational complexity problems. Still, there is a need of some good research, which can efficiently target these eye-catching issues as they directly affects the performance of the encoding process. A brief of related work in the field of video encoding presented in the following section.

A Novel Bitrate-Saving and Fast Coding technique for depth video in 3D HEVC presented in [15] to reduce synthesis error, which produced while processing of depth, and color map encoding. This technique can be very effective to get 3-D view of high definition (HD) videos. However, it degrades the bitrate largely, which is quite difficult to achieve back completely. In [13], Segmentation and Scheduling video encoding based on distributed Map Reduce architecture presented to speed up the encoding process by using advanced distributed technique. In this paper, video segmentation technique used to provide better efficiency and to change the encoding segment order an efficient scheduling scheme adopted. However, it generates high bitrate and reduces the efficiency of the content differentiation process, which can led to performance degrade in encoding process. In [16], a Video Encoding technique based on Content-Aware approach presented to provide a consistent quality of encoded video and required bit-rate to satisfy bandwidth necessities. The main idea behind this is to reduce

computational complexity by searching frame wise optimal quantization factors, which rely upon content-based features and single pass encoding method to analysis distortion prototype. However, it requires high computational speed, high storage capacity and modern machinery with multicore design to handle large computations. In [14], a Real-Time Video Encoding for Heterogeneous Systems adopted based on Adaptive Scheduling Framework to provide real-time encoding by using a combination of CPU+GPU cores parallel for heterogeneous devices. However, heterogeneous devices consists of unified optimization problem in video encoding and provides high computation complexity. In [17], Key point Encoding for Improved Feature Extraction from Compressed Video at Low Bitrates performed to detect key-points and descriptor calculation, which helps in detecting different variations, occurs in video encoding, to lower bitrates. In this paper, they have reduced bitrate to a sufficient amount for a single scene of a video. However, it can be very complex process for multiple scenes of a video and bitrate can be very high to encode multiple scenes of a video. In [18], HEVC Coding presented based on Fast Intra Prediction to reduce the computational complexity and bitrate while video encoding. In this paper, Screen content coding (SCC) used to provide low latency and fast transmission with HEVC technique. However, it consists of larger prediction units, which can cause very high complexity, and the motion estimation of a video can be an expensive process. In addition, it requires high storage capacity and real camera capture contents to produce efficient outcomes in terms of bitrate and computational complexity. In [19], HEVC Live Video Encoding presented to reduce high computational complexity produced in HEVC technique and maintain quality of video encoding after reducing computational complexities. However, transmission bandwidth and delay consideration should be minimum (super-fast) for live video encoding using HEVC, which is very hard to achieve. Therefore, it can concluded from the above literature in the encoding field that still a healthy amount of work required countering all the video encoding issues effectively.

In this section, many algorithms presented as a related work. However, each algorithm has its own issues. The basic issues are high storage capacity, limited battery power of high definition devices, long encoding time, large bitrate degradation, and high computational complexity. Therefore, to overcome the issues we have proposed here video compression technique modified HEVC coding based on a saliency features. In this model, first we extract features on the raw data and then compress the raw data largely. This technique can increase the compression rate, takes less time to execute, and requires low computational complexity. This factor makes our proposed model more efficient compare to other existing techniques.

3. Proposed Methodology

3.1 Video Encoding based on Saliency Features

In recent years, the demand of high-definition videos has taken drastic growth in real practical world due to the presence of high-definition devices. However, high-definition videos takes large space and bandwidth spectrum. Therefore, there is a need of an efficient video encoding which can maintain quality of a video without any data loss. Therefore, we have implemented a video compression technique based on a saliency features. This technique can used in field of medical, photography, satellite imaging, HDTV, stereoscopic video processing, face recognition and video coding or encoding to estimate saliency and compress high-definition videos largely. Our video compression provides fast computation for large training database such as SFU dataset [18] and HEVC video_database [19]. There are multiple factors, which makes HEVC architecture efficient and helps in enhancing the performance of the system.

