Indonesian Journal of Electrical Engineering and Computer Science
Vol. 32, No. 1, October 2023, pp. 392~402
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v32.i1.pp392-402 ❒ 392

Video perception enhancement using effective encoding optimization in future generation wireless network

Aruna Kumar Bananaik Thulajanaik[1], Naganaik Manjanaik[2]
[1]Department of Computer Science and Engineering, Gowdara Mallikarjunappa Institute of Technology, Davangere, India
[2]Department of Electronics and Communication Engineering, University Brahmappa Devendrappa Tavanappanavar College of Engineering, Davangere, India

Article Info

Article history:

Received Dec 23, 2022
Revised Jun 26, 2023
Accepted Jul 2, 2023

Keywords:

Compression
Error detection
Fifth generation network
Perceptual quality
Sparse feature
Sptial temporal
Video encoding

ABSTRACT

The cumulative effects of network transmission link imperfections and real-time limitations of video data result in multiple challenges for video transmission in massive millimeter wave (mmwave) network. The challenges, together with the increasing expectations of users for good-quality videos, further extend the complexity particularly in dense areas. To address the issues this work proposes perceptual quality aware video encoding (PQAVE) scheme in two phases: in phase 1 an effective perceptual quality encoding method leveraging low-rank approximation to reduce overall video size is done. In phase 2, a novel bitstream optimization technique is introduced to improve perceptual quality of videos. Experiment are conducted using standard video dataset show the proposed PQAVE model attain better bit error rate (BER), symbol error rate (SER), error vector magnitude (EVM), and coding gain in comparison with existing perceptual video encoding scheme.

Corresponding Author:

Aruna Kumar Bananaik Thulajanaik
Department of Computer Science and Engineering, Gowdara Mallikarjunappa Institute of Technology
Davangere, Karnataka, India
Email: arunakumarbt@gmail.com

1. INTRODUCTION

Video applications are expanding in the direction of giving more accurate and engaging visuals in conjunction with the technological advances in the fields of capturing, presenting, and computation. Volumetric video-based augmented-reality (AR) as well as the virtual-reality (VR) systems are growing in popularity due to the increased accuracy and interactivity they offer in comparison to traditional methods. Nevertheless, the massive rise in the amount of data of these videos poses a serious problem for video streaming/transmission [1], processing and storage. While fifth generation (5G) networks continue to expand their capacities, video traffic and the amount of bandwidth they consume are also growing at an exponential rate [2]. Therefore, it continues to be difficult to find an appropriate video-compression strategy to use with the goal to decrease the necessary bandwidth as well as storage space for streaming services. There have been a number of developments in video compression techniques in the past few years, including high-efficiency-video-coding (HEVC) as well as versatile-video-coding (VVC). Traditional objective metrics, like mean-squared-error (MSE), are used in the process of encoding prescribed by these standards in order to increase the peak signal-to-noise-ratio (PSNR). Nevertheless, PSNR as well as the MSE do not accurately portray the qualities of the human-visual-system (HVS). Hence, it has been proposed to use perceptual-video-coding (PVC) techniques to get rid of perceptual repetitions that are not observed by the HVS [3], [4].

When it comes to PVC methods, region-based-bitrate/quality allotment represents one of the greatest effective and popular options. These methods capitalize on the reality that HVS sharpens in on a small number of region-of-interests (RoI) while blurring the background. Using these techniques, the quality of the image is not compromised while additional bits are allocated to the important parts of the image and less bits are utilized for the background. Region-based quality allotment can help with the massive volume of video data broadcast, which is useful for programs like cloud-gaming (CG) [5] as well as 360° video [6]. One of the most difficult aspects of the aforementioned techniques is figuring out which level of quality provides the highest possible output at the lowest possible bitrate whilst yet maintaining the desired level of perceptual quality. It is important to keep in mind that RoIs do not always correspond to squared blocks, but they can be estimated by squared blocks provided the level of detail is high sufficiently as well as the size of the block is sufficiently tiny (as in HEVC-encoding). For PVC, for instance, the map of saliency is divided into blocks that are 64 by 64 pixels in [7] and [8]. Just-noticeable difference (JND) assessment at the block level is further justified by this consideration [9].

The JND depends on the idea that people are able to distinguish between a few distinct levels of quality, as opposed to a quality that varies constantly across the entire range that comprises the quantization-parameter (QP) [10]. To this end, JND-based methods seek to identify the highest distortion levels undetectable in the HVS and then utilize this information to remove the largest possible number of redundant perceptual signals without degrading the visual experience [11]. The quality factor (QP) linked to this degree of distortions is commonly known to represent the JND level. That's why JND-based methods can provide the best feasible compression while maintaining a high enough quality for human perception [7], [8]. Several significant problems with the existing JND-based PVC approaches are described at length in section 2. For the most part, present JND techniques fail to integrate visual perception techniques alongside JND, simply anticipate the level of JND for every frame or certain individual frames in a video-sequence [10], ignoring the various perceptual significance of video areas inside every frame [12], ignoring the effect of temporal video dynamics, and consider design using spatial feature only [13] which necessitates longer encoding periods [14]. In addressing this paper introduce a two-phase model. The phase 1, is focused on jointly exploiting both temporal and spatial power spectrum relationship of received video sequence to achieve higher coding efficiency under high density wireless network. In achieving better coding efficiency using principal component analysis (PCA) low-rank approximation is performed by taking relationship of temporal-spatial power spectrum. Then, to enhance video perception quality with minimal bits a novel perception-quality enhancement aware coding is presented; where different quantization bits are added according to noticeable distortion considering video dynamics to each frame of video sequence. The significance of the research work is given as:

- The research work introduces a novel two-phase architecture that brings good tradeoffs between reducing overall video size with enhanced perceptual quality of videos.
- The perceptual quality aware video encoding (PQAVE) is very efficient in achieving higher coding gain assuring perceptual quality considering varied videos frame and quality factor.

