Webology (ISSN: 1735-188X)
Volume 19, Number 2, 2022

Visual Saliency Estimation Using Combined Feature Of PDDL And MRF

Vinay C Warad[1*] and Dr. Ruksar Fatima[2]

[1]Assistant Professor, Department of Computer Science and Engineering, Khaja Bandanawaz College of Engineering, Gulbarga-585105, Karnataka, India.

[2]Head of the Department, Computer Science and Engineering, Khaja Bandanawaz College of Engineering, Gulbarga-585105, Karnataka, India..

Abstract-Last few decades has seen the big growth in visual saliency due to its usability in the real time scenario. In past several models have been proposed, despite of research variety the existing model fails to match the human level accuracy. In this research work, we have proposed a technique named as SEIPM –video encoding, in here the video size is compressed without degrading any quality in the video. Moreover, SEIPM is based on the saliency prediction of spatial and temporal features; these features are combined by PDDL (Processing Division of Description Layer) and MRF (Markov Random Field). MRF is used for analyzing the saliency frame to recognize movement and motions in the frame and PDDL is used for the feature extraction which are very much eye attracting. Once the region of movement and the particular region, which attracts the eye, is recognized, then encoding is done from the previous frame only and hence we tend to achieve the quality video. Moreover, SEIPM-video encoding is evaluated by considering the standard dataset used in existing protocol by considering the various performance metric and compared with the various state-of-art technique and our models outperforms the other state-of-art technique.

1 INTRODUCTION

The widespread utilization of inexpensive and portable video capture devices, such as camcorders and mobile phones, coupled with the surge in surveillance cameras, has led to a remarkable increase for data. Besides, thanks to the ubiquitous Internet and various video sharing sites (e.g., YouTube, Douyin, Tik Tok, etc.), many users are pleased to upload their videos on the web as social media.

Twenty first century has been the era of technologies specially in multimedia Last two decades has been revolution for capturing devices such as mobile phone, camcorders and the surveillance camera, these devices are cost effective and can be used for the various purpose. Meanwhile with growth of internet, video sharing site such as Tik Tok, YouTube, and Instagram has been very popular in a while; however, the video quality plays an important part. Moreover, for better appealing of the videos usually video owner edit the video such as cropping, scaling, and insertion of frame as well as deletion of frame. The alteration in video is performed in the pixel domain such that it can be decoded and re encoded if needed. Recent year has seen the major change in compression technologies and video coding standards. Hence, the video content has a bigger role to play in the daily life with increasing demand for high quality video such as HDTV, Blu-ray and UHD. Moreover high quality video needs to be compressed.

To make videos more appealing, video owners usually edit their videos before uploading, such as scaling, cropping, frame-deletion/insertion, etc. Each of these alterations needs to be done in the pixel domain, so an altered video has to be decoded and then re-encoded inevitably. Because of the advent of the multimedia and information era, the past two decades have seen the major developments of video coding standards and video compression technologies in significant compression gains and impressive deployment of services and applications. Meanwhile, the introduction of video content has continued to become an increasing presence in our lives, with increasing diversification of usage models along with ever-increasing demands for higher quality. Consumers now expect higher resolution for their entertainment video, as standard-definition analog broadcast TV and VCR movies have given way to HDTV, DVD, and Blu-ray, and UHD video has emerged on the prospect. Due to the rapid growth of the Internet and wireless technologies, video compression becomes essential for reducing the bandwidth for transmission and storage in many applications. In the case of limited bandwidth and storage resources. However, new requirements have been raised for the current video coding standard, such as higher resolution and image quality [Muller et al. 2013]. Moreover, several researcher have focused in developing video compression technique, since video compression helps in minimizing the bandwidth for storage and transmission for various applications. Rise in demand for image quality and high resolution [Muller et al. 2013] has encouraged video coding standards. Meanwhile the above scenarios has led to the development of HD and UHD [Sullivan et al. 2012].