- a) This architecture helps to achieve lower bitrate, which helps in enhancing the performance of the encoding process.
- b) It can effectively estimate video saliency by extracting features efficiently.
- c) It takes less time while compression of large datasets such as SFU dataset [18] and HEVC video_database [19].

3.2 Video Compression Using Saliency Features

There are very few techniques, which can effectively estimate saliency and compress a high definition video largely without any loss in the quality. Therefore, to detect saliency

efficiently and to provide high quality video after compression, we have implemented a video compression technique as modified HEVC based on a saliency features.

3.3 Compressed Domain Features

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. In recent years, many techniques evolved which tried to compress videos effectively without any loss of quality. Therefore, these factors becomes more advance with the evolution of new techniques. The outcomes of these factors provides compression domain features. These compressed domain features can work efficiently with the H.264/AVC standard [20] and HEVC (High Efficiency Video Coding) [21]. Our methodology involves 16×16 blocks, INTER, SKIP or INTRA processing blocks for different sizes (4×4 , 8×8) and approximation of Transform for HEVC coding standard.

3.3.1 Motion vector entropy (MVE)

In video compression field, Motion Vector (MV) plays a very important role related to variations in the scene. It is a two-dimensional block vector (D_a, D_b) which balances the effect of best-matching template of a reference frame. Motion detection helps to find best-matching template. The approximation of Optical Flow provides a Motion Vector field.

Many different type of Motion Vectors are generated while video compression. Many different MVs created, when a moving object passed to a definite section in the video scene. These MVs generated in the corresponding spatial-temporal adjacent region. Some MVs represent moving object itself and some represents the background environment, out of these generated MVs. The Motion Vector of object can differ from each other. The Motion Vector of the background can cover entire section of the scene and provides consistent MVs. It derived from motion of camera. This change in MVs in the spatial-temporal adjacent region can utilized to recognize the existence of moving objects.

The processing of blocks inside a frame described as follows. This process carried out only for INTER blocks. However, SKIP blocks contains a zero Motion vector. All the Motion Vector (MVs) inside a frame mapped into 4×4 blocks. For example, we consider a 8×8 block for which a MV assigned. This MV can be distributed to all four constitutes blocks of size 4×4 .

The motion vector entropy can described by the equation 1 as follows,

$$mve(x) = -\log M^{-1} \sum_{j=k(\theta(x))} m_j \cdot M^{-1} \cdot \log m_j \cdot M^{-1} \quad (1)$$

Here, x represent a block and $\theta(x)$ represents a motion cube linked with 4×4 block, m_j is the number of Motion Vectors present in the bin index j . The parameter $\log M^{-1}$ in (1) lies between 0 (min) and 1 (max).

$$\log M^{-1} = \begin{cases} 1 & \text{max; } m_j = m_k, \forall j, k \\ 0 & \text{min; } m_j = 0, \forall j \neq 1 \end{cases} \quad (2)$$

Background of a video scene can designed by large blocks and smaller blocks used to represent the moving objects with the help of motion estimation [22].

$$S_m = F_d \odot F_n \quad (3)$$

Here, Equation (1) represents the element wise multiplication of motion magnitude F_d and global angle F_n . The matrix for motion magnitude derived from a constant duration of a scene. Here, motion vectors used to construct saliency map. For every motion vector, a temporal-spatial motion magnitude calculated.

$$F_d(m) = \begin{cases} \frac{z}{\mathcal{J}} & z \leq \mathcal{J} \\ 1 & z < \mathcal{J} \end{cases} \quad (4)$$

Here, Equation (2), represents the normalization to filter the inline motion vectors (MVs). In equation (2), z is the motion magnitude after filtering of block m . \mathcal{T} Represents the threshold. When F_d becomes 1, then it shows that the corresponding blocks consists large motion magnitude.

$$F_n(m) = \frac{-\mathcal{P}(\varphi) \cdot \log \mathcal{P}(\varphi)}{\log m} \quad (5)$$

Here, Equation (3) represents the normalized entropy to calculate the coherence global angle for motion vector. Here, $\mathcal{P}(\varphi)$ represents the normalized histograms of motion vectors and m denotes the number of histograms presents. To keep normalized histogram range in between 0 and 1, $\log m$ used in denominator. If the normalized histograms of motion vector is 0 then it shows that camera motion is very high and if it is 1 then it shows motion inconsistency.