In section 2, studies various encoding mechanism designed in recent time to improve perceptual quality of video is presented. In section 3, the research methodology of proposed 2-phase perceptual video quality enhancement technique is presented. In section 4, the performance of perceptual video encoding scheme of proposed and existing methods is studied using different performance metrics. Lastly, the research is concluded with research significance and future work to enhance the perceptual quality.

2. LITERATURE SURVEY

Numerous schemes for video-compression or encoding were introduced in recent years with an effort to lower video-stream sizes and satisfy network efficiency needs [15]. VVC, HEVC, moving picture experts group (MPEG) -1/2 and H.264/advanced-video-coding (AVC) [16] are examples of popular video compression methods. With the goal of encoding 8K UHD video in the future, the joint-video-experts-team (JVET) was formed to establish a VVC [17]. There are various advancements in coding technology over the last few years, all aimed at increasing compression performance. A literature review on the factors (brightness adjustment, contrasting masking, patterned masking, as well as visual sensitivities) which influence the evaluation of the JND in visual perspective was presented in [18]. Lin and Ghinea [19] they compared and contrasted custom-built algorithms with machine learning techniques for JND. Athar and Wang [20], the authors surveyed image-quality-assessment (IQA) measures for two-dimensional images while presenting their results across multiple datasets. Min *et al.* [21], the researchers examined the properties of display contents images through various viewpoints of humans, systems, and other contexts. After that, they assessed quality evaluation methodologies and datasets for these features. Although this survey provides useful context, they are primarily concerned with perceptual frameworks for evaluating picture and video quality; improvements for video-coding has been neglected in this study. Zhang *et al*. [22] examined improvements for video-coding utilizing machine-learning from the following perspectives: complexity, compression-ratio, and visual-quality. In order to further enhance the performance of video-compression, Dong *et al*. [23] studied a good representation deep-learning centered video-coding (DLVC) methods for key-coding functions like deep inter/intra anticipation, cross-channel anticipation, transformation, in/post-loop filtration, likelihood anticipation, and up/down-sampling. To improve video-encoding efficiency, they used learning techniques, particularly deep-learning. Nevertheless, visual similarities were not given much consideration in this study. Manjanaik *et al*. [24], they improved coding efficiency (in terms of PSNR and bitrate) by developing a new method (Gaussian pulse) for Intra frame coding with diagonal down left intra prediction mode.

Hurst *et al*. [25], characteristics and difficulties of building 3D virtual-conferencing solutions taking the user-experience into account were investigated; the paper focuses on the utilization of virtual-reality for communicating with display-based 3D conferencing environment. One potential approach to resolving the issues posed by the bandwidth requirements of 3D transmission of video involves the transmission of mixed-resolution, multi-view videos. Tang and Zhang [26], they developed a method for decreasing delay simultaneously synchronizing audio/video feeds from numerous peers in real-time. This technique enables the deployment of internet-based conversational systems that combine and rebroadcast mixed-live conversational feeds across multiple participants to numerous viewers. Gandam and Sidhu [27], an approach to low-bitrate HEVC video-transmission using an adaptable deblocking filtering constructed using fuzzy logic is given. To successfully eliminate quantized noise, blocked noise, and edge outliers from HEVC decoding videos, they devised a four-step-fuzzy-based adaptable deblocked filtering decision-making method. Using simulation, they were able to prove that their approach worked. According to their findings, both the PSNR as well as subjective evaluation of videos created using their approach has been significantly improved. Malekzadeh [28], a two-stage approach for mobile video-transmission has been presented, consisting of video-related-settings (VRS) as well as network-related-settings (NRS). The researcher employed H.2.64, MPEG-4, H.265, VP9 and VP8, five different transcoding methods, to evaluate the importance of the lines of transmission and transcoding methods for high-quality video transmission with respect to quality of service and quality of experience. The reported experimental findings demonstrate that subscribers of LTE networks typically attain a higher data throughput, structural similarity index measure (SSIM), PSNR as well as the least loss-rate independent of the video compression techniques employed. Prashnani *et al*. [29], the researchers modeled a novel massive dataset annotated using the likelihood that people will choose a particular picture over other. Islam *et al.* [30], the researchers develop an approach for real-time beneath the water image augmentation using conditionally generative-adversarial-networks (GAN). They design a function that is objective to monitor the adversary training, and this function takes into account various factors such as global-content, hue, localized appearance, and visual appearance of the image. They also conduct a series of both quantitative and qualitative assessments which provide support for the idea that the suggested approach can improve underwater picture quality through matched and unmatched training. Moreover, the enhanced images show superior performances in comparison with the existing traditional approaches developed for detecting submerged objects, estimating human poses, and predicting saliency.

Liu *et al*. [31], they have presented a new self-supervised motion-perception (SMP) approach as well as a pretext-task for the purpose of learning spatial-temporal representations. The SMP approach relies on collaboratively capturing information pertaining to movement dynamics as well as appearances using video footage of varying temporal-resolutions using discriminating as well as generated perception of motion methods. Further, they present differential as well as convolutional motion-attention (MA), that pushes the generated perception of motion method mainly focused on movement-related physical appearance, and additionally use multiple-granularity perceptions (MG) for extracting correct movement dynamics, all with the goal of improving their collaboration. In order to improve visual-coding, the researchers of [32] suggest using collaborative-perception. In order to replicate the eye's response features in an adaptable manner, this method initially isolates the multi-modal spatial-temporal aspects of the source video. Furthermore, it takes the input-stimulus and utilizes the fundamental functionality to create a stimulus-matrix that incorporates several sensory modalities. A spike sequence is generated when the stimulus-matrix is reformed by filtering both downstream and upstream.