Moreover, to meet the industry norms several technique has been proposed, these techniques have discussed the several issue regarding the same. To start with two premier organizations i.e. ITU-T and ISO/IEC, they developed the two standardized video coding for establishing the JCT(Joint Collaborative Team-VC), this tends to provide the new standard of HEVC and it was pretended that it’s a natural evolution of AVC. Later a journal CFP was proposed for the coding performance on HD pictures where the size of the pictures were up to low delay, 8Kx4K and minimized complexity [Hsueh-Ming Hang et al. 2010]. Moreover, it is observed that video content has grown threefold by 2021. However, it is also observed that video compression technique has not been changed over the last two decades [Vanne et al 2012]. Moreover, few of existing method are thoroughly tuned and fine engineered, however they are hard coded this makes unadoptable for the increase in demand in video spectrum for various use cases such as VR streaming, object detection and social media sharing. Meanwhile last two years we observed the large impact of ML (Machine Learning) in image compression, however still in more of infancy stage for the real time scenario. However Rate Control algorithm has been one of the major step in compression. Rate control is one of the coding tools in VCS(Video Coding Standards). A rate control algorithm first allocates a bit budget to the group of pictures, frames, or the coding units depending on the level at which the rate control is being applied [Ahmad and Luo 2006]. The allocated bit budget for each unit at its level is then used to calculate the model parameters, which in turn are used to encode the video for achieving the target bit rate. In the theory of video compression, the quantization step size primarily determines the degree of spatial detail retained in the video. Rate control is a well-studied topic in video compression and several efficient schemes have been widely used in standards. However, numerous new algorithms continue to be reported in the literature, typically improving on the past results and taking into account the complexity of the new video coding such as HEVC [Shanshe et al 2013].

Considering all the above discussion and the survey we have observed that there isn't an ideal method which can provide the encoding without degrading quality, hence in this paper we propose a technique named SEIPM-video encoding which compresses the video in efficient manner, the contribution of the research work has been highlighted through the below points:

- This research works aims at proposing the technique for video compression without degrading any quality.
- We propose a technique SEIPM-video encoding for the video compression; this is based on the spatial and temporal features.
- Features are combined by PDDL (Processing Division of Description Layer) and MRF (Markov Random Field). MRF is used for analyzing the saliency frame to recognize movement and motions in the frame and PDDL is used for the feature extraction which are very much eye attracting. Once the region of movement and the particular region, which attracts the eye, is recognized, then encoding is done from the previous frame only and hence we tend to achieve the quality video.
- SEIPM is evaluated by considering the various state-of-art technique with various parameter.

2 LITERATURE SURVEY

Regarding to the predecessors and development which was largely collaborative effort so that the industries as well as community related to the academic in to the extensible design as well as each coherent knowledge which is collective in compression occurs major advance as this is enabled by the HEVC standard. ITU-T in VCEG has joint video project recently is HEVC and ISO/IEC in MPEG standard organization, they are working together as the JCT-VC [Bross et al 2012]. In January 2013, the primary version of high efficiency of video coding is supposed to be finalized, in aligned text it will be resulted and ISO/IEC along with ITU-T has been published it. In order to assist the scenarios consisting range which is expanded and that expanded range will be utilized with the color format support, enhanced precision and scalable video coding as well for some additional application the standard will be expanded as it is planned. The part 2 (ISO/IEC 23008-2) of MPEG-H along with ITU-T will become high efficiency of video coding. It is similar as ITU-T Recommendation H.265. in order to exhibit the history of video compression, there are many researchers as well as scholars done a lot of research for potential characteristics in video compression [Singh, and Aggarwal 2018] like [Jain 1989] DCT coefficient distribution as well as blocking MPEG videos' artifacts DCT coefficients consists 1st digit law, features that are based on Markov, [Jain 1989] prediction has been done based on the HEVC videos' unit feature, more to say, in video compression, one of the task which is common is video transcoding. Basically, temporal filtering techniques or considering spatial or combination of both will perform de-noising and this is given in literature that some of the architecture of video de-noising previously exist. In fact, there are many spatial techniques that are Wiener, [Zhan and Karam 2003] wavelet filtering, it tends to be more and more accurate for the identical images, although, because of their nature, spatio-temporal technique as well as temporal technique is utilized to get appropriate video signals because of the temporal correlation and this will exist in between adjacent images. And such type of methods is found in detailed. There are two variants for such methods such as motion along with non-motion compensated filters, and in order to filter the current image it may suppose or not motion estimation and methods for compensation. As compared to the less visually significant regions there are the techniques that are proposed with the concept of the predicted attention-grabbing having area around it and that area will be encoded by those techniques. The fact of mall region having 2–5◦ of visual angle as it is having on every side of it is center of gaze which is perceived in together with the resolution and the reason behind this is distribution of photoreceptors which is highly non-uniform as on the retina of the human and then that fact will assists spatial prioritization [Itti 2004]. In the field of computer vision, Human action recognition is considered to be more and more significant because in human-computer interaction, video indexing and video surveillance systems it has many possible applications. Although,