3.3.2 Smooth Residual Feature

The motion-estimation residual detects that the current block is not same as the best-matching block in the reference frame. This shows that the block translation process cannot estimate motion of current block. This is due to either occlusions or large motion. However, most often it occurs due to occlusions. Occlusion can occur when suddenly any moving object comes in between the scene or hide behind the object. Therefore, large residuals can detect the sections where occlusions can occur.

The size of residual depends on the normalization function n_0 which is a non-zero variable. Residual Normalization feature can expressed in Equation (3) as,

$$RF(N) = 256^{-1} \| N \|_0 \quad (6)$$

Here, N represent the residual transformation of 16×16 macro block. Then, the spatial smoothness of residual normalization feature map calculated using 3×3 filter and temporal smoothness obtained by utilizing average moving filter over total existing frames. Therefore, a SRF (Smooth Residual Feature) map generated using 4×4 block.

3.3.3 Proposed Saliency Estimation

Here, we explain that how to estimate saliency using our proposed method. The detection of saliency can done without reconstructing the full video. This can verified by identifying two visual correlates of video fixation features for compressed territory. These two visual correlates are Motion Vector Entropy (MVE) and Smooth Residual Feature (SMF). Here, Figure 3.1 demonstrates the block diagram to compute and estimate saliency. For every frame, Motion Vectors (MV) and SRF (Smooth Residual Feature) and BCM (Block Coding Maps) generated. Here, *RaceHorseD* video data taken to obtain saliency map from HEVC video_database [19], [23, 24]. Both the features are necessary to obtain a saliency map of an input video. In fig 3.1 all, the processing stages described to find saliency map.

To detect saliency map, many feature extraction approaches examined [25, 26]. There are two factors which decides the accuracy of feature extraction process such as feature independency and behavior of their mutual activities which can affect the saliency map generation. MVE AND SRF values does not corresponds to each other i.e. one can be high and other can be low or vise-versa. The combination of both the features MVE and SRF can be very crucial which can give a better saliency map. If both the features are having huge values then there is very high possibility of moving objects in that region as well as it can contain sudden objects. Therefore, our model consists of a combination of both MVE and SRN features. Equation (7) represent the saliency map estimation.

$$SM = \eta (mve + RF(N) + mve \odot RF(N)) \quad (7)$$

Here, \odot represents the pointwise multiplication and $\eta(\cdot)$ represents the norm of range [0, 1]. This proposed method is a combination of MVE and SRF to generate and estimate saliency map.

Figure 1 represent the block diagram to find saliency map using the combination of MVE and SRF feature. All the processing can done on HEVC coding standard. Most of the

popular existing techniques, first they extract HEVC (High Efficiency Video Coding) feature then they compress the compressed outcome of HEVC feature. Therefore, there is very less area remain to further, compress on already compressed data. It also takes more time to extract features and encode the huge size data. These led to presence of high computational complexity. Therefore, to eliminate these drawbacks here we have first compressed our input raw video, then we extract features using modified HEVC coding standard on the processed input data. These phenomenon's increase our encoding performances by saving computational time and reducing computational complexity. Due to these factors, the compression rate will become very high. Therefore, our proposed model becomes more efficient than the existing techniques which verified by our experimental results shown in following sections.