The local-global-frequency based on features method (LGFM) is a highly innovative as well as a successful image-quality-assessment (IQA) method developed in [33] for HDR-images that uses frequencies difference as a primary IQA metric. The Butterworth as well as the Gabor filtering technique are put in place upon the brightness element contained in the HDR-image to acquire both global and local frequencies characteristics, respectively. In order to get the projected overall quality rating, the frequencies characteristics go through a similarity assessment and characteristic pooling approach. Wan *et al*. [34], S3D H.264/AVC-compressed video is used to probe how the human-visual system (HVS) handles perception of depth as well as its associated features. The findings show that there is a frequency and orientation-dependent HVS reaction in perception of depth, as well as that video coding-induced abnormalities can lead to a reduction or fluctuation in perception of depth. Huang *et al*. [35], a human-machine-friendly video-compression

(HMFVC) technique is proposed to meet the needs of humans as well as viewers and automated systems. To begin with, they suggest a learnt-semantic-representation (LSR) technique for acquiring semantic data from adjacent frames in time. LSR has potential applications in visual evaluation for computer perception as well as signal restoration for consumption by users. Second, based on the suggested LSR, they create an approach for optimizing video-compression in a way that simultaneously improves user perceptual quality, computer evaluation accuracy, as well as compression-efficiency. Finally, a high-mobility, low-complexity (HMFVC) codec is created, that outperforms both conventional video-codecs and learning video-compression methods in terms of motion recognition efficiency and reconstructive quality [36].

3. PROPOSED METHOD

3.1. System model

This work presents a perceptual quality aware video encoding (PQAVE) technique for achieving bringing good tradeoffs between improved video reconstruction error reduction, higher coding efficiency and enhanced video reconstruction quality in massive multiple-input multiple-output (MIMO) wireless network. The work uses a similar architecture as defined [14]; Thulajanaik and Manjanaik [14] focused in reducing the video sequence size and transmit the video in MIMO wireless network. However, the encoding model presented in [14] failed to consider perceptual quality enhancement. This paper introduces a novel bit-optimization design that enhance the perceptual quality of video with limited amount bits leveraging tempo spatial power spectrum and sensitivity measure. Here the video sequence transmitted by different user are pre-coded using fast fourier transformation and cycle prefix code is appended to the video sequence. The remote radio head (RRH) receives time-domain orthogonal frames with varying noise considering video dynamics interference with other users.

3.2. Perceptual quality aware video encoding technique-phase 1

In this section, a perceptual quality aware video encoding technique has been presented which makes use of the optimal encoding error detection method for enhancing the efficiency of the compression while simultaneously enhancing the quality of the bitstreams reconstruction at the receiving end. The fronthaul compression technique for the MIMO networks which comprises of different users and multiple RRHs in this work has been achieved by using the time-domain video-sequence. The RRHs comprises of O antennas each of which is responsible for receiving the video-sequence from the multiple users. The following equation is used for representing the bitstreams of the video-sequence a_o received by the o^{th} antenna:

$$a_o[p] = \sum_w z_w[p] * j_{o,w}[p] + y_o[p] \quad o \in \{1,2,\dots,O\}, p \in \{0,1,2,\dots\} \tag{1}$$

where, z_w is used to represent the frames of the orthogonal frequency-division multiplexing (OFDM) as bitstreams for the w^{th} user. The noise characteristic of the channel assigned to the w^{th} user by the o^{th} antenna is defined using the expression $j_{o,w}$. Further, for the o^{th} antenna, the Gaussian-Noise is represented using the y_o, and the convolution operator is represented using $(*)$. The above equation can be used to represent the received video-sequence as the matrix $A \in \mathbb{E}^{P*O}$. The P is used for representing the number of frames which are needed to be compressed and O is used for representing the size of the RRH antenna. The given below equation approximates the matrix A utilizing a low-rank computational technique [15], and it is noticed that the column $a_k, k \in \{1,2,\dots,O\}$ are highly correlated with one another.

$$A = A_0 + G \tag{2}$$

After the approximation of the low-rank technique, it yields a noise-free matrix defined by $A_0 \in \mathbb{E}^{P*O}$. $G \in \mathbb{E}^{P*O}$ is a used for representing the realistic noise-matrix that is defined by the use of the Gaussian-representation. The input bitstreams a of the video-sequence, the behavior of the channel k, and $G \in \mathbb{E}^{P*O}$, all these information is comprised in the approximated matrix. The goal of the PQAVE technique is to lower the matrix-size A by adopting a low-rank approximation computational technique. This will ensure that only a smaller number of video-sequences as bitstreams will need to be communicated over a MIMO network. In the preceding equation, the A_0 is used for representing the matrix having the low-rank of the size $P * O$. Also, $P \gg O > N$. From this the matrix A can be defined using (3).

$$A'' = \underset{\mathcal{R}ank(\hat{A})=N}{\operatorname{argmin}} \left\|A - \hat{A}\right\|H \tag{3}$$

By utilizing the Frobenius distance-normalization method represented as $\|.\|H$ in the preceding equation, the matrix $\widehat{A}$ can be acquired. The $\widehat{A}$ is represented using (4):

$$A'' = W_N \beta_N X_N^J \tag{4}$$

with,

$$W_N = [w_1 \quad w_2 \quad \dots \quad w_N]$$

$$X_N = [x_1 \quad x_2 \quad \dots \quad x_N]$$

$$\beta_N = \mathcal{D}iag[\alpha_1 \quad \alpha_2 \quad \dots \quad \alpha_N]$$

where A'' is used for representing the singular-value-decomposition (SVD) [15] which further is utilized for defining the A'' inside the conjugated transpose which is denoted using $(.)^J$. The $w_k \in \mathbb{E}^P$ and $x_j \in \mathbb{E}^P$ which is used for denoting the left and right eigen-vectors respectively. The α_N is used for representing the values which are singular diagonally.