In order to identify actions of human appropriately and it is challenging due to obstruction, movements of camera as well as illumination changes, motion styles and so on., there are many existing techniques which are suffering from some obstacles, in the area of action recognition, their representation's discriminability will be diminished. Into the video level representations which are discriminative, to improve transform spatio-temporal local features as well as to face the obstacle.

This research work is organized in any well-mannered research work as first section discuss the background and the various phenomena, second section discuss the various existing technique, third section gives us the detailed description about the SEIPM technique. Moreover, the performance is evaluated in the fourth section.

3 PROPOSED WORK

In this part of the article, we are implementing a new Efficient Video-coder using feature responses of DLOC's (Description-Layer-of- Operational-Chunk) MRF (Markov-Random-Field) technique.

3.1 Description Layer of Operational Chunk (DLOC) Feature Responses:

Usually encoding of the video includes prediction of the motion and assumption of the compensation of motion, along with transformation, intra-estimation, quantization, motion vector, and estimation residuals' coding by entropy. Many steps was in place as the traditional standards of the video coding, though getting classier before time. Whereas, for the purpose of compactness, we aim on the coding standard H.264/AVC, the evaluations of the feature implemented here can be used for some other standard of video coding also, even the newest HEVC standard (High-Efficiency-Video-Coding). Because of aiming on the video coding standard H.264/AVC, our "chunk" is of a × a-pixel of macro-chunk denoted as MC.

The DLOC is evaluated immediately from the outcome of the entropy-decoder that is the first chunk that is processed by the video-decoder. No additional decoding on the bit-stream that is compressed is required. The amount of bits used for encoding each MC is extracted and assigned with the interval values [0, 1]. Here 0 and 1 is represented as the MC(s) needing the least and most bits to code respectively with all the MCs within the frame. The DLOC map, which is normalized, is smoothed by complexity using 2-Dimentional Gaussian, where the value of standard deviation is 2° of angle of view. While the DLOC map that is smoothed spatially is compact saliency evaluation, we see that an extra enhancement in the performance of estimation of saliency can be done by doing the smoothing LOC mapby temporally. This observe with the knowledge regarding the biological vision. Here, temporal filtering is assumed to be present in the first layers of cortex. Particularly, we simply use typical temporal mean for some milliseconds to get a derived feature DLOC.

3.2 Markov Random Field technique:

While the technique of compression of the video are very classy predictors of content's local information, they give the estimation of only local information, while all the operations are temporally and spatially localized to MC unit. Simultaneously, saliency consist of both global content as well as local content. Such as, many designs of saliency develop reserve of reverse tools that overwhelm the locations of image of saliency around the peaks of saliency. To include these kinds of effect, we depend on Markov-Random-Field (MRF) technique.

More precisely, the detection issues of saliency is designed as one of assuming the highest a posteriori (HAP) solution of Markov-Random-Field with Temporal-Spatio (MRFTS) technique. This is described based on the classification issue of binary, the salient chunks of pixel size having 16x16 is classified as class-1 and chunks having non-salient is classified as class-0. The aim is to define the labels of the class as $\mathcal{L}^{f} \in \{0, 1\}$of the frame's chunk f, where the labels are $\mathcal{L}^{1\cdots f-1}$ of the last frames, and overall compressed information that are lastly seen is denoted as$s^{1\cdots f}$. The finest label allotment $\mathcal{L}_*^{f}$is the one that increases the chances of posteriori to the highest $\mathrm{Proby}(\mathcal{L}^{f}|\mathcal{L}^{1\cdots f-1}, s^{1\cdots f})$. Using Bayes rule, this can be denoted as

$$\begin{aligned}\mathrm{Proby}(\mathcal{L}^{f}|\mathcal{L}^{1\cdots f-1}, s^{1\cdots f}) &\cong \mathrm{Proby}(\mathcal{L}^{1\cdots f-1}|\mathcal{L}^{f}, s^{1\cdots f}) \times \mathrm{Proby}(\mathcal{L}^{f}|s^{1\cdots f}) \\ &\cong \mathrm{Proby}(\mathcal{L}^{1\cdots f-1}|\mathcal{L}^{f}, s^{1\cdots f}) \times \mathrm{Proby}(s^{1\cdots f}|\mathcal{L}^{f}) \times \mathrm{Proby}(\mathcal{L}^{f}),\end{aligned} \qquad (1)$$