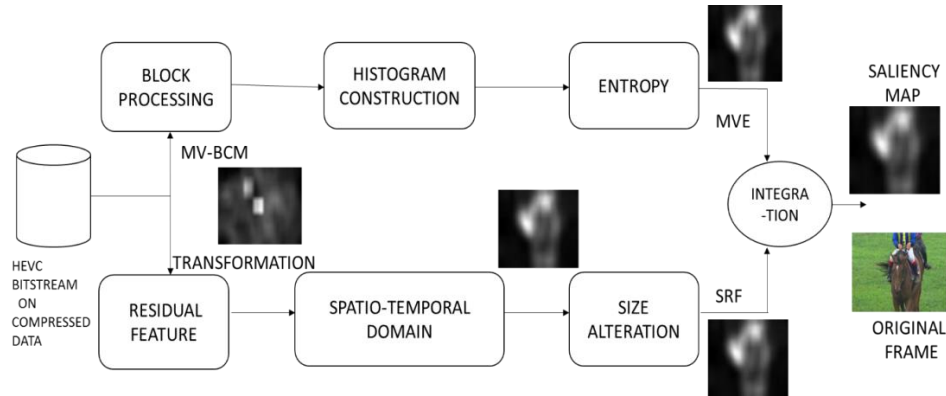


Figure 1. Block Diagram to Find Saliency Map using HEVC standard on compressed data

3.4 Accuracy Evaluation

There are many parameters, which used to calculate the accuracy of visual saliency maps. Some of the parameters are AUC (Area under Curve), LCC (Linear Correlation Coefficient), JSD (Jensen-Shannon divergence), KLD (Kullback -Leibler divergence) and NSS (Normalized scan path saliency) in terms of gaze point information [27-29]. All these parameters has its own importance with respect to evolution of performance. A prototype, which provides, better scores for all these mentioned parameters considered as an accurate prototype.

3.4.1 Area Under Curve (AUC)

AUC represents the properties of ROC (Receiver Operating Characteristic), more accurately, area under ROC curve, which can evaluated from the TPR (True Positive Rate) and FPR (False Positive Rate) at different threshold values. This parameter used to estimate performance and saliency map using gaze points. The AUC lies between 0 and 1. The lower value of AUC represents that the weaker saliency prediction and the higher value represents the better correspondence of saliency map.

3.4.2 Kullback -Leibler Divergence (KLD)

The parameter KLD used to find the divergence of two probability functions. It can defined as the relative entropy from one to another distribution. The higher values of KLD and JD can predict better saliency map or gaze points to evaluate accuracy.

$$kld(\mathcal{P} \parallel \mathcal{Q}) = \sum_k^l \mathcal{P}(k) \cdot \log c \frac{(\mathcal{P}(k))}{(\mathcal{Q}(k))} \quad (8)$$

$$jd(\mathcal{P} \parallel \mathcal{Q}) = kld(\mathcal{P} \parallel \mathcal{Q}) + kld(\mathcal{Q} \parallel \mathcal{P}) \quad (9)$$

3.4.3 JSD (Jensen-Shannon Divergence)

Jensen-Shannon divergence (JSD) is an upgraded version of KLD (Kullback -Leibler divergence) to remove the drawbacks of KLD AND JD. Its value ranges from 0 to 1. Equation (10) represents the Jensen-Shannon divergence to evaluate performance using Q and \mathcal{P} probability distributions.

$$jd(\mathcal{P} \parallel Q) = \frac{kld(\mathcal{P} \parallel \mathcal{R}) + kld(Q \parallel \mathcal{R})}{2} \quad (10)$$

$$\mathcal{R} = \frac{\mathcal{P} + Q}{2}$$

3.4.4 Normalized scan path saliency (NSS)

Normalized scan path saliency (NSS) used to find the average normalized saliency values at gaze positions. NSS consider parallelism to get standard deviation as 1 and mean as 0. A positive value of normalized saliency at a definite gaze point shows the similarity of gaze points with the estimated salient fields while negative value of normalized saliency shows that the gaze points placed into a non-salient region. There is no relation between estimation and gaze points if value becomes zero.

3.4.5 Linear correlation coefficient (LCC)

Linear Correlation Coefficient used to evaluate the linear relationship between an estimated saliency and ground truth. The ground truth is defined as the convolution of gaze points and Gaussian functions with SD=1. The linear correlation coefficient can evaluated as,

$$corr(G, \mathcal{S}_m) = \frac{conv(G, \mathcal{S}_m)}{\sigma_G \sigma_{\mathcal{S}_m}} \quad (11)$$

Here, $conv(G, \mathcal{S}_m)$ represents covariance of σ_G and $\sigma_{\mathcal{S}_m}$. LCC ranges from -1 to 1. In LCC, 0 shows no correction whereas ± 1 shows strong bonding.