By utilizing the realistic noise matrix, ranked matrix N can be attained. After this the proposed PQAVE technique utilizes the SVD to extract the principal components X_N from the N which has identical eigen-vectors. Further, it is used for multiplying it to the X_N to get a transformed matrix A. Lastly, the extracted video-frame vector a_k having the relation with the actual space O and novel space N is said to have highly uncorrelated with the video-sequence under consideration. The given (5) describes the optimized matrix R_N where $R_N \in \mathbb{E}^{P*N}$:

$$R_N = AX_N = W_N \beta_N \tag{5}$$

where, k in the R_k is described as $k \in \{1,2, \dots, N\}$ which is used for representing the k^{th} column for the R_N. Also, the k^{th} column is the a_k which has been decorrelated. The following equation defines how the base-band unit collects the expected low-rank-matrix A'' is given (6).

$$A'' = R_N X_N^J = W_N \beta_N X_N^J. \tag{6}$$

Further, the ranked matrix A_0 comprises of the input video-frames which contains noise, hence, by eliminating the $(O - N)$, the noise can be removed. By reconstructing the matrix A'' described in (4), we can evaluate the efficacy of the noise removal process using (7).

$$A'' = A_0 + \delta \tag{7}$$

In the preceding equation, δ is the detected error in the video-frame. δ also represents the low-ranked approximated matrix. The δ parameter comprises of the detected error from the residual noise and the data loss from the video-sequence during the computation of the low-ranked approximated matrix. Once encoding is completed it induce visual perception video dynamic i.e., distortion in each frame. The work identifies this distortion and perform bit optimization to enhance the perceptual quality of video sequence in next section.

3.3. Perceptual quality enhancment technique-phase 2

In this section, the perceptual quality enhancement (PQE) technique has been presented. The PQE technique improves the performance of video encoding by quantizing the video-sequence matrix R_N and X_N. Each element of the matrix R_N and X_N containing source and reference video-frames is subjected to this quantization process and the quantizers used in this work are generally assumed to be having uniform quantizers. For attaining an enhanced perceptual video reconstruction quality, in this work, the matrix R_N is considered to be the source components $Src(R_N)$ or reference components $Ref(R_N)$. For the matrix X_N also, the same is considered. Given that the matrix R_N is considered to be the linear transformed-matrix of the matrix A, by using the low-ranked approximated encoding component of the matrix X_N, each transformed video-sequence variable r_k and its accompanying low-ranked approximated encoding variable x_k can be quantized individually, allowing us to employ a smaller set of quantization bit-streams while maintaining a high perceptual quality compression. This process is called as high perceptual transformation coding. The challenge in this situation is to find the optimal number of bit-streams for every quantizer, such that the overall quantization error is minimized while yet achieving the desired perceptual quality assured compression ratio. The bit-streams optimization constraint for a targeted compression ratio is used in conjunction with our technique to solve the bit-streams allocation challenge for the individual quantization in order to maximize the overall coder's perceptual quality performance.

For solving the perceptual quality assured bit-streams optimization issue, in this technique our goal is to decrease the total weighted distortion variable F of the quantizing matrix R_N for a particular bitstreams optimization constraint D in respect of the parameter d which is denoted as $d = [d_1\ d_2\ \dots\ d_N]$. The d_k is used for defining the number of quantization-bitstreams for the transformed video-sequence variable r_k. The eigenvalues ε_k are used for providing weights to the perceptual video quality distortion variable F. Moreover, for finding the d which will help to reduce the $F(d)$, we solve the perceptual quality tradeoffs operation using (8).

$$\begin{aligned} &\min F(d) = \textstyle\sum_{k=1}^{N} \varepsilon_k\, Y_k(d_k) \\ &S.T\ \textstyle\sum_{K=1}^{N} d_k = D, d_k \in \mathbb{A}_+ \end{aligned} \tag{8}$$

In the preceding equation, $\mathbb{A}_+$ is used for representing the non-negative integer. $Y_k(d_k)$ is used for defining the mean-squared-error experienced during the quantization process for the transformed video-sequence variable r_k having d_k bitstreams. Consider that the R_k is a random parameter for video-sequence samples in the transformed video-sequence variable r_k. From this the $Y_k(d_k)$ can be approximated using the given (9).

$$Y_k(d_k) \approx (R_k) j_k 2^{-2d_k} \tag{9}$$

The $diff(R_k)$ is the difference of the constant j_k and the random parameter R_k. Further the j_k is evaluated using the pdf $f\bar{r}_k(r)$ from the normalized random parameter $\bar{R}_k = R_k\sqrt{diff(R_k)}$ using (10).

$$j_k = \frac{1}{\alpha}\left\{\int_{-\infty}^{\infty} [f\bar{r}_k(r)]^{1/3} dr\right\}^3. \tag{10}$$

The approximated $diff(R_k)$ as well as the j_k can be evaluated using the empirical distribution of video sequence samples in the transformed video-sequence variable r_k. α is the empirical set value. A perceptual quality aware bitstreams optimization model is given in Algorithm 1 for solving (8). In every iteration, for the quantizer having the highest weighted mean-square error $\varepsilon_k Y_k(d_k)$, the Algorithm 1 iteratively maximizes the bitstreams d_k by 1 until all the D bits are allocated. This perceptual quality aware bitstreams optimization algorithm is shown in Algorithm 1.