Here, ≅represents the similarity up to the constant of normalization. Assuming the logarithm's similarity, the solutions of HAP for the labels of saliency $\mathcal{L}^{f}$is later represented as:

$$\begin{aligned}\mathcal{L}_*^{f} = \arg\min_{\theta\in\omega^{f}} \{-1 \\ \times (\log \mathrm{Proby}(s^{1\cdots f}|\theta) + \log \mathrm{Proby}(\theta) \\ + \log \mathrm{Proby}(\mathcal{L}^{1\cdots f-1}|\theta, s^{1\cdots f}))\},\end{aligned} \qquad (2)$$

Here, ω^{f}represents the collection of all that can be configured label for f number of frames. According to the theorem of Hammersley-Clifford, the chances of posteriori in equation (2) can be represented as$\mathrm{Proby}(\mathrm{i}) = \frac{1}{\mathrm{Q}} \exp \frac{-\mathrm{Power}(\mathrm{i})}{\mathbb{C}}$, here, $\mathrm{Power}(\mathrm{i})$is the function of power, constant is denoted as$\mathbb{C}$, few of the times it represents as "temperature" and Q a separation function. This allows the redesigning of the prediction problem of HAP as

$$\begin{aligned}\mathcal{L}_*^{f} = \arg\min_{\theta\in\omega^{f}} \{\mathbb{C}_s^{-1} * \mathrm{Power}(\theta; s^{1\cdots f}) + \mathbb{C}_c^{-1} * \mathrm{Power}(\theta) + \mathbb{C}_f^{-1} \\ * \mathrm{Power}(\theta; \mathcal{L}^{1\cdots f-1}, s^{1\cdots f})\}\end{aligned} \qquad (3)$$

The modules$\mathbb{C}_s^{-1} * \mathrm{Power}(\theta; s^{1\cdots f})$, $\mathbb{C}_c^{-1} * \mathrm{Power}(\theta)$, and $\mathbb{C}_f^{-1} * \mathrm{Power}(\theta; \mathcal{L}^{1\cdots f-1}, s^{1\cdots f})$ of the function of power, calculates the amount of temporal presence for

- The labels of saliency.
- Coherence among labels and observation of the features, and
- The label field's compactness of the spatial.

A highly specific description of the above three modules is described in the following below divisions. At the last, the issue of reduction is resolved by the technique of Modes of Conditional Loop (MCL).

3.3 Presence of Temporal:

Provided a chuck at the location of the image $b = (i, j)$of $\mathscr{f}$ number of frames, the surrounding of temporal-spatio $S(b)$ be given as collection of chunks $n = (i', j', \mathscr{f}')$. Where, $1 \geq |i - i'|, 1 \geq |j - j'|$and $\mathscr{f} > \mathscr{f}' > f - X$ for some value of X. The field of label's presence of temporal is calculated locally, with the help of

$$\mathrm{Power}\left(\theta; \mathcal{L}^{1\cdots\mathscr{f}-1}, \mathscr{s}^{1\cdots\mathscr{f}}\right) = \sum_{b} \mathrm{Power}_{\mathscr{f}}(b), \tag{4}$$

Here, $\mathrm{Power}_{\mathscr{f}}(b)$is non-presence amount in $S(b)$ that effects the assignments of non-present labels of temporal, i.e. $\mathcal{L}^{\mathscr{f}'}(i', j') \neq \mathcal{L}^{\mathscr{f}}(i, j)$.

