Unequal Error Protection Aided Region of Interest Aware Wireless Panoramic Video

Yongkai Huo, Xu Wang, Peichang Zhang, Jianmin Jiang and Lajos Hanzo, *Fellow, IEEE*

***Abstract*—Panoramic video with its flawless immersive tele-presence is considered to be the near-future video format of choice, since they carry 360 degree coverage of the designated scenes. However, viewers may focus their specific attention on perfectly lip-synchronized video as part of the panoramic video scene, hence only have a peripheral vision of the remaining parts of a frame. Therefore, it is intuitive to allocate stronger protection to the panoramic video region of interest. As a solution, we propose Region Of Interest Aware Unequal Error Protection (ROI-UEP) for wireless transmission of high efficiency video code (HEVC) sequences. Specifically, the ROI of a panoramic frame may be deemed to be within the 120^0 angular range of the viewing center, which can be estimated from the viewing trajectory of a head mounted display. Then, the most appropriate unequal forward error correction (FEC) coding rates will be found for the ROI signals by minimizing the expected video distortion. Moreover, the so-called weighted peak signal-to-noise ratio (WPSNR) is proposed for evaluating the quality of the reconstructed panoramic video, where the weights of pixels are taken into account for calculating the distortion caused by the related pixels. Our simulation results show that the ROI based equal error protection (ROI-EEP) scheme substantially outperforms the EEP by a WPSNR of more than 10 dB, while the ROI-UEP scheme further improves its ROI-EEP counterpart by a WPSNR of 9.4 dB at a channel E_b/N_0 of 6 dB.**

I. Introduction

A. Panoramic Video

Panoramic video with its flawless immersive tele-presence is considered to be the near-future video format of choice, since it is capable of projecting a 360 degree coverage of the designated scenes as exemplified in Fig. 1. In [3], Schreer *et al.* reviewed the technical details of ultrahigh-resolution panoramic video production. A field-programmable gate array (FPGA) based panoramic video generation system was developed by Xu *et al.* [4]. Fu *et al.* [5] proposed a mapping scheme representing panoramic video for the sake of improving the visual quality, stability and compression efficiency. Alface *et al.* [6] characterized a personalized transmission scheme,

This work was supported in part by the National Natural Science Foundation of China (No.61702335,No.61601304), in part by the Science and Technology Innovation Commission of Shenzhen (JCYJ20170302154149766,JCYJ20170302142545828), in part by the Foundation of Shenzhen University under Grant 2016057. L. Hanzo would also like to acknowledge the financial support of the EPSRC projects EP/Noo4558/1, EP/PO34284/1, COALESCE, of the Royal Society's Global Research Challenges Fund and of the ERC's Advanced Fellow Award Quantcom.

Y. Huo, X. Wang and J. Jiang are with the National Engineering Laboratory for Big Data System Computing Technology, and the Research Institute for Future Media Computing, School of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China. (e-mail:ykhuo@szu.edu.cn, wangxu@szu.edu.cn, jianmin.jiang@szu.edu.cn)

P. Zhang is with the College of Information Engineering, Shenzhen University, Shenzhen 518060, China. (e-mail: pzhang@szu.edu.cn)

L. Hanzo is with the School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, UK. (e-mail: lh@ecs.soton.ac.uk)

Corresponding author: Dr Yongkai Huo

(a) Spherical Frame

(b) Planar Frame, 4096×2048

Figure 1: A spherical panoramic frame and its corresponding planar frame of the RaceVR sequence using the equirectangular projection [1], [2].

where the quality is controlled in spherical regions depending on their specific likelihood to be viewed during a live user interaction. The so-called Cube2Video was proposed by Zhao *et al.* [7] for navigating between cubic domains in a video-viewing mode, which eliminates the discontinuities between cube faces. In [8], tiling was applied by Gaddam *et al.* in a real-time interactive panoramic video system, which progressively increases the quality toward the point of specific region of viewer-focus, while managing to reduce the bandwidth requirement at the cost of a slight quality of experience (QoE) reduction compared to a full-resolution panoramic system. To meet the requirements of lower latency and massive data transmission both in augmented reality (AR) and virtual reality (VR) applications, a software-defined networking architecture is proposed for 5G small cell networks in [9]. In [10], the principle of scalable video coding (SVC) [11] is applied to full-panoramic video coding for the sake of reducing the required bandwidth, while a two-stage cooperative VR transmission scheme is analyzed in [12].

B. Unequal Error Protection

It is intuitive to differently protect the video bits having different importance for the sake of improved error-resilience.

Explicitly, unequal error protection (UEP) was first proposed by Masnick and Wolf [13], where a stronger forward error correction (FEC) was allocated to the more important bits than to the less important bits. Four categories of UEP techniques were reviewed in [14], namely UEP based transceivers schemes [15], packet-level FEC arrangements [16], bit-level FEC schemes [17], [18] and cross-layer operation aided solutions [19]. Here we focus our attention on the family of bit-level FEC schemes.

Marx and Farah [20] minimized the video distortion by non-uniformly allocating the redundancy imposed by a turbo code among successive video frames, which were encoded by the H.263 video codec. UEP assisted Low-density parity-check (LDPC) coded schemes were investigated in [21]. UEP based data-partitioned [22] H.264/AVC video streaming using recursive systematic convolutional (RSC) codes was investigated in [23], while UEP aided turbo coded modulation [24] was evaluated in [25]. Chang *et al.* [26] considered both the unequal importance of successive video-frames in a group of pictures (GOP) and the uneuqal significance of the diverse macroblocks (MBs) within a video frame. However, only three protection classes were employed in [26], which limits the attainable performance. Motivated by the fact that the side information (SI) values within different positions of the Wyner-Ziv (WZ) frames may have different error probability, Micallef *et al.* [27] non-uniformly allocated the FEC redundancy to these SI values for the sake of reducing the required bitrate in the context of distributed video coding (DVC) [28]. The authors of [29] applied UEP on the wireless communications of the future holographic image.

The authors of [14], [30] proposed the so-called bit-level inter-layer FEC (IL-FEC) [17] for layered wireless video relying on soft-decoded FEC, where the systematic bits of the base layer (BL) are implanted into the enhancement layers (ELs) at the transmitter. At the receiver, the implanted bits of the BL may be beneficially exploited for improving the error-resilience of the BL. In the above-mentioned IL-FEC technique of [17], the UEP philosophy was also investigated for the sake of further improving the system performance. In [18], the authors developed an algorithm for finding the optimized coding rates "on-the-fly", which optimized the IL-FEC coded system's performance. Stereoscopic video relying on the depth-map format was investigated in [31] for communication over noisy channels, where different FEC coding rates were applied both to the color and to the depth quantization parameters. The so-called power-based layer-division multiplexing scheme was studied in [32] as a means of providing UEP in digital terrestrial television (DTT) systems by adopting an information-theoretic approach. A cross-layer operation aided UEP scheme was conceived in [33] for low-complexity handheld devices. Song *et al.* [34] investigated various multilayer video representations, such as scalable videos as well as simulcast streaming, and proposed an UEP scheme for striking an improved tradeoff between the video storage and post-enhancement cost.

Recently, Zheng *et al.* [35] described a database containing video assessment scores of 70 traditional video test sequences, which are corrupted when transmitted over a wireless third generation (3G) long term evolution (LTE) network simulator. Since there is paucity of literature on panoramic video communications, we consider the scenario of transmitting panoramic video over wireless channels. We are motivated by the fact that panoramic video viewers tend to be more interested in certain parts of the panoramic frame [8] at a specific instant. Hence it is intuitive to only refresh and protect that particular part of the panoramic video frame, which is located in the region of interest (ROI) [36]. Moreover, different video signals within the ROI may also have different importance, hence UEP should be employed for the sake of improving the performance. In this treatise, for each panoramic frame, we propose ROI aware UEP (ROI-UEP) for wireless transmission of high efficiency video coded (HEVC) streaming. Specifically, each panoramic frame is divided into multiple blocks, which may have unequal importance depending on the viewers' viewing direction. We design our objective function (OF) by carefully considering these blocks' importance, their impact on the video distortion and their packet loss ratio. This OF aims for maximizing the weighted peak signal-to-noise ratio (WPSNR) of the viewers' ROI signals. By optimizing this OF relying on both the Lagrange Multiplier and Newton's down-hill method, a specific set of blocks with specific FEC coding rates may be selected and transmitted for minimizing the expected video distortion. The rationale and novelty of this paper is summarized as follows.

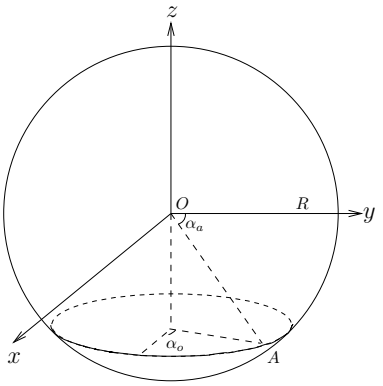
1) *We conceive an efficient wireless panoramic video streaming scheme.*
2) *The weights of panoramic pixels were calculated as the basis of UEP, where the panoramic video is expressed in planar format.*
3) *The OF is designed by considering both the weights of panoramic pixels and the related distortion under the constraint of a specific overall coding rate.*
4) *The WPSNR metric is developed for the sake of characterizing the panoramic video quality.*
5) *The OF is minimized with the aid of the Lagrange Multiplier combined with Newton's down-hill method for finding the most appropriate set of coding rates for the sake of improving the WPSNR.*
6) *The ROI-EEP scheme outperforms the conventional EEP arrangement by a channel SNR of 5 dB, while ROI-UEP is capable of further improving the ROI-EEP by a WPSNR of 9.4 dB.*

The structure of this paper is detailed below. Specifically, the equi-rectangular projection concept is briefly introduced in Section II. Section III details our system architecture, followed by optimizing the coding rates of ROI-UEP in Section IV. Section V characterizes the performance of our proposed scheme, where a RSC codec is employed for encoding the RaceVR panoramic video sequence. Finally, the paper is concluded in Section VI.