Algorithm 1. Perceptual quality aware bitstreams optimization model
Step 1. Start
Step 2. Initialize $d_k = 0\ \ \forall\, k \in \{1,2,\dots N\}$
Step 3. Input D, ε_k, Y_k, $k \in \{1,2,\dots N\}$
Step 4. while $\sum_{j=1}^{N} d_k < D$ do
Step 5. Find $m = \text{argmax}_{\,k\in\{1,2,\dots N\}}\, \varepsilon_k\, Y_k(d_k)$
Step 6. $d_m = d_m + 1$
Step 7. end while
Step 8. Return $d = [d_1, d_2, \dots, d_N]$
Step 9. Stop

As the low-ranked approximated encoding parameter x_k has an identical weight to the eigenvalues ε_k as it is correlated with the transformed video-sequence variable r_k, the proposed technique allocates similar d_k perceptual quality aware bitstreams optimization for the quantized x_k. In Figure 1, the complete process for the compression by using the low-rank approximated and perceptual quality aware transformed coding with bitstreams optimization has been given. When the quantization process is completed, the quantized x_k and r_k are sent to base-band unit through the accessible front-haul link. For the overall number of video-sequence as the bitstreams for the quantization of the x_k and the r_k, the presented video perceptual quality aware quantization technique decreases the total count of the overall quantization bit-streams from $(O + P)Nd_{UF}$ to $(O + P)\sum_{k=1}^{N} d_k$, where, d_{UF} is used for denoting the total count of the standard quantization-bitstreams. Hence, from this the compression-gain of the perceptual quality aware transformation coding using the bit-streams optimization can be evaluated using (11).

$$\mathcal{CR}_{\mathcal{P}erception} = \frac{Nd_{UF}}{\sum_{k=1}^{N} d_k}. \tag{11}$$

By not taking into consideration the total number of bits which have to be transmitted for the quantization-side information (QSI), the total compression-ration for both the low-ranked approximated and perceptual quality aware transformed coding with bitstreams optimization changes to $\mathcal{CR}_{\mathcal{P}}\ \mathcal{CR}_{\mathcal{P}erception}$. The QSI has all the ranges for the quantization of the d_k, x_k, r_k, and raked information. Every quantization range can be represented by a separate d_{UF} bitstream, and all of this information can be communicated using a $2Nd_{UF}$ video-sequence as a bitstream. Since the largest value of d_k is d_{UF}, and ranked information could be expressed using $\log_2 O$ bits, a video-sequence including information for every d_k, $k \in \{1, 2, \cdots, N\}$ takes $N\log_2 d_{UF}$ bitstreams. Hence, from this the compression-ratio can be evaluated using:

$$\mathcal{CR}_{\mathcal{P}\_\mathcal{P}erception} = \frac{OP\ d_{UF}}{(O+P)\sum_{k=1}^{N} d_k + d_{QSI}} \tag{12}$$

with,

$$d_{QSI} = N(2d_{UF} + \log_2 d_{UF}) + \log_2 O \tag{13}$$

in the preceding equation, the d_{QSI} is the count of video sequence as a bitstreams for QSI. For the non-compression situation, the worst-case compression-ratio is given by (11), which considers there is no QSI which has to be transmitted. Given that d_{QSI} is nearly equivalent to the ranked N, and that P is significantly higher than N from the presumption; $P \gg O > N$, d_{QSI} becomes insignificant relative to the remaining terms in the (12). Hence, in this case, the total coding gain changes to $\mathcal{CR}_{\mathcal{P}\_\mathcal{P}erception} \approx \mathcal{CR}_{\mathcal{P}}\ \mathcal{CR}_{\mathcal{P}erception}$. In the next section, the results and analysis section has been presented which evaluates the proposed technique using the Foreman video-sequence and has been compared with the other existing techniques.

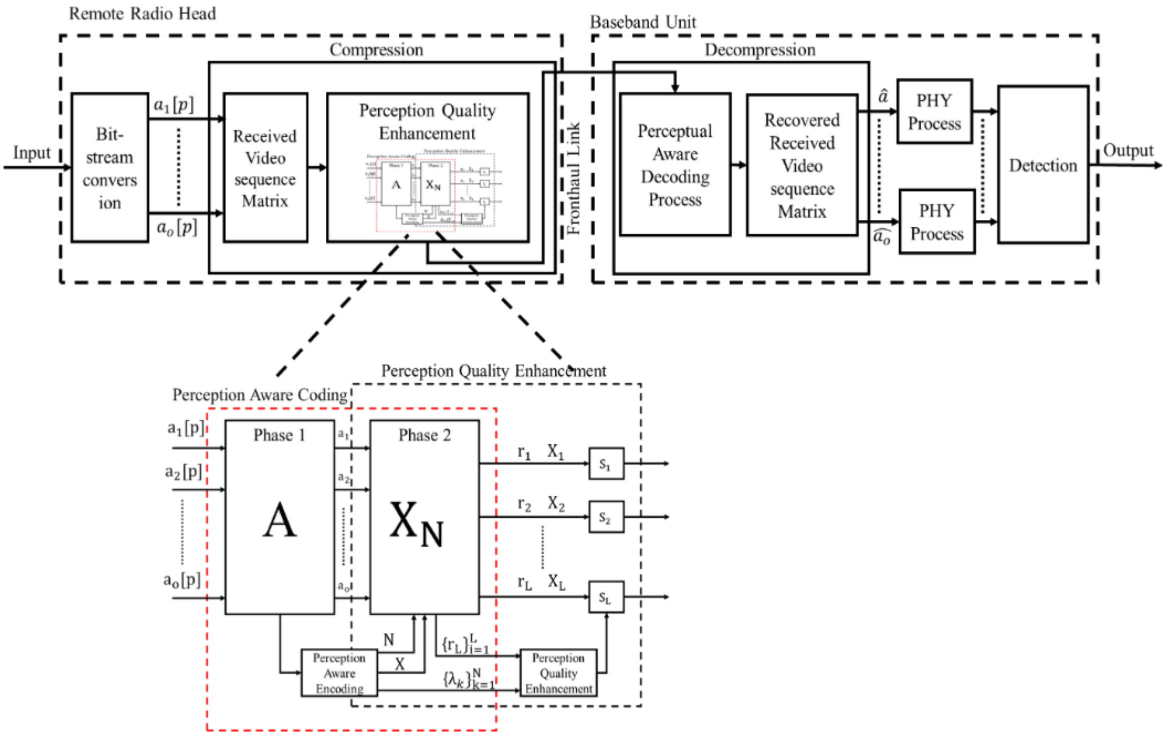

Figure 1. Architecture of proposed perceptual quality aware bitstream optimization model in massive MIMO network

4. RESULTS AND DISCUSSIONS

In this section, the results for the proposed technique have been discussed. The proposed technique has been coded in Python language and the MATLAB environment has been used for the implementation. The experimentation has been done on the system running on the Windows 10 having Intel i7 quad core processor having 16 Gigabytes of RAM. The proposed technique has been evaluated using the bit error rate, symbol error rate, error vector magnitude and coding gain. This work takes into account an experimental configuration arrangement similar to that presented in [37]. Table 1 provides details on the parameters used for the validation of the proposed PQAVE technique. The JCV video sequence dataset [37] have been considered for validating the proposed perceptual quality enhancement technique. A sample sequence of JCV dataset is given in Figure 2. This dataset contains 30 5-second video sequences of resolution 1920×1080 pixels from different genres, semantics, and features.