The label of saliency $\mathcal{L}(a)$of chunk a is considered as distribution of Bernoulli with ratio of parameter to the features' strength $\mathscr{s}(a)$, i.e. $\mathrm{Proby}\big(\mathcal{L}(a)\big) = \left(1 - \mathscr{s}(a)^{1-\mathcal{L}(a)}\right) \times \mathscr{s}(a)^{\mathcal{L}(a)}$. It represents that the probability $\mathrm{bind}(b, a)$, where chunk a will bind with chunk b (i.e. include $\theta(b)$) is

$$\mathrm{bind}(b, a) = \left(1 - \mathscr{s}(a)^{1-\mathcal{L}(a)}\right) \times \mathscr{s}(a)^{\mathcal{L}(a)}. \tag{5}$$

With the help of similarity function, the presence amount of this probability is weighted, depending on function of Gaussian for distance among b and a,

$$\begin{aligned} \mathrm{dist}(b, a) \cong \exp&\left(-0.5 * \rho_{\mathrm{spatial}}^{-2} * \|a - b\|_2^{\mathrm{spatial}}\right) \\ &\times \exp\left(-0.5 * \rho_{\mathrm{temporal}}^{-2} * \|a - b\|_2^{\mathrm{temporal}}\right) \end{aligned} \tag{6}$$

Here, $\rho_{\mathrm{spatial}}^{-2}$and $\rho_{\mathrm{temporal}}^{-2}$are normalization parameters, and $\|*\|_2^{\mathrm{spatial}}$and $\|*\|_2^{\mathrm{temporal}}$are the distances of Euclidean for the spatial and temporal measurement, respectively. The assumed presence between two places is represented as

$$\mathrm{presence}(b, a) = \left(\frac{\sum_{q \in S(b)}\big(\mathrm{dist}(b, a) \times \mathrm{bind}(b, a)\big)}{\mathrm{dist}(b, a) \times \mathrm{bind}(b, a)}\right)^{-1} \tag{7}$$

Here, qdescribes a prior assumption for the labels' presence depending on the features that were seen$\mathscr{s}(a)$. This power function then effects the labeling of non-presence, uniformly to this prior assumption of presence

$$\mathrm{Power}_{\mathscr{f}}(b) = \sum_{a \in S(b)} \mathrm{presence}(b, a) \times \left(1 - \mathcal{L}(a)^{\theta(b)}\right) \times \mathcal{L}(a)^{1-\theta(b)} \tag{8}$$

Remember that, $\mathrm{Power}_{\mathscr{f}}(b)$is between [0, 1]. The value is 0, when all surrounding chunks $a \in S(b)$have similar label as chunk b, and the value is 1, when the surrounding chunks have label dissimilar to $\theta(b)$.

3.4 Observation of Features and Labels of Coherence:

The error among fields of label and seen features at time $\mathscr{f}$ is calculated with the function of power $\mathrm{Power}(\theta; \mathscr{s}^{1\cdots\mathscr{f}})$. Whereas, this helps the requirement of $\mathcal{L}^{\mathscr{f}}$ on all prior seen features$(\mathscr{s}^{1\cdots\mathscr{f}-1})$, we consider that the present labels are based only on present seen features$(\mathscr{s}^{\mathscr{f}})$. Error is then calculated by the power function

$$\mathrm{Power}(\theta; \mathscr{s}^{1\cdots\mathscr{f}}) = \sum_{b} \left(\left(1 - \sup_a(\mathscr{s}(a))\right)^{\theta(b)} \times \left(\inf_a(\mathscr{s}(a))\right)^{1-\theta(b)} \right), \tag{9}$$

Here, sup is supremum and inf is infimum that are defined based on $a = (i', j')$ in a way that $1 \geq |j - j'|$ and $1 \geq |x - x'|$. This also range from 0 to 1 and effects the labeling of chunk b as un-salient, i.e. if$\sup_a(\mathscr{s}(a))$, the value of feature's supremum is lesser, then $\theta(b) = 1$.Similarly, if $\inf_a(\mathscr{s}(a))$, the value of feature's infimum is higher or salient, then $\theta(b) = 0$.