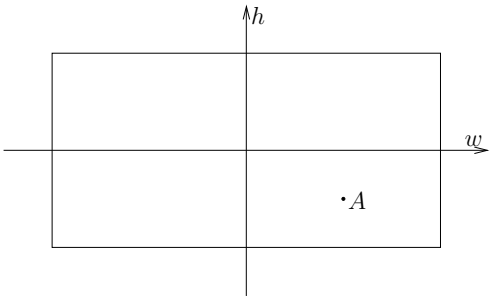
II. Equi-rectangular Projection

A number of techniques may be utilized for creating panoramic video for sequences recorded by multiple traditional two-dimensional cameras [1], [3]. A spherical

panoramic frame is displayed in Fig. 1a, while its planar panoramic counterpart is shown in Fig. 1b. Let us now briefly introduce the equi-rectangular projection (ERP) [1], [2], namely the geographic projection.



(a) Spherical Coordinates



(b) Planar Coordinates

Figure 2: Sphere and the corresponding planar coordinate.

A. Transforming spherical coordinates into planar coordinates

We assume that point A of Fig. 2a is an arbitrary point on the sphere, where the $\alpha_o\,(-\pi \leq \alpha_o \leq \pi)$ and $\alpha_a\left(-\frac{\pi}{2} \leq \alpha_a \leq \frac{\pi}{2}\right)$ are the longitude and latitude of A with coordinates (x_a, y_a, z_a), respectively. The longitude α_o may be calculated as

$$\alpha_o = \begin{cases} -\frac{\pi}{2} + \arctan\frac{y_a}{x_a}, & 0 \leq x_a \leq R \\ \frac{\pi}{2} + \arctan\frac{y_a}{x_a}, & -R \leq x_a < 0 \end{cases}, \tag{1}$$

where R indicates the radius of the sphere, while the latitude α_a may be expressed as

$$\alpha_a = \arcsin\frac{z_a}{R}, -R \leq z_a \leq R. \tag{2}$$

With the derived longitude and latitude, we may readily express the projected location of point A on the planar coordinates as (w_a, h_a), where we have

$$\begin{aligned} w_a &= \alpha_o \cdot R \\ h_a &= \alpha_a \cdot R \end{aligned}. \tag{3}$$

Correspondingly, we can derive the width and height of the projected frame size as $W = 2\pi R$, $H = \pi R$.

B. Transforming planar coordinates into spherical coordinates

We assume that point A of Fig. 2b is an arbitrary point on the planar rectangle with the coordinates (w_a, h_a), where $-\frac{W}{2} \leq w_a \leq \frac{W}{2}$ and $-\frac{H}{2} \leq h_a \leq \frac{H}{2}$. The $\alpha_o\,(-\pi \leq \alpha_o \leq \pi)$ and $\alpha_a\left(-\frac{\pi}{2} \leq \alpha_a \leq \frac{\pi}{2}\right)$ are the projected longitude and latitude of A on the sphere. The longitude α_o may be calculated as

$$\alpha_o = \frac{w_a}{R}, \quad -\frac{W}{2} \leq w_a \leq \frac{W}{2}, \tag{4}$$

while the latitude α_a may be expressed as

$$\alpha_a = \frac{h_a}{R}, -\frac{H}{2} \leq h_a \leq \frac{H}{2}. \tag{5}$$

With the derived longitude and latitude, we may readily express the projected location of point A on the spherical coordinates as (x_a, y_a, z_a), where we have

$$\begin{aligned} x_a &= R \cdot \cos\alpha_a \cdot \cos\alpha_o \\ y_a &= R \cdot \cos\alpha_a \cdot \sin\alpha_o\,. \\ z_a &= \sin\alpha_a \cdot R \end{aligned} \tag{6}$$

III. System Architecture

In Section III-A we first briefly introduce the panoramic video streaming scenarios considered. Then we detail the proposed ROI-UEP scheme conceived for HEVC encoded panoramic video streaming over wireless channels, which is seen in Fig. 5.

A. Panoramic Video Streaming Scenarios

In the panoramic video streaming/broadcasting scenario considered in Fig. 3, the video server such as Youtube transmits the original panoramic video to multiple panoramic video playback terminals. The playback terminals can only display the ROI pixels falling for example within 120° of the viewing center, while the rest of the areas are ignored. Hence it is a natural desire to only transmit the compressed ROI signals instead of the full panoramic video for conserving the network's resources, such as its transmit power and/or bandwidth. More specifically, the terminals feed back their viewing trajectories based on their ROI, which will then be exploited by the video server for the sake of compressing the ROI signals. Hence, the bitrate of the video source may be reduced to 3 Megabits per second (Mbps) from 10 Mbps under perfect networking conditions. Alternatively, the remaining 7 Mbps may be used for the protection of the 3 Mbps source signals in a realistic error-prone network.

Again, the ROI is determined by the viewing center of the panoramic frame considered, since the viewer tends to focus his/her attention on the angular range of say 120° centered at the focal point of the viewer's eye. Here we assume that the estimated viewing center v of Fig. 5 is perfectly known at the transmitter, which may however be estimated using the popular deep learning tool of [37], [38]. The ROI and viewing center are exemplified in Fig. 4. In practice the viewing trajectory has to be fed back to the video server, as indicated in Fig. 3, but again, in this paper our focus is on the wireless streaming aspects.

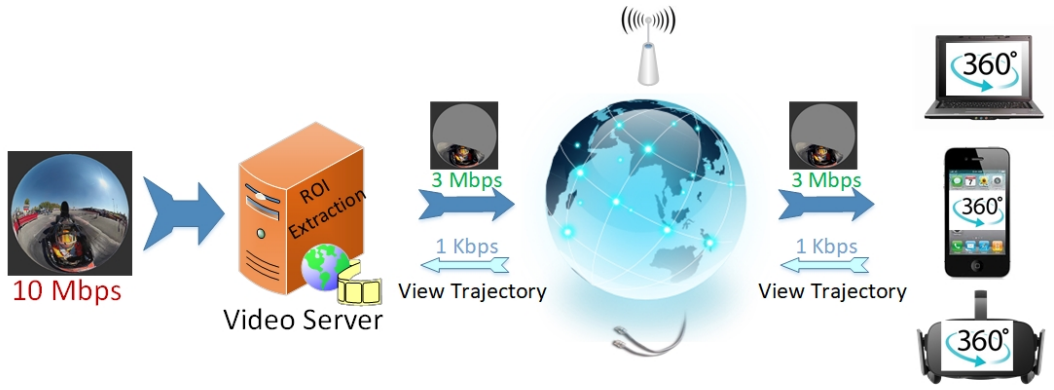


Figure 3: Panoramic video streaming scenarios considered.

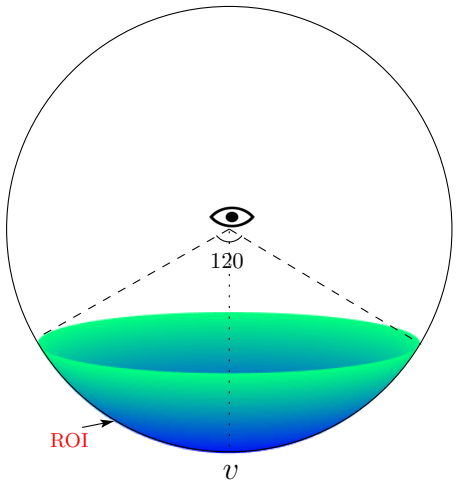


Figure 4: The 120° ROI of a frame is determined by the viewing center v.

Symbol	Definition
U	the panoramic frame considered
n	number of blocks created from the panoramic frame considered
$\mathcal{S}$	the estimated SNR for configuring the transmission of the panoramic video
u_i	the original YUV block i created from the panoramic frame considered
b_i	the bitstream representing the block u_i using the HEVC encoder
v	estimated viewing center of the frame considered
γ_i	FEC coding rate of block b_i
x_i	FEC encoded version of block b_i
y_i	the received version of sequence x_i
$\hat{b}_i$	the decoded version of block b_i
$\hat{u}_i$	the decoded bitstream of block u_i
$\hat{U}$	the reconstructed panoramic frame

Table I: Symbol definitions of system seen in Fig. 5, where $1 \leq i \leq n$ is the block index.

B. Proposed ROI-UEP Architecture

The architecture of the proposed ROI-UEP is shown in Fig. 5. This section focuses on the general architecture of the transmitter and receiver, while the "Minimize Distortion" block of Fig. 5 will be detailed in Section IV. Let us commence by defining the notation of Fig. 5 in Table I.

1) Transmitter Model: At the transmitter of Fig. 5, the panoramic frame U considered is split into blocks $u_1, \cdots, u_n$, which are expressed in YUV formats and compressed by the HEVC encoder, generating the bitstreams $b_1, \cdots, b_n$. Meanwhile, the information of blocks $b_1, \cdots, b_n$ is entered into the "Minimize Distortion" block of Fig. 5, which optimizes the coding rates $\gamma_1, \cdots, \gamma_n$ for the blocks $b_1, \cdots, b_n$, respectively. Then, the resultant n bitstreams $b_1, \cdots, b_n$ will be encoded as follows:

- The n bit sequences $b_1, \cdots, b_n$ are encoded by the FEC encoders $1, \cdots, n$ of Fig. 5, where the coding rates $\gamma_1, \cdots, \gamma_n$ generated by the "Minimize Distortion" block are employed, respectively. This results in the encoded bit sequences $x_1, \cdots, x_n$, respectively.
- The bit sequences $x_1, \cdots, x_n$ are then concatenated into a single bitstream for transmission.

The joint bit sequence is finally transmitted through the antenna.

2) Receiver Model: The received panoramic signals are processed by the wireless receiver of Fig. 5 as follows:

- Following demodulation, the soft information of the sequences $x_1, \cdots, x_n$, namely $y_1, \cdots, y_n$, will then be generated by the receiver of Fig. 5.
- The soft information y_i is then decoded by the FEC decoder i of Fig. 5, generating the estimated bit sequence $\hat{b}_i$, which represents the estimated version of layer b_i.