4. Result and Analysis

We compute our outcomes with the similar dataset as used in [18-19] to compare the performance and efficiency of our model to the existing techniques discussed in the related work. Our model trained on different large dataset like SFU dataset [18] and HEVC video_database [19], [23, 24]. Here, we have shown results only for HEVC video_database. Testing results shows that our model outperforms most of the existing techniques in terms of PSNR, feature extraction, compression rate and the estimation of saliency map. We have tested our model for different coding standards HEVC and H.264/AVC. Our results shows accuracy, speed (bit rate) and Compression ratio enhancement largely. Our model needs less amount of execution time to provide effective video compression. Our 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* [26], *Surprise* [30], *Judd* [31], *PQFT* [32], *Rudoy*[33], *Fang* [34] and *OBDL* [35] existing techniques.

4.1 Implementation Details

We have implemented our extensive experiments on large video SFU dataset [18] and HEVC video_database [19], [23, 24]. In modern era, the availability of 4K monitors is highly increased. Therefore, there is a huge demand of low-resolution videos to high-resolution videos in market. However, there is a huge problem of limited storage capacity and bandwidth spectrum. Therefore, there is need of efficient compression without loss of any kind of data and maintaining the high quality of the video. In this paper, we have first compressed the raw videos to large extent and then we extract HEVC features to get high quality of a video after effective compression. This technique saves large computational time and reduces high computational complexities. To compute the performance of proposed modified HEVC model, here we have

used 33 raw videos from the testing dataset [21, 24] and 15 raw videos from the testing dataset [18]. All the experiments undertaken on the MATLAB 2016b framework.

4.2 Comparative Study

In this paper, we have compared our experimental results with many existing techniques such as *Itti* [26], *Surprise* [30], *Judd* [31], *PQFT* [32], *Rudoy*[33], *Fang* [34] and *OBDL* [35] 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 35 dB). The resolution of input raw videos vary from 416 × 240 (240p) and 1920 × 1080 (1080p). Performance metrics comparison is given in Table 1.

Table 1. Comparison of Performance Matrices with Existing Techniques for All Videos in Hevc Video_Database Dataset with Our Ps

PARAMETER	<i>Itti</i> [24]	<i>Surprise</i> [25]	<i>Judd</i> [26]	<i>PQFT</i> [27]	<i>Rudoy</i> [28]	<i>Fang</i> [29]	<i>OBDL</i> [30]	HEVC[19]	PS
AUC	0.668	0.752	0.816	0.750	0.785	0.797	0.785	0.823	0.846
NSS	0.445	1.078	1.427	1.300	1.401	1.306	1.511	1.658	1.702
LCC	0.119	0.272	0.187	0.311	0.386	0.370	0.352	0.438	0.388
KL	0.104	0.183	0.285	0.239	0.269	0.266	0.236	0.300	0.532