3.5 Compression of Features:

Basically, the chances of chunk getting salient with label must grow high if its surrounding are mostly salient. The second component of power in Equation-3 use this kind of habit. It is represented as

$$\mathrm{Power}(\theta) = \sum_{b} (1 - \mathrm{saliency}(b))^{\theta(b)} \times (\mathrm{saliency}(b))^{1-\theta(b)}, \tag{10}$$

Here, $\mathrm{saliency}(b)$ is the amount of saliency in the b surrounding, which is represented as

$$\mathrm{saliency}(b) = A \sum_{a \in S^{+}(b)} \theta(a) + B \sum_{a \in S^{\times}(b)} \theta(a), \tag{11}$$

Here, $S^{+}(b)$and $S^{+}(b)$ are first (East, West, North, South) and second-order (Southeast, Northeast, South-West, North-West) surrounding of chunk b, respectively. While testing our technique, we assign $A = 0.1666$ and$B = 0.0833$, to provide maximum weight to the surrounding of first-order.

3.6 Mapping of Saliency:

The operations represented above creates chuck labels that are having more chances, a posteriori, plotting of salient. To underline the location where the drawing of attention is high, the DLOC of the chunk that is given as non-salient by the Markov-Random-Field in decreased/increased based on the DLOCs in its surrounding. The operation is equated as

$$\mathrm{Map}_{\mathrm{salient}}(b) = \frac{(\sup_a\{\mathrm{dist}(a,b) \times \mathscr{s}(a)\})^{\theta(b)}}{\left(1 - \sup_p\{\mathrm{dist}(a,b) \times (1 - \mathscr{s}(a))\}\right)^{(\theta(b)-1)}} \tag{12}$$

Here, $a = (i', j', \mathscr{f}')$ is denoted as the group of chunks. For example, $1 \geq |i - i'|, 1 \geq |j - j'|$and $\mathscr{f} \geq \mathscr{f}' > f - X$ and $\mathrm{dist}(a, b)$as denoted in (6). By this way, chunk b represented as salient by MRF implication is given a saliency that is equal to the value of highest feature among its surrounding, included with its distance from b as its starting point. Similarly, for chunk b, which is represented as non-salient, this task is utilized to the value required of the values of saliency amongS_b. The required value of the saliency value is then given as the new saliency value of b.

4 PERFORMANCE EVALUATION

In this research work, we have proposed a methodology named as SEIPM-VIDEO ENCODING aka Saliency Estimation using Integration of Processing Division of Description layer and MRF for video encoding. Moreover, this section presents the evaluation of proposed technique, meanwhile for evaluation ideal configuration of 8GB RAM, MATLAB as programming language. Moreover, the system is loaded with 4GB of Nvidia graphics card.

4.1 Dataset and evaluation

Dataset plays major for the evaluation, absolute dataset might end up for good prediction, whereas non-ideal dataset might be bad for research as it might end up predicting false result although if the technique is good. In this research work we have used the database of Eye Tracking on the Raw Videos from existing methodologies, Here, all the raw videos are from test cases [Deng and Wang 2014]. Moreover, our model is compared against the various state-of-art technique, these methods are ITTI [14 Itti et al 1998], surprise [Itti and P. Baldi 2009], JUDD [Judd et al 2009], PQFT [Guo and Zhang 2010], Rudoy [Rudoyet al 2013], Fang [Fang et al 2014], HEVC [Jiang et al 2017] and Compressed HEVC [Zhou et al 2018].

4.2 Parametric metrics

In order to prove the efficiency of proposed model, we have compared with the 7 other state-of-art technique discussed in the earlier section, this comparison has been done based on the various performance metrics such as AUC, NSS, CC and KL, these are discussed later in the same section.

4.2.1 AUC

AUC aka Area under Curve is one of the performance metric considered for the comparison, the comparison value has been depicted in the below table. AUC values lies from the range of 0.5 to 1, the higher value indicates the model is more efficient. In here we observe that the other methods like Itti, Surprise, JUDD, PQFT, Rudoy, Fang,, HEVC and Compressed HEVC achieves the AUC value of 0.668, 0.752, 0.816, 0.75, 0.785, 0.797,0.823 and 0.775 whereas proposed model achieves the value of 0.84. Moreover, through the comparative analysis, observation can be made that proposed model outperforms the other state-of-art technique.