Finally, the estimated bitstreams $\hat{b}_1, \cdots, \hat{b}_n$ are then decoded by the HEVC decoder of Fig. 5, resulting in the YUV blocks

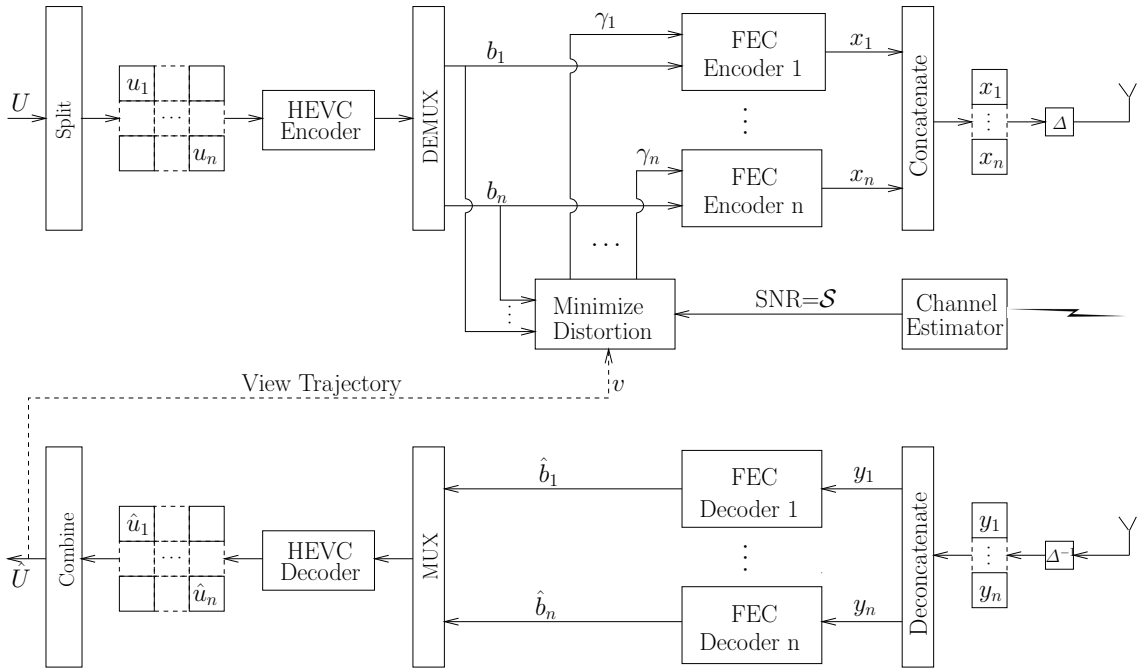


Figure 5: Architecture of the proposed ROI-UEP aided wireless panoramic communications system, where the symbols are defined in Table I. The "Minimize Distortion" block will be illustrated in Section IV.

$\hat{u}_1, \cdots, \hat{u}_n$, which are finally combined for reconstructing the panoramic video frame $\hat{U}$.

IV. Optimized ROI-UEP Coding Rates

Symbol	Definition
Γ	overall coding rate of panoramic frame U
$\|b_i\|$	number of bits in the bitstream i in the panoramic frame U
$\|x_i\|$	number of bits in the FEC encoded bitstream i
$d(b_i)$	video distortion caused by the corruption of block b_i
$p\left(s,l,\gamma\right)$	the expected packet error probability at the receiver, where l indicates the length of the input bit sequence to the FEC encoder with rate γ and transmitted at SNR s

Table II: Symbol definition, where $1 \leq i \leq n$ indicates the block index.

In this section, we detail the "Minimize Distortion" block of Fig. 5, which decides the coding rates $\gamma_1, \cdots, \gamma_n$ for encoding the different-significance blocks $b_1, \cdots, b_n$ of the panoramic frame U. The "Minimize Distortion" block is designed for the sake of minimizing the distortion of the reconstructed video expected at the receiver. More specifically, we consider the overall coding rate of Γ at the transmitting SNR of $\mathcal{S}$. Hence, the generated coding rates $\gamma_1, \cdots, \gamma_n$ satisfy the following condition

$$\frac{\sum_{i=1}^{n} |x_i|}{\Gamma} = \sum_{i=1}^{n} |x_i|. \tag{7}$$

Below, we minimize the video distortion of the blocks b_1,$\cdots$,b_n by deriving the specific FEC coding rates $\gamma_1, \cdots, \gamma_n$. Based on the symbol definitions of Table I, we characterize our algorithm by employing the notations in Table II.

We quantify the video distortion according to the peak signal-to-noise ratio (PSNR) degradation $\mathcal{D}\left(\gamma_1, \cdots, \gamma_n\right)$, caused by the n erroneous blocks considered, when the coding rates of $\gamma_1, \cdots, \gamma_n$ and SNR=$\mathcal{S}$ are employed. In our proposed scheme, our objective is to derive the coding rates $\gamma_1, \cdots, \gamma_n$, which minimize the expected degradation $\mathcal{D}\left(\gamma_1, \cdots, \gamma_n\right)$. For the derivation of the expected frame distortion, we employ the following assumptions for the block b_i

- $d\left(b_i\right)$: denotes the distortion caused by the corruption of block b_i;
- $p\left(s, |b_i|, \gamma_i\right)$: the packet error ratio (PER) of the block b_i using the coding rate γ_i and wireless transmit power s;
- $w\left(b_i\right)$: the weight of the block b_i.

Then, the distortion caused by the corruption of block b_i may be expressed as $d\left(b_i\right) \cdot p\left(\mathcal{S}, |b_i|, \gamma_i\right) \cdot w\left(b_i\right)$. The expected distortion $\mathcal{D}\left(\gamma_1, \cdots, \gamma_n\right)$ of the panoramic frame may be expressed as

$$\begin{aligned} &\mathcal{D}\left(\gamma_1, \cdots, \gamma_n\right) \\ &= \sum_{i=1}^{n} d\left(b_i\right) \cdot p\left(S, |b_i|, \gamma_i\right) \cdot w\left(b_i\right). \end{aligned} \tag{8}$$

Figure 6: The FEC decoding process at the receiver.

Hence, our objective function may be formulated as

$$\begin{aligned} &\arg \min_{\gamma_1,\cdots,\gamma_n} \mathcal{D}(\gamma_1,\cdots,\gamma_n) = \\ &\arg \min_{\gamma_1,\cdots,\gamma_n} \sum_{i=1}^{n} d(b_i) \cdot p(S,|b_i|,\gamma_i) \cdot w(b_i), \end{aligned} \tag{9}$$

subject to the condition of

$$\frac{\sum_{i=1}^{n} |b_i|}{R} = \sum_{i=1}^{n} \frac{|b_i|}{\gamma_i}, \tag{10}$$

where Eq. (10) indicates the overall bitrate limit of transmitting the n encoded blocks.

In Sections IV-A and IV-B, we resolve the components of Eq. (9), namely the panoramic video distortion $d(b_i)$ and the PER $p(S,|b_i|,\gamma_i)$, respectively. Afterwards, the solution of the OF in Eq. (9) is detailed for the sake of determining the coding rates in Section IV-D. Finally, the transmission overhead imposed by our proposed scheme is discussed in Section IV-E.

A. Estimation of the Block Distortion $d(\cdot)$

The distortion $d(b_i)$ caused by the corruption of the block b_i is estimated using a similar solution to that of [14], [18], [39]. Explicitly, in this paper, the distortion $d(b_i)$, $1 \leq i \leq n$ is determined by decoding the bitstream in the total absence of block b_i [18]. Alternatively, the solutions of [26], [40], [41] may be employed in our system.

B. Estimation of the PER $p(\cdot)$

The FEC decoder of Fig. 5 is shown in Fig. 6, where the soft information of length $|y| = \frac{\lambda}{\gamma}$ is input into the FEC decoder, which outputs the estimated bit sequence $\hat{b}$ of length λ. Moreover, r indicates the coding rate of the FEC codec and the signals are received at SNR s. Based on the constant value l, the PER of $\hat{b}$ in Fig. 6 depends on the parameters s and γ, which may be expressed as $p(s,l,\gamma)$.

It has been shown in [18], [42] that the occurrence of burst errors encountered by non-iterative codecs remains unaffected by the packet length. For illustrating the burst error distributions, we consider the scenario that a packet carrying $(n_1 \times n_2)$ bits is generated by a non-iterative FEC decoder. Furthermore, this $(n_1 \times n_2)$-bit segment may be considered as either n_1 packets, associated with n_2 bits each or n_2 packets each carrying n_1 bits [43]. As detailed in [43], we have

$$p(n_1) = 1 - [1 - p(n_2)]^{\frac{n_1}{n_2}}. \tag{11}$$

By substituting n_1, n_2 of Eq. (11) with $|b_i|$, l, the PER $p(s,|b_i|,\gamma_i)$ component of the OF in Eq. (9) may be formulated as

$$p(s,|b_i|,\gamma_i) = 1 - [1 - p(s,l,\gamma_i)]^{\frac{|b_i|}{\lambda}}, \tag{12}$$

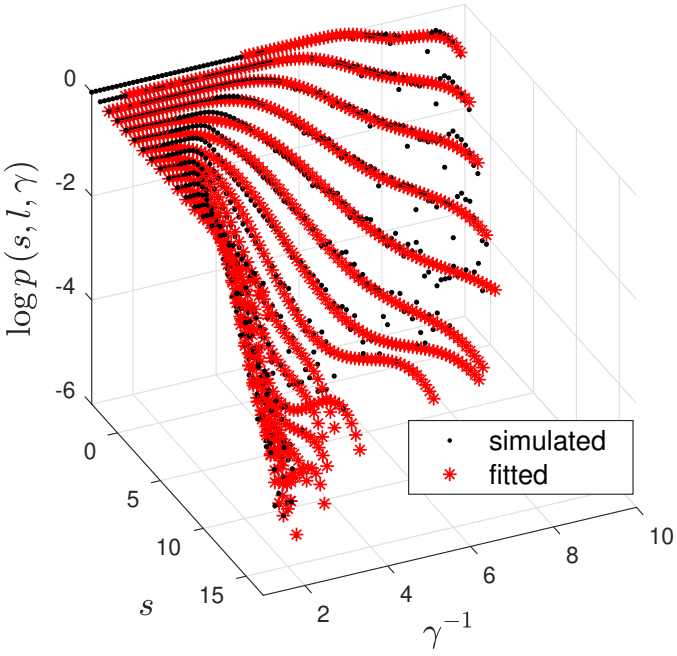


Figure 7: Simulated surface versus the mathematical fitted surface of $p(s,\lambda,\gamma)$ using the model of $-1.1^{\left(d\cdot\gamma^{-3}+c\cdot\gamma^{-2}+b\cdot\gamma+a\right)}$, where we employ $\lambda = 1000$.

where l is the number of bits in the packet input to the FEC decoder of Fig. 6. For further information concerning this calculation, we may refer to [14], [18].