Table 1. Simulation parameter

Parameters	Values
Coverage area	200 m * 200 m
Transmission BW	10 MHz
Number of resource blocks	50
Radio antennas height	6 m
User antennas height	1 m
Sub-carrier spacing	3.5 Ghz
FFT Size	1024
CP Lengths	72
No. of OFDM symbols per subframe	7
Modulation	QPSK, 16-QAM, 64-QAM
Channel	AWGN
Number of RRH antennas	32

Figure 2. A sample sequence of JCV dataset [37]

4.1. Communciation perforamcne

Bit error rate (BER) is a metric used to measure the percentage of video-frames that are correctly received at the receiver end in a wireless network. It is defined as the ratio of the number of bits received in error to the total number of bits transmitted. The lower value denotes that the performance has been improved. Further, in this experimentation, the SNR has been varied from -6 to +6. Also, in the experimentation, the size of the RRH antenna is considered to be 32. The results attained for the proposed PQAVE technique and the existing perception video encoding (EPVE) [37] has been given in the Figure 3. The results show that the PQAVE technique has enhanced its performance and has attained an improvement of 32.55% in comparison to the EPVE technique for the BER.

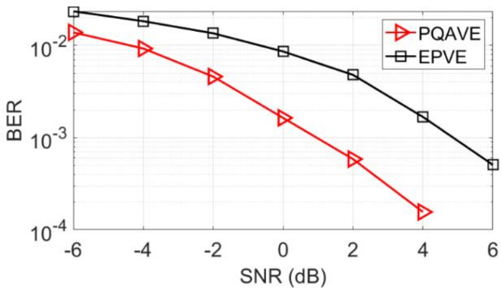

Figure 3. BER vs SNR for 32 RRHx

Symbol-error-rate (SER) is a metric used to measure the percentage of the video-frames that are incorrectly received at the receiver side in a wireless network. It represents the probability of a video-frame being incorrectly received at the receiver. The lower value denotes that the performance has been improved. Further, in this experimentation, the SNR has been varied from -6 to +6. Also, in the experimentation, the size of the RRH antenna is considered to be 32. The results attained for the proposed PQAVE technique and the EPVE technique [37] has been given in the Figure 4. The results show that the PQAVE technique has enhanced its performance and has attained an improvement of 86.22% in comparison to the EPVE technique for the SER.

The error-vector-magnitude (EVM) is a metric which is utilized for detecting the error performance of the compression technique. The lower value denotes that the performance has been improved. Further, in this experimentation, the SNR has been varied from -6 to +6. Also, in the experimentation, the size of the RRH antenna is considered to be 32. The results attained for the proposed PQAVE technique and the EPVE technique [37] has been given in the Figure 5. The results show that the PQAVE technique has enhanced its performance and has attained an improvement of 9.04% in comparison to the EPVE technique for the SER.

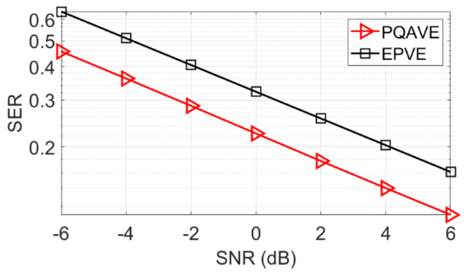

Figure 4. SER vs SNR for 32 RRH

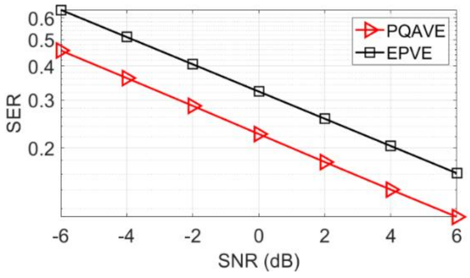

Figure 5. EVM vs SNR for 32 RRH

4.2. Coding gain

The coding gain is used for measuring the perceptual quality transmitted video frames. The higher the value the better the perceptual quality and performance. The and Figure 6 shows the coding gain performance achieved using EPVE [37] and PQAVE under varied frame size. The Figure 7 shows the coding gain performance achieved using EVQE and PQAVE under varied quality factor. From both Figures 6 and 7 we can conclude that the PQAVE achieve much perceptual quality in comparison with EVQE.

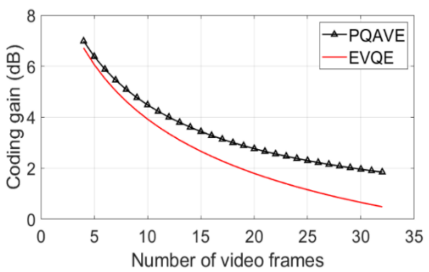

Figure 6. Coding gain with varying frame size

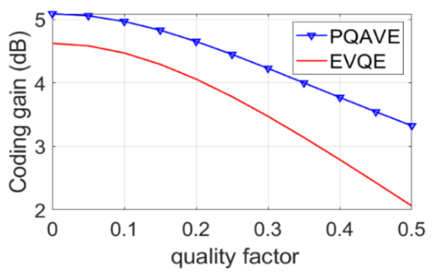

Figure 7. Coding gain with varying quality factor

5. CONCLUSION

This research presents an effective approach of video encoding that assure perceptual quality for large MIMO networks. The PQAVE technique comprises of two phases, the first phase presents an efficient encoding using low-ranked approximation technique that keeps the bits size as low as possible with minimal bit error under dynamic SNR conditions; the phase two, enhance the perceptual quality of image with minimal bits at the block-level of each frame through enhanced quantization technique. Experiment is carried out utilizing the standard video-sequence dataset while working within a large MIMO network. Comparing the PQAVE to the EPVE technique reveals significant improvements in SER, BER, EVM, and compression gain. The future work would consider inducing diverse mobility model representing realistic fast fading environment to study PQAVE and perform further optimization.