Methodology	AUC
Itti	0.668
Surprise	0.752
JUDD	0.816
PQFT	0.75
Rudoy	0.785
Fang	0.797
HEVC	0.823
Compressed HEVC	0.775
Proposed	0.84

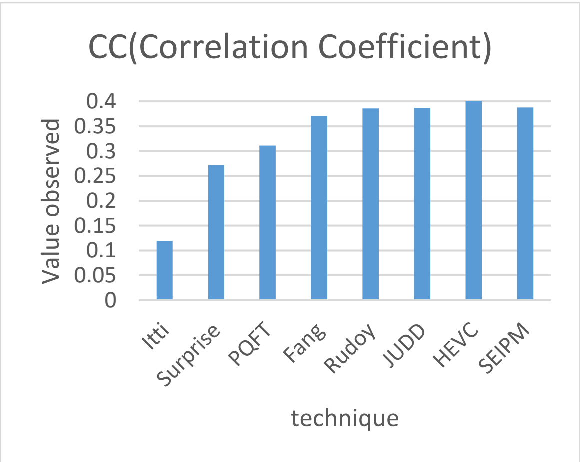
4.2.2 NSS(Normalized Scanpath Saliency)

NSS is one of the evaluation, which have been used for the saliency evaluation, NSS is defined and computed as the average of the RV (Response Values) in a saliency map, which has been normalized to achieve zero, the higher value of NSS is expected when compared to the other technique. Below table shows the comparison of various model based on the NSS, here Itti, Surprise, JUDD, PQFT, Rudoy, Fang and HEVC achieves the value of 0.445, 1.078, 1.427, 1.300, 1.401, 1.306 and 1.658 whereas proposed model achieves slight better than the existing model.

Methodology	NSS
Itti	0.44500
Surprise	1.078
JUDD	1.427
PQFT	1.300
Ruoy	1.401
Fang	1.306
HEVC	1.658
Proposed	1.702

4.2.3 CC (Correlation Coefficient)

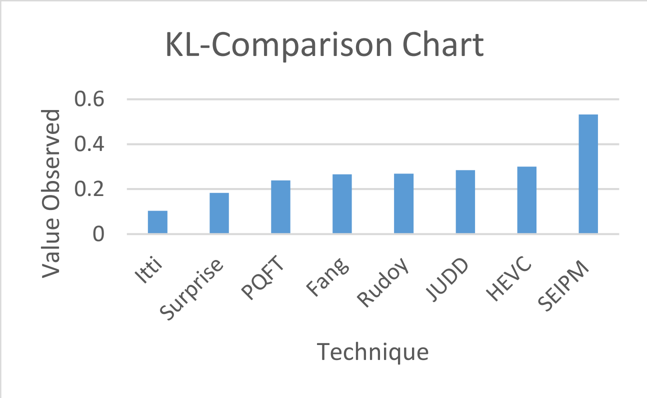
In general, Correlation Coefficient is defined as the measure of bond between the two variables, i.e. how strong the bond is between the two variable and it is mostly used in LR (Linear Regression). In here, it is used as one of the comparison metric for evaluation of SEIPM technique. The below graph shows the comparison of various technique compared with the SEIPM, the value observed by Itti, Surprise, PQFT, FANG, Rudoy, JUDD and HEVC are 0.119, 0.272, 0.311, 0.37, 0.386, 0.387 and 0.438. Moreover, SEIPM observes the value of 0.388, which is slightly lesser than the existing model.



4.2.4 Kullback-Leibler divergence

KL-divergence or the relative entropy is nothing but measuring the difference of two probability distribution, moreover in here KL is measured on the saliency map. The below graph represents the comparison of various model. Here Itti, Surprise, PQFT, Fang, Rudoy, JUDD model observes the 0.104, 0.183, 0.239, 0.266, 0.269, 0.285 respectively. Moreover existing model i.e. HEVC observes the value of

0.3 and our model i.e. SEIPM observes the value of 0.532 and it is marginally higher than the existing model.



5 CONCLUSION

In this paper, we have proposed a methodology named SEIPM-video encoding for compressing the video without any degradation in quality. SEIPM extract the feature of compressed videos by taking account of spatial and temporal information, moreover the main aim of this technique isto achieve the quality video and this is achieved through two important factor i.e. focusing on the movement region and the region, which capture the eye most. Moreover, SEIPM has been evaluated by considering the standard parameter such as Kullback-Leibler divergence, Correlation coefficient, Normalized Scanpath Saliency and Area under Curve. These performance metric shows the model efficiency, hence we compared with various state-of-art technique we observe that our model achieves the value of 0.532, 0.388, 1.702 and 0.84 respectively for the above metric. When the comparative analysis is done, it is observed that our model outperforms the other state- of-art technique in terms of all metric except Correlation Coefficient; however, the value is slightly less.