For the sake of determining the function $p(s,l,\gamma)$ of Eq. (12), we firstly simulate the decoding process of Fig. 6 employing the variables s, γ along with the constant value $l = \lambda = 1000$. This generates the look-up table (LUT) $\hbar(s,\gamma)$ as indicated by the "simulated" surface of Fig. 7. Then, as indicated by the "fitted" surface of Fig. 7, the LUT $\hbar(s,\gamma)$ is mathematically modeled as

$$\log \hbar(s,\gamma) = -1.1^{\left(d\cdot\gamma^{-3}+c\cdot\gamma^{-2}+b\cdot\gamma+a\right)}, \tag{13}$$

where a,b,c and d may be readily found, when s is known. Note that some "fitted" curves of Fig. 7 wiggle up and down in the lower PER range, such as $p(s,l,\gamma) = 10^{-5}$. This is because the "simulated" surface is generated employing 10^5 simulated packets, which leads to inaccurate "fitted" curves. Hence, longer simulations are required for generating smoother "fitted" surface. Based on the definitions of $\hbar(s,\gamma)$ and $p(s,l,\gamma)$, the probability $p(s,l,\gamma)$ may be expressed as

$$p(s,l,\gamma) = \hbar(s,\gamma). \tag{14}$$

Then the PER estimation $p(s,|b_i|,\gamma_i)$ in Eq. (12) may be formulated using the function $\hbar(s,\gamma)$ as

$$p(s,|b_i|,\gamma_i) = 1 - [1 - \hbar(s,\gamma_i)]^{\frac{|b_i|}{\lambda}}. \tag{15}$$

By substituting Eq. (15) into Eq. (9), the expected video

distortion $\mathcal{D}(\gamma_1, \cdots, \gamma_n)$ may be finally formulated as

$$\begin{aligned}&\mathcal{D}(\gamma_1, \cdots, \gamma_n)\\&= \sum_{i=1}^{n} d(b_i) \cdot \left[1 - [1 - \hbar(s_i, \gamma_i)]^{\frac{|b_i|}{\lambda}}\right] \cdot w(b_i).\end{aligned} \tag{16}$$

C. Estimation of the Weight $w(\cdot)$

The panoramic frame U is converted from the sphere E, where the pixels are displayed on the panoramic displays, such as Oculus rift [44]. If we ignore the unequal importance of visual information associated with different viewing directions as well as with different visual contents, the pixels on the sphere E have equal importance, while the pixels on U have different importance. However, in practical panoramic applications, such as Youtube, panoramic videos are delivered in the planar format, but they are viewed in a spherical format. Hence, we have to evaluate the weight, namely the importance, of spherical panoramic blocks, which are generated from planar panoramic blocks u_i, $1 \leq i \leq n$. It may be a challenge to derive their weight analytically, since they are deformed rectangles. Our methodology is to generate M uniformly distributed points on the sphere E, where each pixel is denoted as $e()$. By counting the number of these points positioned in a planar block u_i, we are able to estimate the importance of block u_i. A number of points may be uniformly distributed on sphere E as follows:

- In the latitude direction of the sphere, we employ $\frac{\pi}{\delta}$ equally spaced parallel circles with an angle of δ, which are exemplified in Fig. 8a.
- We denote the angle $\angle COB$ of Fig. 8 as α. For each circle, the radius may be calculated as $r = R \cdot \sin\alpha$. On each circle, we employ all equally spaced points with an angle of δ, which are exemplified in Fig. 8c. Note that $|AB|$ belongs to both the triangle ACB and triangle AOB. The length $|AB|$ may be calculated as $2r \cdot \sin\frac{\angle ACB}{2}$ or $2R \cdot \sin\frac{\delta}{2}$. Hence, we have $\angle ACB = 2\arcsin\frac{\sin\frac{\delta}{2}}{\sin\alpha}$. Then, we have $\frac{2\pi}{\angle ACB}$ points on the circle considered, each of which may be calculated as

$$\begin{cases} x = r \cdot \sin\left(2i \cdot \arcsin\dfrac{\sin\frac{\delta}{2}}{\sin\alpha}\right) \\ y = r \cdot \cos\left(2i \cdot \arcsin\dfrac{\sin\frac{\delta}{2}}{\sin\alpha}\right) \\ z = R \cdot \cos(\pi - \alpha). \end{cases} \tag{17}$$

We generate equally distributed points as exemplified in Fig. 9 with the aid of the above mentioned process.

We denote the set of generated points on sphere E as Φ_e and the number of points as $|\Phi_e|$. An arbitrary point $(x, y, z) \in \Phi_e$ satisfies $x^2 + y^2 + z^2 = R^2$, which may be mapped to a rectangle point according to the equi-rectangular projection of [1], as discussed in Section II-A.

The related points on the rectangle of Fig. 9 are shown in Fig. 10.

We denote the points on the rectangle as the set Φ_γ and the number of points as $|\Phi_\gamma|$. Since the points on the sphere E are uniformly distributed, the number of points located within the subblock u_i indicates the importance of u_i, namely its weight $w(\cdot)$. Hence, the weight $w(i)$ of block u_i of Fig. 5 may be calculated as

$$w(i) = \frac{|\phi|}{|\Phi_\gamma|}, \phi = \{(w,h) \mid (w,h) \in u_i, \forall (w,h) \in \Phi_\gamma\}. \tag{18}$$

In the definition of Eq. (18), $w(i)$ represents the percentage of points located within the subblock u_i, while (w, h) represents the planar coordinates of a rectangular pixel. Correspondingly, we have $\sum_{i=1}^{n} w(i) = 1$ based on Eq. (18).

In our proposed UEP-ROI solution, the invisible pixels outside the ROI should not be transmitted, since they are ignored by the viewers. Hence, they are counted when calculating the weight of block u_i, leading to the following equation

$$\begin{aligned}w(i) =& \frac{|\phi|}{|\Phi_\gamma|}\\ &\phi = \{(w,h) \mid (w,h) \in u_i \wedge (w,h) \in ROI, \forall (w,h) \in \Phi_\gamma\}.\end{aligned} \tag{19}$$

With this definition of $w(\cdot)$, the blocks outside ROI have weights of 0, indicating that they will not be transmitted. The weights of the blocks within ROI will be decided depending on how many spherical ROI pixels they carry.

D. ROI-UEP Coding Rates

Given the OF defined in Eq. (9) and the coding rate constraint of Eq. (9), we have the Lagrange multiplier equation as in Eq. (20),

where λ is the Lagrange multiplier. The corresponding set of partial derivatives $\nabla_{\gamma_1,\cdots,\gamma_n,\lambda}\mathcal{L}(\gamma_1, \cdots, \gamma_n, \lambda) = 0$ may be readily derived as Eqs. (21).

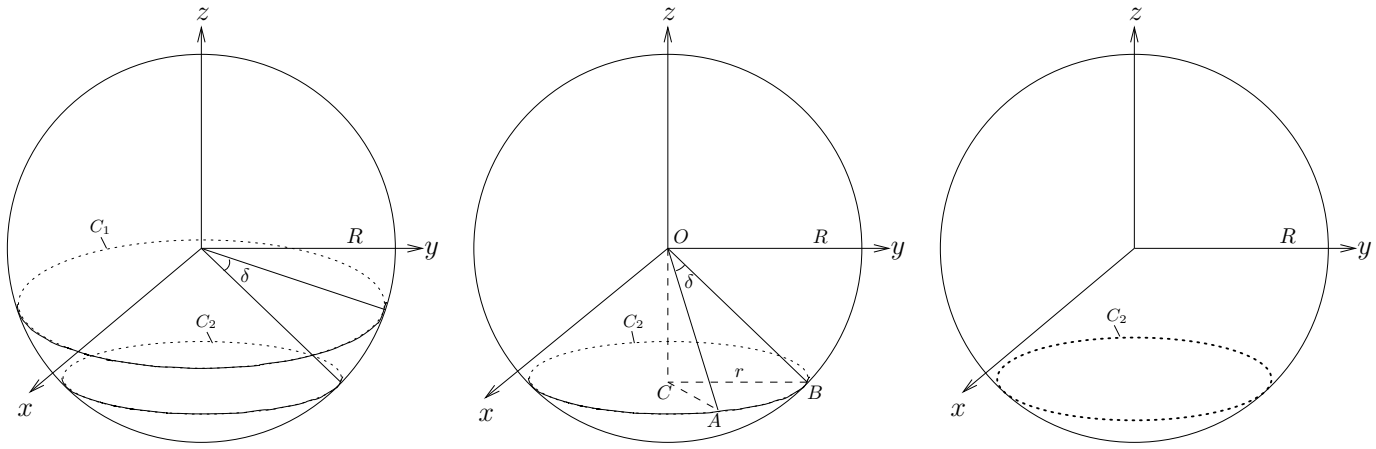
The set of Eqs. (21) was then solved using the classic Newton down-hill method as detailed in Algorithm 1.

Algorithm 1 Newton down-hill method for solving Eq. (21).

inputs:
  $\mathcal{S}, |b_1|, \cdots, |b_n|, w(b_1), \cdots, w(b_n), d(b_1), \cdots, d(b_n)$
initialize:
  $dist \leftarrow +\infty$
for each initial point $\{\hat{\gamma}_1, \cdots, \hat{\gamma}_n\}$ **do**
  ▷ determining the code rates using Newton down-hill
  $\hat{\gamma}_1, \cdots, \hat{\gamma}_n \leftarrow \nabla_{\hat{\gamma}_1,\cdots,\hat{\gamma}_n,\lambda}\mathcal{L}(\hat{\gamma}_1, \cdots, \hat{\gamma}_n, \lambda)$
  $tmp \leftarrow \mathcal{D}(\mathcal{S}, b_1, \cdots, b_n, \hat{\gamma}_1, \cdots, \hat{\gamma}_n)$
  if $tmp < dist$ **then**
    $dist \leftarrow tmp$
    $\gamma_1, \cdots, \gamma_n \leftarrow \hat{\gamma}_1, \cdots, \hat{\gamma}_n$
  end if
end for
outputs:
  $\gamma_1, \cdots, \gamma_n$

E. Overhead

All the above-mentioned operations are performed at the transmitter of Fig. 5. The overhead imposed by this optimiza-



(a) Parallel circles c_1,c_2 (b) Generation of points on c_2 (c) Example of equally spaced points on c_2

Figure 8: Generation of uniformly distributed points on sphere E.