REFERENCES

[1] R. Lee, S. I. Venieris, and N. D. Lane, "Deep neural network based enhancement for image and video streaming systems: A survey and future directions," *ACM Computing Surveys*, vol. 54, no. 8, pp. 1–30, Nov. 2022, doi: 10.1145/3469094.

[2] "Cisco annual internet report (2018–2023) white paper," Cisco, 2020, [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html.

[3] T. Tian, H. Wang, L. Zuo, C. C. J. Kuo, and S. Kwong, "Just noticeable difference level prediction for perceptual image compression," *IEEE Transactions on Broadcasting*, vol. 66, no. 3, pp. 690–700, Sep. 2020, doi: 10.1109/TBC.2020.2977542.

[4] H. Liu *et al.*, "Deep learning-based picture-wise just noticeable distortion prediction model for image compression," *IEEE Transactions on Image Processing*, vol. 29, pp. 641–656, 2020, doi: 10.1109/TIP.2019.2933743.

[5] H. Ahmadi, S. Zadtootaghaj, F. Pakdaman, M. R. Hashemi, and S. Shirmohammadi, "A skill-based visual attention model for cloud gaming," *IEEE Access*, vol. 9, pp. 12332–12347, 2021, doi: 10.1109/ACCESS.2021.3050489.

[6] C. L. Fan, S. C. Yen, C. Y. Huang, and C. H. Hsu, "Optimizing fixation prediction using recurrent neural networks for 360° video streaming in head-mounted virtual reality," *IEEE Transactions on Multimedia*, vol. 22, no. 3, pp. 744–759, Mar. 2020, doi: 10.1109/TMM.2019.2931807.

[7] X. Jiang, T. Song, D. Zhu, T. Katayama, and L. Wang, "Quality-oriented perceptual HEVC based on the spatiotemporal saliency detection model," *Entropy*, vol. 21, no. 2, p. 165, Feb. 2019, doi: 10.3390/e21020165.

[8] S. Zhu and Z. Xu, "Spatiotemporal visual saliency guided perceptual high efficiency video coding with neural network," *Neurocomputing*, vol. 275, pp. 511–522, Jan. 2018, doi: 10.1016/j.neucom.2017.08.054.

[9] S. Nami, F. Pakdaman, and M. R. Hashemi, "Juniper: A jnd-based perceptual video coding framework to jointly utilize saliency and JND," in *2020 IEEE International Conference on Multimedia and Expo Workshops, ICMEW 2020*, Jul. 2020, pp. 1–6, doi: 10.1109/ICMEW46912.2020.9106036.

[10] M. Takeuchi *et al.*, "Perceptual quality driven adaptive video coding using JND estimation," in *2018 Picture Coding Symposium (PCS)*, Jun. 2018, pp. 179–183, doi: 10.1109/PCS.2018.8456297.

[11] X. Zhang, H. Wang, and T. Tian, "Perceptual video coding with block-level staircase just noticeable distortion," in *Proceedings - International Conference on Image Processing, ICIP*, Sep. 2019, pp. 4140–4144, doi: 10.1109/ICIP.2019.8803454.

[12] D. Yuan, T. Zhao, Y. Xu, H. Xue, and L. Lin, "Visual JND: a perceptual measurement in video coding," *IEEE Access*, vol. 7, pp. 29014–29022, 2019, doi: 10.1109/ACCESS.2019.2901342.

[13] B. Mallik, A. Sheikh-Akbari, P. B. Zadeh, and S. Al-Majeed, "HEVC based frame interleaved coding technique for stereo and multi-view videos," *Information (Switzerland)*, vol. 13, no. 12, p. 554, Nov. 2022, doi: 10.3390/info13120554.

[14] A. K. B. Thulajanaik and N. Manjanaik, "An adaptive compression technique for efficient video reconstruction in future generation wireless network," *International Journal of Intelligent Engineering and Systems*, vol. 16, no. 2, pp. 497–505, Feb. 2023, doi: 10.22266/ijies2023.0430.40.

[15] Z. Wang, Y. Gong, and K. Yang, "Perception of temporal quality fluctuation for HEVC coded surveillance videos," in *ACM International Conference Proceeding Series*, Mar. 2022, vol. Par F18047, pp. 110–116, doi: 10.1145/3532342.3532355.

[16] B. Bross *et al.*, "Overview of the versatile video coding (VVC) standard and its applications," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 10, pp. 3736–3764, Oct. 2021, doi: 10.1109/TCSVT.2021.3101953.

[17] B. Bross, J. Chen, J. R. Ohm, G. J. Sullivan, and Y. K. Wang, "Developments in international video coding standardization after AVC, with an overview of versatile video coding (VVC)," *Proceedings of the IEEE*, vol. 109, no. 9, pp. 1463–1493, Sep. 2021, doi: 10.1109/JPROC.2020.3043399.

[18] J. Wu, G. Shi, and W. Lin, "Survey of visual just noticeable difference estimation," *Frontiers of Computer Science*, vol. 13, no. 1, pp. 4–15, Feb. 2019, doi: 10.1007/s11704-016-6213-z.

[19] W. Lin and G. Ghinea, "Progress and opportunities in modelling just-noticeable difference (JND) for multimedia," *IEEE Transactions on Multimedia*, vol. 24, pp. 3706–3721, 2022, doi: 10.1109/TMM.2021.3106503.