Table 2 Comparison of Jsd for Different Videos of a Video_Database Dataset

PARAMETER	RACEHORSED	MALE	BQTERRACE	SLIDESHOW
JSD	0.5033	0.480	0.4109	0.615

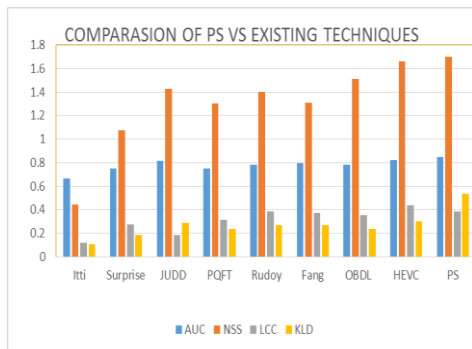


Figure 2 . comparison of our proposed model with existing techniques

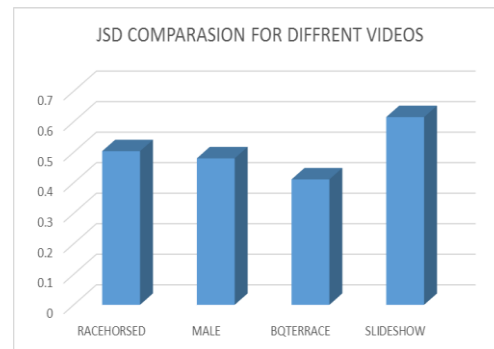


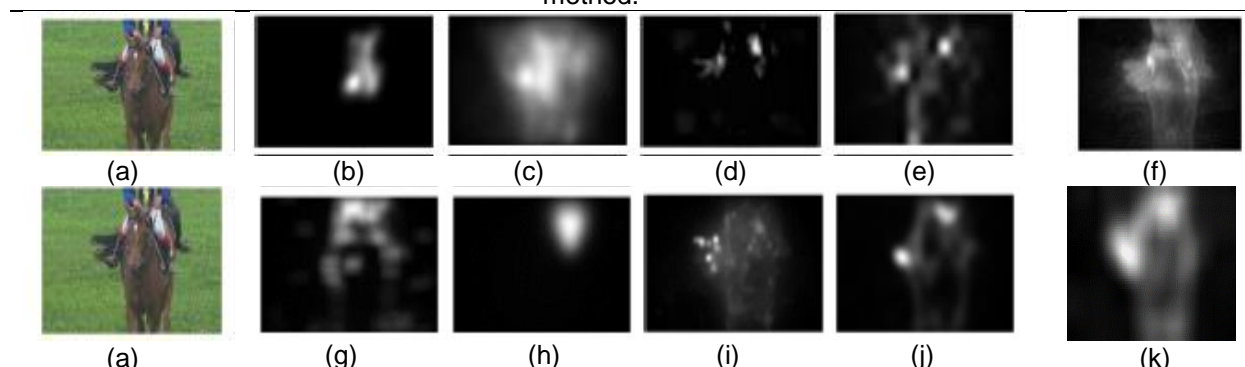
Figure 3. JSD comparison of different videos using our proposed method

Table 2 represents the performance comparison of our model with other existing techniques in terms of performance matrices such as AUC, NSS, LCC and KL. The average AUC (Area under Curve) value is very high using our proposed method compare to other existing techniques considering HEVC video_database dataset. The average AUC value by our proposed method is 0.846. Similarly, the average NSS (Normalized scan path saliency) and KLD (Kullback -Leibler divergence) value is very high using our proposed method compare to other existing techniques considering HEVC video_database dataset. The average NSS and KLD value by our proposed method are 1.702 and 0.532 respectively. However, Linear Correlation Coefficient (Kullback -Leibler divergence) is an exceptional case in our proposed method as its value is little less. Overall our model performs very efficiently and outperforms all the existing techniques in terms of saliency map generation and compression. Similarly, we have also computed the advance JSD (Jensen-Shannon divergence) parameter to detect saliency map more precisely. Here, in table 2 JSD parameter computed for four videos. The Slideshow video gives highest JSD as 0.615. Figure 2 shows the graphical comparison of different matrices for

all the existing techniques with proposed model. Similarly, figure 3 shows the JSD comparison for different videos using our proposed method.

Table 3 Saliency maps of RaceHorseD video selected from the first time of our cross validation experiments. The maps were yielded by our and other 7 methods as well the ground-truth human fixations. Note that the results of only one frame are shown for selected video. (a) Input. (b) Human. (c) hevc. (d) ltti. (e) Surprise. (f) Judd. (g) PQFT. (h) Rudoy. (i) Fang. (j) OBDL. and (k) author proposed work.. Our experimental results shows the better outcomes in comparison either other techniques. Therefore, it produces a better saliency map.

Table 3 shows the saliency map comparison of all the existing techniques with our proposed method.