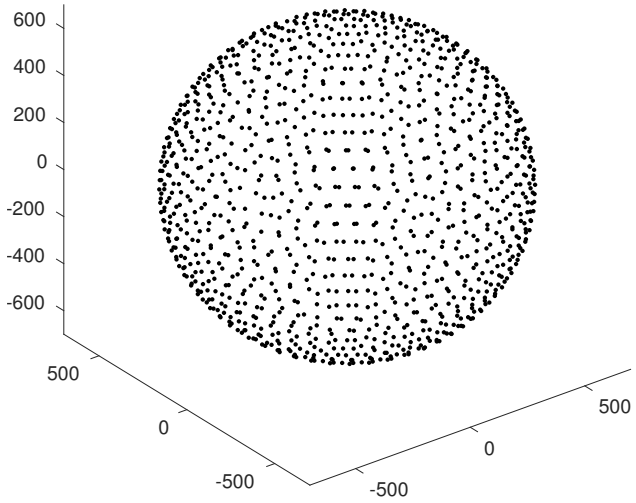


Figure 9: Uniformly distributed points on sphere E.

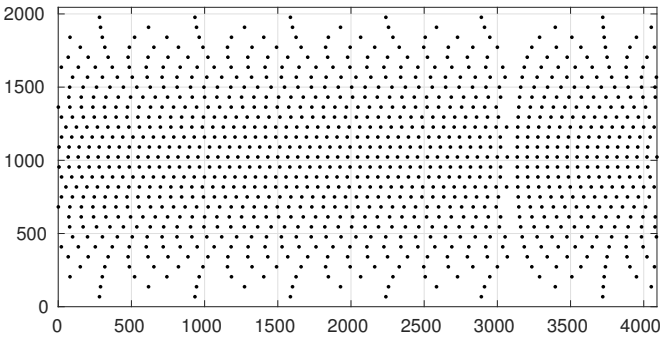


Figure 10: Uniformly distributed points on sphere E mapped to rectangle.

tion process will be briefly illustrated below, noting that no overhead is imposed at the receiver side. More specifically, the overhead imposed includes the estimation of the distortion $d(\cdot)$, estimation of the PER $p(\cdot)$, estimation of the weight $w(\cdot)$ and determination of the coding rates.

1) Estimation of the distortion $d(\cdot)$: As discussed in [14], [18], [39], the complexity imposed by estimating the distortion $d(b_i)$ is linearly proportional to n.

2) Estimation of the PER $p(\cdot)$: As detailed in Section IV-B, the PER estimation mainly includes the LUT $\hbar(s,\gamma)$ generation process, which is specific for the FEC decoder of Fig. 6. Again, the LUT $\hbar(s,\gamma)$ is obtained by simulating the decoding process of Fig. 6, which is carried out during the offline design process. *Moreover, the LUT is independent of the video sequences employed and dependent on the channel, on the modulator and on the FEC employed.* The size of the LUT may be expressed as $(n_s \times n_\gamma)$, where n_s and n_γ indicate the number of variables s and γ, respectively.

3) Estimation of the weight $w(\cdot)$: As detailed in Section IV-C, the weight estimation mainly includes generation of the M uniform points, which is carried out during the offline design process.

4) ROI Estimation: Note that each of the 60 viewing trajectory points defines the corresponding estimated viewing center [37], [38], which will be utilized for determining the ROI of the panoramic frame considered, as exemplified in Fig. 4. This process would only impose a modest complexity. Here we refrain from detailing this process, since it is beyond the scope of this paper.

5) Determination of the coding rates: In our proposed system, the coding rates are determined using the classic Newton down-hill method, which imposes a modest complexity. Alternatively, the adaptive particle swarm optimization (APSO) technique of [45] may be employed for finding the coding rates.

V. System Performance

Below, we benchmark our proposed ROI-UEP system against the equal error protection (EEP) and the ROI based

$$\begin{aligned}\mathcal{L}\left(\gamma_1,\cdots,\gamma_n,\lambda\right) &= \mathcal{D}\left(\mathcal{S},b_1,\cdots,b_n,\gamma_1,\cdots,\gamma_n\right)+\lambda\cdot\left(\sum_{i=1}^{n}\frac{|b_i|}{\gamma_i}-\frac{\sum_{i=1}^{n}|b_i|}{\Gamma}\right)\\ &= \sum_{i=1}^{n} d\left(b_i\right)\cdot p\left(S,|b_i|,\gamma_i\right)\cdot w\left(b_i\right)+\lambda\cdot\left(\sum_{i=1}^{n}\frac{|b_i|}{\gamma_i}-\frac{\sum_{i=1}^{n}|b_i|}{\Gamma}\right)\end{aligned} \tag{20}$$

$$\nabla_{\gamma_1,\cdots,\gamma_n,\lambda}\mathcal{L}\left(\gamma_1,\cdots,\gamma_n,\lambda\right)=\begin{cases}\frac{\partial\mathcal{L}}{\partial\gamma_1} & = d\left(b_1\right)\cdot w\left(b_1\right)\cdot\frac{\partial p(S,|b_1|,\gamma_1)}{\partial\gamma_1}-\lambda\cdot\frac{|b_1|}{\gamma_1^2}=0\\ & \vdots\\ \frac{\partial\mathcal{L}}{\partial\gamma_n} & = d\left(b_n\right)\cdot w\left(b_n\right)\cdot\frac{\partial p(S,|b_n|,\gamma_n)}{\partial\gamma_n}-\lambda\cdot\frac{|b_n|}{\gamma_n^2}=0\\ \frac{\partial\mathcal{L}}{\partial\lambda} & = \sum_{i=1}^{n}\frac{|b_i|}{\gamma_i}-\frac{\sum_{i=1}^{n}|b_i|}{\Gamma}=0\end{cases} \tag{21}$$

Parameters	RaceVR, RollerCoaster
Representation	YUV 4:2:0
Format	4096×2048
Bits Per Pixel	8
FPS	30
No. of Frames	60
No. of Blocks/Frame	64
Video Codec	HEVC
GOP	1
Bitrate (Mbps)	69.2, 89.2
QP	24
Error-Free Y-PSNR (dB)	44.3, 46.2
FEC	RSC
QAM	BPSK
Channel	Rayleigh Fading
Simulations Repeated	100
Error Concealment	Frame-copy

Table III: Parameters used for transmitting the RaceVR and RollerCoaster sequences

EEP (ROI-EEP) systems. The parameters of the RaceVR [46] and RollerCoaster [47] sequences employed in the simulations are detailed in Table III. Specifically, the 4:2:0 YUV format (4096×2048)-pixel resolution based RaceVR and RollerCoaster video clips were encoded by the HEVC reference software, where the low-complexity "frame-copy" based error concealment was employed for replacing the corrupted panoramic frames. For the sake of simplicity, only IDR/CDR frames were employed in our simulations. However our algorithm may be readily extended both to B and P frames. Additionally, the video sequences were encoded by the H.265/ HEVC scheme using the standard quantization parameter (QPs) set of 24. These configurations jointly resulted in a bitrate of 69.2 Mbps and 89.2 Mbps at 30 frames per second (FPS). Furthermore,

Symbol	Definition
$U\left(i,j\right)$	original pixel at location (i,j) of frame U
$\hat{U}\left(i,j\right)$	estimated pixel at location (i,j) of frame U
$I\left(i,j\right)$	weight/importance of pixel $U\left(i,j\right)$
(u_x,u_y,u_z)	the transformed pixel on sphere from pixel $U\left(i,j\right)$
(c_x,c_y,c_z)	point of view on the spherical frame
v	angle of the view

Table IV: Symbol definition employed for WPSNR estimation.

in the absence of transmission errors, Y-PSNR of 42.5 dB and 46.2 dB may be achieved by reconstructing the RaceVR and RollerCoaster sequences.

Apart from the panoramic video parameters of Table III, the transmission and FEC parameters are detailed as follows. Specifically, a RSC codec [23] configured by the octal generator polynomials of $[031,027,027,027,027,035,035,035,035,033]$ was employed as the FEC codec resulting in a minimum coding rate of 0.1. Moreover, binary phase shift keying (BPSK) was employed for transmitting the FEC encoded bitstream. Moreover, each HEVC coded bitstream was RSC encoded, BPSK modulated and then transmitted on a network abstract layer unit (NALU) by NALU [48] basis. The simulations were repeated 100 times for the sake of generating statistically sound performance curves. The remaining parameters are listed in Table III.

A. PSNR Evaluation

Bearing in mind the discussions in Section II, we know that different pixels on the planar frame have different importance. Correspondingly, we weight the pixels for evaluating the PSNR of the frame U. We employ the symbol definitions listed in Table IV. Specifically, we denote the weight, namely importance, of the pixel $U\left(i,j\right)$ as $I\left(i,j\right)$. Furthermore, we

$$I(i,j) = \begin{cases} 0 & , d < \sqrt{(u_x - c_x)^2 + (u_y - c_y)^2 + (u_z - c_z)^2} \\ \frac{2\pi \cdot \sqrt{u_z^2 + u_y^2}}{W} & , d \geq \sqrt{(u_x - c_x)^2 + (u_y - c_y)^2 + (u_z - c_z)^2} \end{cases} \quad (22)$$

assume that the point of view on the sphere is (c_x, c_y, c_z) with the viewing angle of υ. Hence the longest distance of pixels on the sphere located within the viewing range may be expressed as

$$d = R \cdot \sqrt{2 \cdot (1 - \cos \upsilon)} \ . \quad (23)$$

Recall from Section II that the pixels in the same row of Fig. 2b are on the same latitude of Fig. 2a and vice versa. Hence, the importance of pixel $U(i,j)$ may be calculated as the ratio between the related latitude circumference and the width W, namely $I(i,j)$ in Eq. (22).

The weighted mean square error (WMSE) employed is defined as

$$WMSE = \frac{\sum_{i=0}^{W} \sum_{j=0}^{H} \left[\hat{U}(i,j) - U(i,j)\right]^2 \cdot I(i,j)}{\sum_{i=0}^{W} \sum_{j=0}^{H} I(i,j)} \ , \quad (24)$$

while the corresponding weighted peak signal-to-noise ratio (WPSNR) may be calculated as

$$WPSNR = 10 \cdot \log_{10} \frac{MAX^2}{WMSE} \ , \quad (25)$$

where we have $MAX = 2^m$ for m-bit pixel scenarios.