[20] S. Athar and Z. Wang, "A comprehensive performance evaluation of image quality assessment algorithms," *IEEE Access*, vol. 7, pp. 140030–140070, 2019, doi: 10.1109/ACCESS.2019.2943319.

[21] X. Min *et al.*, "Screen content quality assessment: overview, benchmark, and beyond," *ACM Computing Surveys*, vol. 54, no. 9, pp. 1–36, Dec. 2022, doi: 10.1145/3470970.

[22] Y. Zhang, S. Kwong, and S. Wang, "Machine learning based video coding optimizations: A survey," *Information Sciences*, vol. 506, pp. 395–423, Jan. 2020, doi: 10.1016/j.ins.2019.07.096.

[23] L. I. U. Dong, L. I. Yue, L. I. N. Jianping, L. I. Houqiang, and W. U. Feng, "Deep learning-based video coding: a review and a case study," *ACM Computing Surveys*, vol. 53, no. 1, pp. 1–35, Jan. 2020, doi: 10.1145/3368405.

[24] N. Manjanaik, B. D. Parameshachari, S. N. Hanumanthappa, and R. Banu, "Intra frame coding in advanced video coding standard (H.264) to obtain consistent PSNR and reduce Bit rate for diagonal down left mode using Gaussian pulse," *IOP Conference Series: Materials Science and Engineering*, vol. 225, no. 1, p. 012209, Aug. 2017, doi: 10.1088/1757-899X/225/1/012209.

[25] W. Hurst, A. Withington, and H. Kolivand, "Virtual conference design: features and obstacles," *Multimedia Tools and Applications*, vol. 81, no. 12, pp. 16901–16919, May 2022, doi: 10.1007/s11042-022-12402-4.

[26] D. Tang and L. Zhang, "Audio and video mixing method to enhance WebRTC," *IEEE Access*, vol. 8, pp. 67228–67241, 2020, doi: 10.1109/ACCESS.2020.2985412.

[27] A. Gandam and J. S. Sidhu, "Fuzzy based adaptive deblocking filters at low-bitrate HEVC videos for communication networks," *Computers, Materials and Continua*, vol. 66, no. 3, pp. 3045–3063, 2021, doi: 10.32604/cmc.2021.013663.

[28] M. Malekzadeh, "Perceptual service-level QoE and network-level QoS control model for mobile video transmission," *Telecommunication Systems*, vol. 77, no. 3, pp. 523–541, Jul. 2021, doi: 10.1007/s11235-021-00777-y.

[29] E. Prashnani, H. Cai, Y. Mostofi, and P. Sen, "PieAPP: Perceptual image-error assessment through pairwise preference," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 2018, pp. 1808–1817, doi: 10.1109/CVPR.2018.00194.

[30] M. J. Islam, Y. Xia, and J. Sattar, "Fast underwater image enhancement for improved visual perception," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 3227–3234, Apr. 2020, doi: 10.1109/LRA.2020.2974710.

[31] C. Liu, Y. Yao, D. Luo, Y. Zhou, and Q. Ye, "Self-supervised motion perception for spatiotemporal representation learning," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–15, 2022, doi: 10.1109/TNNLS.2022.3160860.

[32] L. An, Z. Yan, W. Wang, J. K. Liu, and K. Yu, "Enhancing visual coding through collaborative perception," *IEEE Transactions on Cognitive and Developmental Systems*, pp. 1–1, 2022, doi: 10.1109/TCDS.2022.3203422.

[33] Y. Liu, Z. Ni, S. Wang, H. Wang, and S. Kwong, "High dynamic range image quality assessment based on frequency disparity," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1–1, 2023, doi: 10.1109/tcsvt.2023.3237702.

[34] W. Wan *et al.*, "Depth perception assessment of 3D videos based on stereoscopic and spatial orientation structural features," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1–1, 2022, doi: 10.1109/TCSVT.2022.3165970.
[35] Z. Huang, C. Jia, S. Wang, and S. Ma, "HMFVC: a human-machine friendly video compression scheme," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1–1, 2022, doi: 10.1109/TCSVT.2022.3207596.
[36] K. Uddin, Y. Yang, and B. T. Oh, "Double compression detection in HEVC-coded video with the same coding parameters using picture partitioning information," *Signal Processing: Image Communication*, vol. 103, p. 116638, Apr. 2022, doi: 10.1016/j.image.2022.116638.
[37] S. Nami, F. Pakdaman, and M. R. Hashemi, "BL-JUNIPER: A CNN-assisted framework for perceptual video coding leveraging block-level JND," *IEEE Transactions on Multimedia*, pp. 1–16, 2022, doi: 10.1109/TMM.2022.3187259.

BIOGRAPHIES OF AUTHORS

Aruna Kumar Bananaik Thulajanaik Assistant Professor Department of Computer Science and Engineering, G. M. Institute of Technology, Davanagere, Karnataka, India. He Completed BE and M. Tech from University B. D. T College of Engineering, Kuvempu University, Shivamogga. 16 Years of Experience in teaching, interested areas are image and video processing and embedded systems. He can be contacted at email: arunakumarbt@gmail.com.

Dr. Naganaik Manjanaik Professor, Department of Electronics and Communication Engineering, University B.D.T College of Engineering, Davanagere, Karnataka, India. Graduated from University B.D.T College of Engineering, Davanagere. Kuvempu University, Shivamogga. A constituent College of VTU Belagavi. Has 22 years of experience in teaching. He completed Ph.D., from Jain University, Bengaluru. He served various positions BOS, BOE member at Visveswaraya Technological University. Published more than 15 journals and guided more than 25 PG students, presently guiding 05 Ph.D., students and 01 student is already awarded. Interested areas are image and video processing, signal processing and pattern recognition. He can be contacted at email: manjubdt2000@gmail.com.