5. Conclusion

21st century has provided an enormous evolution in the field of High Definition videos. However, there are few problems associated with it, which cannot be ignore. Therefore, to reduce these drawbacks we have implemented an efficient video compression technique as modified HEVC coding based on saliency features. In this paper, we have estimated saliency, considering HEVC video_database dataset. HEVC video_database contains total 33 videos. Here, we have shown saliency map comparison with other existing techniques for Race HorseD for videos in Table 3. Our experimental results outperforms all the existing techniques in terms of saliency map detection, AUC, NSS, KLD and JSD which shown in Table 1 and 2. The average AUC, NSS and KLD value by our proposed method are 0.846, 1.702 and 0.532 respectively which is very high compare to other existing technique. Similarly, the Slideshow video gives highest JSD as 0.615 using our proposed method. These results verify that our model is more efficient than any other 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] Cisco. Cisco visual networking index: Forecast and methodology, 2012- 2017. White Paper. 2013.
- [2] J Ostermann et al. Video coding with H.264/AVC: tools, performance, and complexity. *IEEE Circuits and Systems Magazine*. 2004; 4 (4): 7–28.
- [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*. 2017; 6 (1): 104–109. DOI: 10.11591/ijeecs.v6.i1.pp104-109.
- [4] Mengmeng Zhang, Jianfeng Qu, Huihui Bai2. Fast Intra Prediction Mode Decision Algorithm for HEVC. *Indonesian Journal of Electrical Engineering and Computer Science*. 2013; 11 (10): 5703 – 5710. ISSN: 2302-4046.
- [5] S R Smooth, R A Rowe. *Study of DCT coefficients distributions*. Proc. SPIE, San Jose, CA, USA. 1996; 365–368.
- [6] N Kamaci, Y Altunbasak, R M Merereau, “Frame bit allocation for the H.264/AVC video coder via cauchy-density-based rate and distortion models,” *IEEE Trans. Circuits Syst. Video Technol*. 2005; 15 (8): 994–1006.
- [7] Z He, Y K Kim, S. K. Mitra. Low-delay rate control for DCT video coding via p-domain source modelling. *IEEE Trans. Circuits Syst. Video Technol.*, 2001; 11 (8): 928–940.