B. Off-line LUT Generation

For generating the LUT, the vectors of $[-5:1:25]$, $[0:0.1:9.9]$ are utilized for the variables s, γ^{-1} of $\hbar(s,\gamma)$, respectively, resulting in $n_s = 31$, $n_\gamma = 100$. Moreover, a packet-length of $\lambda = 1000$ is employed.

C. Benchmarkers

1) ROI-UEP: The architecture of the ROI-UEP scheme is detailed in Fig. 5.

2) ROI-EEP: In the ROI-UEP scheme the signals within the ROI are unequally protected, while in the ROI-EEP arrangement they have the same protection associated with $\gamma_1 = \cdots = \gamma_n$.

3) EEP: The entire panoramic video signal - including the ROI - is equally protected using the coding rate Γ.

D. Performance

Below, we evaluate the WPSNR video quality, the PER and the BER of the blocks involved.

1) View Trajectory: In the simulations, we employ a view trajectory, namely head-moving track for all 100 rounds of simulations for the sake of observing the performance improvements using our proposed algorithm. The trajectory employed is shown in Fig. 11. Note that the each of the 60 trajectory points represents an estimated viewing center, which may be estimated using the deep learning tools [37], [38].

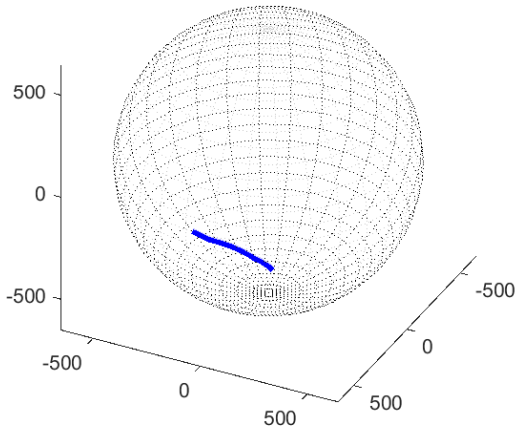


Figure 11: The view trajectory employed in the simulation, which consists of 60 view centers for 60 panoramic frames, respectively.

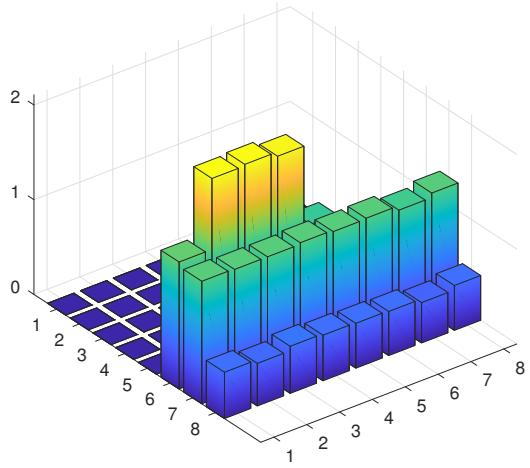
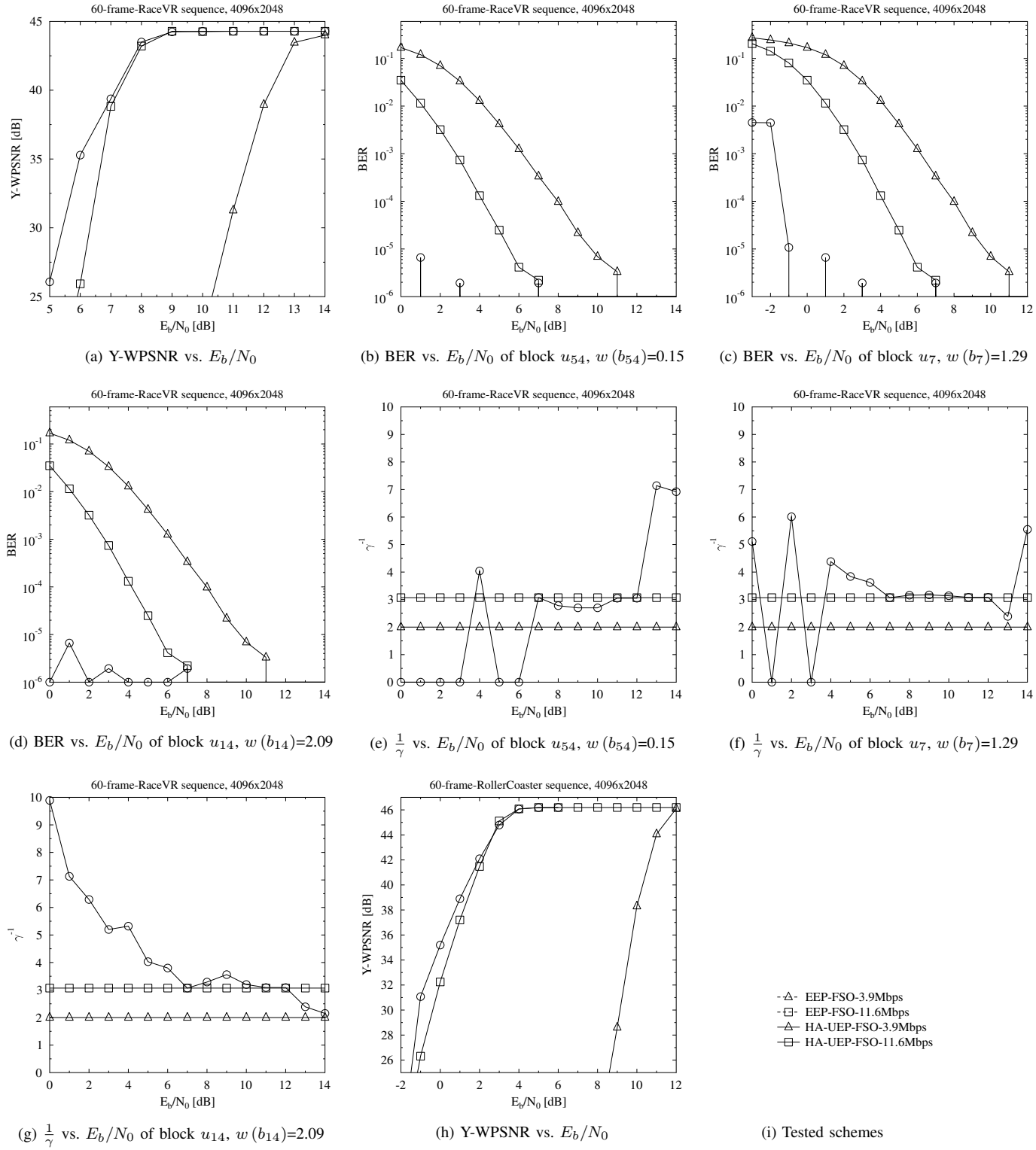


Figure 12: The weights of blocks are computed by the algorithm discussed in Section IV-C, where the panoramic frame is divided into 64 subblocks.

2) Weights of Blocks: The block-weights of the first panoramic frame are shown in Fig. 12, where the panoramic frame is divided into (8×8) subblocks resulting in $n = 64$ blocks.

3) WPSNR Video Quality: The WPSNR versus E_b/N_0 results of RaceVR sequence were recorded in Fig. 13a [1], where the ROI-EEP scheme is seen to substantially outperform the EEP scheme. Specifically, the ROI-EEP and EEP schemes achieve a WPSNR of 38.7 dB at 7 dB and 12 dB channel E_b/N_0, respectively. Alternatively, the ROI-EEP scheme achieves an E_b/N_0 gain of 5 dB, when aiming for a WPSNR

[1] As in traditional PSNR evaluation, Y-WPSNR is used representing WPSNR quality of the YUV sequence.



(a) Y-WPSNR vs. E_b/N_0

(b) BER vs. E_b/N_0 of block u_{54}, $w\,(b_{54})$=0.15

(c) BER vs. E_b/N_0 of block u_7, $w\,(b_7)$=1.29

(d) BER vs. E_b/N_0 of block u_{14}, $w\,(b_{14})$=2.09

(e) $\frac{1}{\gamma}$ vs. E_b/N_0 of block u_{54}, $w\,(b_{54})$=0.15

(f) $\frac{1}{\gamma}$ vs. E_b/N_0 of block u_7, $w\,(b_7)$=1.29

(g) $\frac{1}{\gamma}$ vs. E_b/N_0 of block u_{14}, $w\,(b_{14})$=2.09

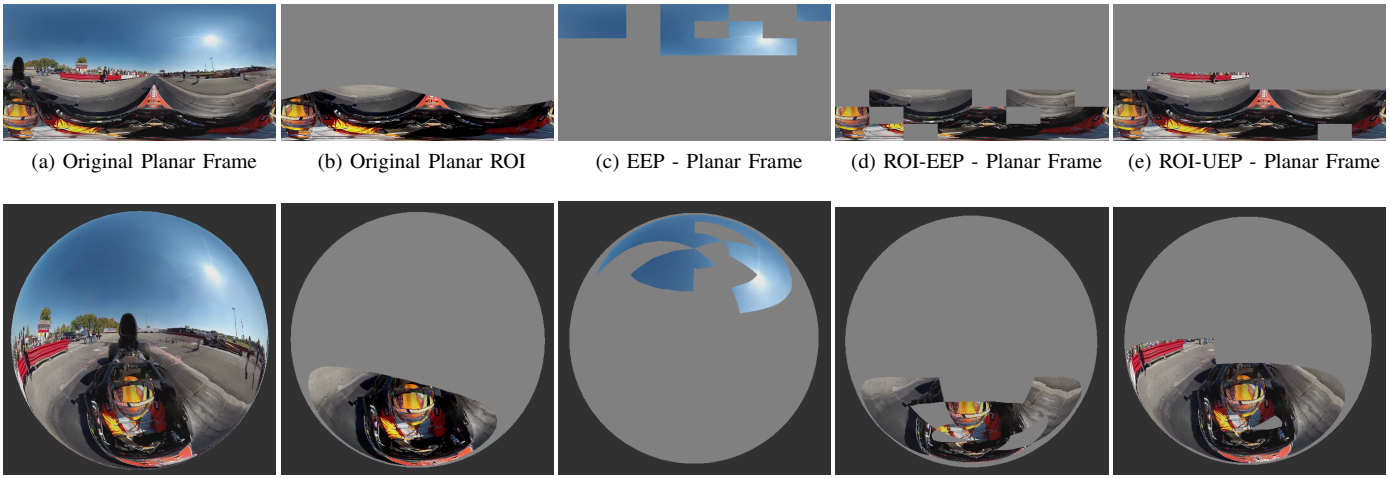
(h) Y-WPSNR vs. E_b/N_0

(i) Tested schemes

Figure 13: Y-WPSNR, BER versus E_b/N_0 comparison of the proposed ROI-UEP scheme, the ROI-EEP scheme, the EEP scheme for the *RaceVR and RollerCoaster* sequences.

of 38.7 dB. This is because the ROI-EEP scheme assigns stronger protection to the ROI signals by sacrificing the regions in the peripheral view of the viewers.