Moreover considering the importance of video quality this research could be prominent benchmark, however yet there has to be seen how it works under different scenario and constraint. Future works includes the more analysis considering various parameter and scenarios.

REFERENCES

Ahmad I and Luo J, "On using Game Theory to Optimize the Rate Control in Video Coding, " IEEE Transactions on Circuits and Systems for Video Technology, vol. 16, pp. 209-219, 2006.

Bross B, Han W, Sullivan G.J, Ohm J.R, and Wiegand T, High Efficiency Video Coding (HEVC) Text Specification Draft 9, document JCTVC-K1003, ITU-T/ISO/IEC Joint Collaborative Team on Video Coding (JCT-VC), Oct. 2012.

Deng M. Xu, Li S, and . Wang Z, “Region-of-interest based conversational HEVC coding with hierarchical perception model of face,” IEEE J. Sel. Topics Signal Process., vol. 8, no. 3, pp. 475–489, Jun. 2014.

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

Guo C and Zhang L, “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.

Hsueh-Ming Hang, Wen-Hsiao Peng, Chia-Hsin Chan and Chun-Chi Chen,” Towards the Next Video Standard: High Efficiency Video Coding”, Proceedings of the Second APSIPA Annual Summit and Conference, pages 609–618, Biopolis, Singapore, 14-17 December 2010.

Itti L, “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.

Itti L,Koch C, and Niebur E, “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.

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

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

Muller K, Schwarz H, Marpe D, et al. 3D high-efficiency video coding for multi-view video and depth data, IEEE Transactions on Image Processing, 2013, 22(9): 3366-3378.

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

Sullivan G. Ohm J, R,. Han W. J, and Wiegand T “Overview of the high efficiency video coding (HEVC) standard,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, pp.1649-1668, 2012.

Shanshe W, Siwei M, Zhang L, Shiqi W, Debin Z, and Wen G, "Multi-layer based rate control algorithm for HEVC, " 2013 IEEE International Symposium on Circuits and Systems (lSCAS), pp.41 -44 May 2013.

Singh, R.D., Aggarwal, N. Video content authentication techniques: A comprehensive survey. Multimed Syst. 2018, 24, 211–240.
Jain A K , fundamentals of Digital Image Processing. Eagle-Wood cliffs NJ: Prentice-Hall; 1989

Vanne J, Viitanen M, Hamalainen T. D, and Hallapuro A. Comparative rate-distortion-complexity analysis of HEVC and avc video codecs. IEEE Transactions on Circuits and Systems for Video Technology, 22(12):1885–1898, 2012..

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

Zhan C. Q and Karam L.J "Wavelet-Based Adaptive Image De-noising With Edge Preservation," Proc. of the IEEE 2003 Int. Conf.onImage Process. MA-L2, Sept.'03.

Zhou W, Bai R and Wei H, "Saliency Detection With Features From Compressed HEVC," in IEEE Access, vol. 6, pp. 62528-62537, 2018.

AUTHORS PROFILE

Author-1

Prof. Vinay Warad is working as Assistant Professor in Department of CSE in Khaja Bandanawaz College of Engineering, Karnataka, since 2014. Master of Engineering from Sathyaba University Chennai India, He has attended many workshops such as Research Methodology conducted K.B.N. College of Engineering, Gulbarga, workshop on Big-Data Analytics and Hadoop and have Published 1 conference paper in IEEE Xplore.

Author-2

Dr. Ruksar Fatima is presently working as Dean Faculty of Engineering and Technology, Khaja Bandanawaz University, Kalaburagi, with an experience of 18 years as a dedicated, resourceful education professional. She has received 4 awards, Award for Best Scientific publication by VGST in 2018, RURLA award for Distinguished Scientist in 2018, Chairman for IETE Gulbarga sub-center and Best Senior Researcher (Female) by international Academic and Research Excellence Awards 2019. She has 31 International publications in reputed journals and technical member and reviewer of many famous international Journals. She has written a book entitled “Image Processing Techniques for classification of Images” https://www.interlinepublishing.com/title cover/large/caf591ac8e3ef54ba4d4046650e5971e.jpg