- [8] J Sun, Y Duan, J Li, J Liu, Z Guo. Rate-distortion analysis of dead-zone plus uniform threshold scalar quantization and its application—Part II: Two-pass VBR coding for H.264/AVC. *IEEE Trans. Image Process.* 2013; 22 (1): 215–228.
- [9] J Hou, S Wan, Z Ma, and L P Chau. Consistent video quality control in scalable video coding using dependent distortion quantization model. *IEEE Trans. Broadcast.*, 2013; 59 (4): 717–724.
- [10] Y H Tan, C Yeo, Z Li. Single-pass rate control with texture and non-texture rate-distortion models. *IEEE Trans. Circuits Syst. Video Technol.*, 2012; 22 (8): 1236–1245.
- [11] C Y Wu and P C Su. A content-adaptive distortion–quantization model for H.264AVC and its applications. *IEEE Trans. Circuits Syst. Video Technol.*, 2014; 24 (1): 113–126.
- [12] M Tagliasacchi, G Valenzise, S Tubaro. Minimum variance optimal rate allocation for multiplexed H.264/AVC streams. *IEEE Trans. Image Process.*, 2008; 17 (7): 1129–1143.
- [13] M Jeon, N Kim, B D Lee. MapReduce-Based Distributed Video Encoding Using Content-Aware Video Segmentation and Scheduling. *IEEE Access*, 2016; 4: 6802-6815. doi: 10.1109/ACCESS.2016.2616540.
- [14] A Ilic, S Momcilovic, N Roma, L Sousa. Adaptive Scheduling Framework for Real-Time Video Encoding on Heterogeneous Systems. *IEEE Transactions on Circuits and Systems for Video Technology.* 2016; 26 (3): 597-611,. doi:10.1109/TCSVT.2015.2402893.
- [15] K L Chung, Y H Huang, C H Lin, J P Fang, "Novel Bitrate Saving and Fast Coding for Depth Videos in 3D-HEVC," in *IEEE Transactions on Circuits and Systems for Video Technology*, 2016 ; 26 (10): 1859-1869,. doi: 10.1109/TCSVT.2015.2473296.
- [16] G Kim, K Yi, C M Kyung, "A Content-Aware Video Encoding Scheme Based on Single-Pass Consistent Quality Control," in *IEEE Transactions on Broadcasting*, 2016; 62 (4): 800-816,. doi: 10.1109/TBC.2016.2569999.
- [17] J Chao, E Steinbach, "Keypoint Encoding for Improved Feature Extraction From Compressed Video at Low Bitrates," in *IEEE Transactions on Multimedia*, 2016; 18 (1): 25-39,. doi: 10.1109/TMM.2015.2502552.
- [18] Hadizadeh H, Enriquez MJ, Bajić IV., Eye-tracking database for a set of standard video sequences. *IEEE Trans Image Process.*, 2012; 21 (2): 898–903.
- [19] M Xu, L Jiang, X Sun, Z Ye, Z Wang, "Learning to Detect Video Saliency With HEVC Features," in *IEEE Transactions on Image Processing*, 2017; 26 (1): 369-385.
- [20] T Wiegand, G J Sullivan, G Bjontegaard, A Luthra. Overview of the H. 264/AVC video coding standard. *IEEE Trans. Circuits Syst. Video Technol.*, 2003; 13(7): 560-576.
- [21] G Sullivan, J Ohm, W J Han, T Wiegand. Overview of the high efficiency video coding (HEVC) standard. *IEEE Trans. Circuits Syst. Video Technol.*, 2012; 22(12): 1649-1668.
- [22] M G Arvanitidou, A Glantz, A Krutz, T Sikora, M Mrak, A. Kondoz. *Global motion estimation using variable block sizes and its application to object segmentation.* Proc. IEEE WIAMIS'09, 2009: 173-176.
- [23] M Xu, X Deng, S Li, Z Wang. Region-of-interest based conversational HEVC coding with hierarchical perception model of face. *IEEE J. Sel. Topics Signal Process.*, 2014; 8 (3): 475–489.
- [24] J R Ohm, G J Sullivan, H Schwarz, T K Tan, T Wiegand. Comparison of the coding efficiency of video coding standards-Including high efficiency video coding (HEVC). *IEEE Trans. Circuits Syst. Video Technol.*, 2012; 22 (12): 1669–1684.
- [25] L Itti. Automatic foveation for video compression using a neurobiological model of visual attention. *IEEE Trans. Image Process.*, 2004; 13(10): 1304-1318.
- [26] L Itti, C Koch, E Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 1998; 20 (11): 1254-1259.
- [27] Borji A, Itti L .State-of-the-art in visual attention modeling. *IEEE Trans Pattern Anal Mach Intel.*, 2013; 35(1): 185–207.
- [28] Borji A, Sihite DN, Itti L, Quantitative analysis of human-model agreement in visual saliency modeling: a comparative study. *IEEE Trans Image Process.* 2013; 22(1): 55–69.
- [29] Itti L, Baldi P. A principled approach to detecting surprising events in video. In: IEEE Computer society conference on computer vision and pattern recognition (CVPR'05), 2005; 1: 631– 637.
- [30] L Itti, P Baldi. Bayesian surprise attracts human attention. *Vis. Res.*, 2009; 49 (10): 1295–1306.
- [31] T Judd, K Ehinger, F Durand, A Torralba. *Learning to predict where humans look.* Proc. ICCV, Sep./Oct. 2009: 2106–2113.
- [32] C Guo, L Zhang. A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression. *IEEE Trans. Image Process.*, 2010; 19 (1): 185–198.
- [33] D Rudoy, D B Goldman, E Shechtman, L Zelnik-Manor. *Learning video saliency from human gaze using candidate selection.* Proc. CVPR, Jun. 2013: 1147–1154.
- [34] Y Fang, W Lin, Z Chen, C M Tsai, C W Lin. A video saliency detection model in compressed domain. *IEEE Trans. Circuits Syst. Video Technol.*, 2014; 24 (1): 27–38.
- [35] S Hossein Khattoonabadi, N Vasconcelos, I V Bajic, Y Shan. *How many bits does it take for a stimulus to be salient?.* Proc. CVPR. 2015: 5501–5510.