Observe from Fig. 13a of RaceVR sequence that the WP-SNR performance is further improved by the ROI-UEP scheme compared to the ROI-EEP scheme, especially in the lower E_b/N_0 range. Specifically, the ROI-UEP scheme outperforms the ROI-EEP scheme by 0.8 dB channel E_b/N_0 at a WPSNR of 35 dB. Moreover, at a channel E_b/N_0 of 6 dB, the ROI-UEP scheme outperforms the ROI-EEP scheme by a WPSNR



(a) Original Planar Frame (b) Original Planar ROI (c) EEP - Planar Frame (d) ROI-EEP - Planar Frame (e) ROI-UEP - Planar Frame

(f) Original Spherical Frame (g) Original Spherical ROI (h) EEP - Spherical Frame (i) ROI-EEP - Spherical Frame (j) ROI-UEP - Spherical Frame

Figure 14: Reconstructed frame comparison of the proposed ROI-UEP scheme, the ROI-EEP scheme, the EEP scheme for the *RaceVR* sequence. The first and second row indicates the planar and spherical frames, respectively. E_b/N_0 =6 dB is employed.

of 9.4 dB, where the ROI-UEP scheme and the ROI-EEP scheme are capable of achieving WPSNR of 35.3 dB and 25.9 dB, respectively. This is because the ROI-UEP scheme dynamically finds a suitable set of FEC coding rates for the non-uniform protection of the ROI signals, where the less important video signals may be sacrificed for improving the protection of the more important panoramic pixels.

Observe from Fig. 13h of RollerCoaster sequence that the trends WPSNR curves are similar to that of Fig. 13a. Note that the ROI-UEP and ROI-EEP curves of Fig. 13h achieve higher WPSNR in comparison to that of Fig. 13a at the same channel E_b/N_0 values. This is because the signals outside ROI of the RollerCoaster sequence carry more bits, which are sacrificed for the ROI signals. Alternatively saying, the ROI of the RollerCoaster sequence are protected with lower FEC coding rates in comparison to that of the RaceVR sequence.

4) BER of Blocks: The BER versus E_b/N_0 results recorded for blocks u_{54}, u_7 and u_{14} are displayed in Fig. 13b, Fig. 13c and Fig. 13d, respectively, where the corresponding weights of blocks are $w\left(b_{54}\right)$=0.15, $w\left(b_7\right)$=1.29 and $w\left(b_{14}\right)$=2.09. As for the ROI-EEP scheme, similar BER versus E_b/N_0 curves are observed for all the blocks u_{54}, u_7 and u_{14}. This is due to the fact that the ROI-EEP scheme allocates equal protection to the ROI including the region of blocks u_{54}, u_7 and u_{14}. Similarly, similar BER versus E_b/N_0 curves are observed for the EEP scheme, since all panoramic bitstream is equally protected including the region of blocks u_{54}, u_7 and u_{14}. Note that the ROI-EEP scheme always outperforms the EEP scheme. This is because the invisible region is sacrificed in the ROI-EEP scheme, while the spare protection is allocated to the ROI.

Moreover, from the results of the ROI-UEP scheme, lower BER values are observed in Fig. 13b, Fig. 13c and Fig. 13d for subblocks of higher weight. Specifically, the BER of block u_{14} is below 10^{-5} for E_b/N_0 values ranging from 0 dB to 14 dB, while the BER of block u_{54} drops to 10^{-5} at 7 dB. This is because the ROI-UEP scheme allocates more FEC protection redundancy to the more important ROI. Note that, in Fig. 13b, Fig. 13c and Fig. 13d, the BER curves recorded for the ROI-UEP scheme fluctuate with the E_b/N_0 values. This may be attributed to the fact that the optimization procedure of the ROI-UEP scheme seen in Algorithm 1 may fail to find the global optimum.

5) Coding Rates: The coding rate versus E_b/N_0 results recorded for blocks u_{54}, u_7 and u_{14} are displayed in Fig. 13e, Fig. 13f and Fig. 13g, respectively. For the ROI-EEP scheme, the same coding rate versus E_b/N_0 curves are observed for all the blocks u_{54}, u_7 and u_{14}, where the coding rates remain unchanged in all E_b/N_0 ranges. This is due to the fact that the ROI-EEP scheme allocates equal protection to the ROI including the region of blocks u_{54}, u_7 and u_{14}. Similar trends are observed for the EEP scheme. Note that the ROI-EEP scheme always has lower coding rates compared to the EEP scheme, since the invisible region is sacrificed in the ROI-EEP scheme.

As for the ROI-UEP scheme, in E_b/N_0 range of $[0, 7]$ dB, we observe that more important blocks are associated with lower coding rates. For example, the most important block b_{14} has lowest coding rates compared to the less important blocks b_7 and b_{54}. Moreover, in E_b/N_0 range of $[0, 7]$ dB, the ROI-UEP scheme even uses higher coding rates for the least important block b_{54} than that of the ROI-EEP scheme and the EEP scheme. This is because the block b_{54} is sacrificed to the block b_{14} for decreasing the panoramic video distortion, as shown in Fig. 13a. Furthermore, the coding rates of the block b_{14} increases with the increasing E_b/N_0, since less protection is required for securing the reconstruction of block b_{14} when more transmit power is consumed. We also note that fluctuations are observed in the coding rate versus E_b/N_0 curves of the the ROI-EEP scheme, as shown in Fig. 13e, Fig. 13f and Fig. 13g. This may be attributed to the fact that the optimization procedure of the ROI-UEP scheme seen in Algorithm 1 may fail to find the global optimum. Hence, we may consider to improve this optimization algorithm in our future work.

6) Panoramic Frames: Our reconstructed frame comparison recorded at a channel SNR of 6 dB is seen for the proposed system and for the benchmarkers in Fig. 14, where the *RaceVR* video sequence is employed. The ROI of the viewers is displayed in Fig. 14b and Fig. 14g, which are more strongly protected. Moreover, the ROI-UEP scheme reconstructed better visual quality in the ROI. Note that there are blank areas in the rightmost three columns in Fig. 14, which represent the reconstructed version of the panoramic frame for the EEP, ROI-EEP and ROI-UEP schemes, respectively. For the ROI-EEP and ROI-UEP schemes, this is because their preceding blocks in the previous panoramic frame are either not transmitted or corrupted, hence the frame-copy tool is unable to conceal the errors. For the EEP scheme, their preceding blocks in the previous panoramic frame are corrupted. We also note that only some of the blue sky region seen at the top of the frame is correctly reconstructed as shown in Fig. 14c and Fig. 14h. This is because this smooth blue sky area requires lower video bitrate in comparison to other areas, such as the area containing the racing car. As illustrated by Eq. (15), a smaller block length leads to a reduced PER. Hence, these blue sky blocks tend to be correctly reconstructed.

VI. Conclusions

Motivated by the fact that viewers tend to focus on certain parts of the panoramic frame, we proposed the ROI-UEP and ROI-EEP concept for video communications over wireless channels. Moreover, the ROI pixels may have unequal importance, hence UEP was employed for reducing the video distortion. Specifically, each panoramic frame was divided into multiple blocks of unequal importance, which was calculated by accumulating each pixel's weight. Our OF was derived considering these blocks' importance, their impact on the video distortion and their PER. This OF aims for maximizing the so-called WPSNR of the viewers' ROI signals. By solving this OF, a specific set of blocks with different FEC coding rates may be selected and transmitted for minimizing the expected video distortion. The simulations demonstrated that our proposed ROI-UEP system outperforms the ROI-EEP benchmarking scheme by a WPSNR margin of 9.4 dB at a channel SNR of 6 dB.

In our future work, we may consider our previous inter-layer FEC technique of [17], [18] for ROI based wireless panoramic video communications. Moreover, we may consider the popular tool of reinforcement learning [49] for optimizing the system performance in a panoramic video streaming/broadcasting scenario.

References

[1] J. P. Snyder, *Flattening the earth: two thousand years of map projections*. University of Chicago Press, Chicago, USA, 1993.

[2] S. H. Lee, S. T. Kim, E. Yip, B. D. Choi, J. Song, and S. J. Ko, "Omnidirectional video coding using latitude adaptive down-sampling and pixel rearrangement," *Electronics Letters*, vol. 53, no. 10, pp. 655–657, 2017.

[3] O. Schreer, I. Feldmann, C. Weissig, P. Kauff, and R. Schafer, "Ultrahigh-resolution panoramic imaging for format-agnostic video production," *Proceedings of the IEEE*, vol. 101, pp. 99–114, January 2013.

[4] Y. Xu, Q. Zhou, L. Gong, M. Zhu, X. Ding, and R. K. F. Teng, "High-speed simultaneous image distortion correction transformations for a multicamera cylindrical panorama real-time video system using FPGA," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, pp. 1061–1069, June 2014.

[5] C. W. Fu, L. Wan, T. T. Wong, and C. S. Leung, "The rhombic dodecahedron map: An efficient scheme for encoding panoramic video," *IEEE Transactions on Multimedia*, vol. 11, pp. 634–644, June 2009.

[6] P. R. Alface, J. F. Macq, and N. Verzijp, "Interactive omnidirectional video delivery: A bandwidth-effective approach," *Bell Labs Technical Journal*, vol. 16, pp. 135–147, March 2012.

[7] Q. Zhao, L. Wan, W. Feng, J. Zhang, and T. T. Wong, "Cube2video: Navigate between cubic panoramas in real-time," *IEEE Transactions on Multimedia*, vol. 15, pp. 1745–1754, December 2013.

[8] V. R. Gaddam, M. Riegler, R. Eg, C. Griwodz, and P. Halvorsen, "Tiling in interactive panoramic video: Approaches and evaluation," *IEEE Transactions on Multimedia*, vol. 18, pp. 1819–1831, September 2016.

[9] X. Ge, L. Pan, Q. Li, G. Mao, and S. Tu, "Multipath cooperative communications networks for augmented and virtual reality transmission," *IEEE Transactions on Multimedia*, vol. 19, pp. 2345–2358, October 2017.

[10] G. He, J. Hu, H. Jiang, and Y. Li, "Scalable video coding based on user's view for real-time virtual reality applications," *IEEE Communications Letters*, vol. 22, pp. 25–28, January 2018.

[11] C. Zhu, Y. Huo, B. Zhang, R. Zhang, M. El-Hajjar, and L. Hanzo, "Adaptive-truncated-HARQ-aided layered video streaming relying on interlayer FEC coding," *IEEE Transactions on Vehicular Technology*, vol. 65, pp. 1506–1521, March 2016.

[12] X. Gu, L. Zhang, B. Leng, Y. Wang, and L. Zhang, "Performance analysis for interconnected virtual reality with two-stage cooperative transmission scheme in 3D UDNs," *IEEE Access*, vol. 6, pp. 28162–28173, 2018.

[13] B. Masnick and J. Wolf, "On linear unequal error protection codes," *IEEE Transactions on Information Theory*, vol. 13, pp. 600–607, October 1967.

[14] Y. Huo, C. Hellge, T. Wiegand, and L. Hanzo, "A tutorial and review on inter-layer FEC coded layered video streaming," *IEEE Communications Surveys and Tutorials*, vol. 17, pp. 1166–1207, 2nd-quarter 2015.

[15] D. Song and C. W. Chen, "Scalable H.264/AVC video transmission over MIMO wireless systems with adaptive channel selection based on partial channel information," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, pp. 1218–1226, September 2007.

[16] C. Hellge, D. Gomez-Barquero, T. Schierl, and T. Wiegand, "Layer-aware forward error correction for mobile broadcast of layered media," *IEEE Transactions on Multimedia*, vol. 13, pp. 551–562, June 2011.

[17] Y. Huo, M. El-Hajjar, and L. Hanzo, "Inter-layer FEC aided unequal error protection for multilayer video transmission in mobile TV," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, pp. 1622–1634, September 2013.

[18] Y. Huo, M. El-Hajjar, R. G. Maunder, and L. Hanzo, "Layered wireless video relying on minimum-distortion inter-layer FEC coding," *IEEE Transactions on Multimedia*, vol. 23, pp. 319–331, January 2014.

[19] S. Cicalo and V. Tralli, "Distortion-fair cross-layer resource allocation for scalable video transmission in OFDMA wireless networks," *IEEE Transactions on Multimedia*, vol. PP, no. 99, pp. 1–1, 2014.

[20] F. Marx and J. Farah, "A novel approach to achieve unequal error protection for video transmission over 3G wireless networks," *Signal Processing: Image Communication*, vol. 19, no. 4, pp. 313–323, 2004.

[21] N. Rahnavard, H. Pishro-Nik, and F. Fekri, "Unequal error protection using partially regular LDPC codes," *IEEE Transactions on Communications*, vol. 55, pp. 387–391, March 2007.

[22] Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, *ITU-T Rec. H.264/ISO/IEC 14496-10 AVC: Advanced Video Coding for Generic Audiovisual Services*, March 2010.

[23] Nasruminallah and L. Hanzo, "EXIT-chart optimized short block codes for iterative joint source and channel decoding in H.264 video telephony," *IEEE Transactions on Vehicular Technology*, vol. 58, pp. 4306–4315, October 2009.

[24] S. X. Ng and J. Y. Chung and P. Cherriman and L. Hanzo, "Burst-by-Burst Adaptive Decision Feedback Equalised TCM, TTCM and BICM for H.263-Assisted Wireless Video Telephony," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, pp. 363–374, March 2006.

[25] M. Aydinlik and M. Salehi, "Turbo coded modulation for unequal error protection," *IEEE Transactions on Communications*, vol. 56, pp. 555–564, April 2008.

[26] Y. C. Chang, S. W. Lee, and R. Komiya, "A fast forward error correction allocation algorithm for unequal error protection of video transmission over wireless channels," *IEEE Transactions on Consumer Electronics*, vol. 54, pp. 1066–1073, August 2008.

[27] J. Micallef, R. Farrugia, and C. Debono, "Correlation noise based unequal error protected rate-adaptive codes for distributed video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. PP, no. 99, pp. 1–1, 2013.

[28] Y. Huo, T. Wang, R. G. Maunder, and L. Hanzo, "Motion-aware mesh-structured trellis for correlation modelling aided distributed multi-view video coding," *IEEE Transactions on Image Processing*, vol. 23, pp. 319–331, January 2014.

[29] Y. Huo, P. T. Kovács, T. J. Naughton, and L. Hanzo, "Wireless holographic image communications relying on unequal error protected bit-planes," *IEEE Transactions on Vehicular Technology*, vol. 66, pp. 7136–7148, August 2017.

[30] Y. Huo, M. El-Hajjar, and L. Hanzo, "Wireless video: An interlayer error-protection-aided multilayer approach," *IEEE Vehicular Technology Magazine*, vol. 9, pp. 104–112, September 2014.

[31] A. Vosoughi, P. C. Cosman, and L. B. Milstein, "Joint source-channel coding and unequal error protection for video plus depth," *IEEE Signal Processing Letters*, vol. 22, pp. 31–34, January 2015.

[32] D. Gómez-Barquero and O. Simeone, "LDM versus FDM/TDM for unequal error protection in terrestrial broadcasting systems: An information-theoretic view," *IEEE Transactions on Broadcasting*, vol. 61, pp. 571–579, December 2015.

[33] M. A. Khan, A. A. Moinuddin, E. Khan, and M. Ghanbari, "Optimized cross-layered unequal error protection for SPIHT coded wireless video transmission," *IEEE Transactions on Broadcasting*, vol. 62, pp. 876–889, December 2016.

[34] X. Song, X. Peng, J. Xu, G. Shi, and F. Wu, "Unequal error protection for scalable video storage in the cloud," *IEEE Transactions on Multimedia*, vol. 20, pp. 699–710, March 2018.

[35] K. Zheng, X. Zhang, Q. Zheng, W. Xiang, and L. Hanzo, "Quality-of-experience assessment and its application to video services in LTE networks," *IEEE Wireless Communications*, vol. 22, pp. 70–78, February 2015.

[36] H. Shen, W. D. Pan, and D. Wu, "Predictive lossless compression of regions of interest in hyperspectral images with no-data regions," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, pp. 173–182, January 2017.

[37] H. Hu, Y. Lin, M. Liu, H. Cheng, Y. Chang, and M. Sun, "Deep 360 pilot: Learning a deep agent for piloting through 360^o sports videos," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1396–1405, July 2017.

[38] M. Xu, Y. Song, J. Wang, M. Qiao, L. Huo, and Z. Wang, "Predicting head movement in panoramic video: A deep reinforcement learning approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, early access,2018.

[39] H. Ha and C. Yim, "Layer-weighted unequal error protection for scalable video coding extension of H.264/AVC," *IEEE Transactions on Consumer Electronics*, vol. 54, pp. 736–744, May 2008.

[40] X. Yang, C. Zhu, Z. Li, X. Lin, G. Feng, S. Wu, and N. Ling, "Unequal loss protection for robust transmission of motion compensated video over the Internet," *Signal Processing: Image Communication*, vol. 18, no. 3, pp. 157–167, 2003.

[41] E. Maani and A. Katsaggelos, "Unequal error protection for robust streaming of scalable video over packet lossy networks," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, pp. 407–416, March 2010.

[42] R. Khalili and K. Salamatian, "A new analytic approach to evaluation of packet error rate in wireless networks," in *Communication Networks and Services Research Conference*, (Halifax, Nova Scotia, Canada), pp. 333–338, May 2005.

[43] Y. Huo, C. Zhou, J. Jiang, and L. Hanzo, "Historical information aware unequal error protection of scalable HEVC/H.265 streaming over free space optical channels," *IEEE Access*, vol. 4, pp. 5659–5672, 2016.

[44] P. R. Desai, P. N. Desai, K. D. Ajmera, and K. Mehta, "A review paper on Oculus rift-a virtual reality headset," *International Journal of Engineering Trends and Technology*, vol. 13, no. 4, 2014.

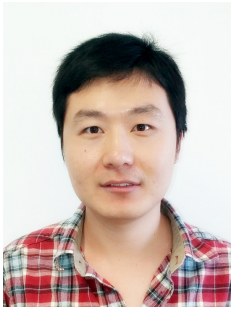
[45] Z.-H. Zhan, J. Zhang, Y. Li, and H.-H. Chung, "Adaptive particle swarm optimization," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 39, pp. 1362–1381, December 2009.

[46] "The Panoramic Race Video." Available: https://youtu.be/CcAi_BjZ-Jc.

[47] "The Panoramic RollerCoaster Video." Available: https://youtu.be/s9njwl_VzZA.

[48] ITU-T, *H.265 : High efficiency video coding*, April 2015.

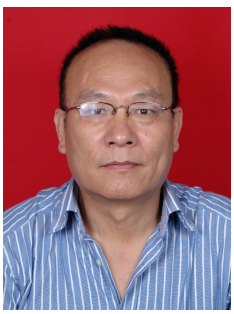
[49] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "Deep reinforcement learning: A brief survey," *IEEE Signal Processing Magazine*, vol. 34, pp. 26–38, November 2017.



Yongkai Huo received the B.Eng. degree with distinction in computer science and technology from Hefei University of Technology, Hefei, China, in 2006 and the M.Eng. degree in computer software and theory from University of Science and Technology of China, Hefei, China, in 2009. In 2014, he was awarded a Ph.D in Wireless Communications group, School of Electronics and Computer Science, University of Southampton, Southampton, UK, where he continued to work as a research fellow until September 2016. He received a scholarship under the China-U.K. Scholarships for Excellence Programme. He is now an associate professor with the College of Computer Science&Software Engineering, Shenzhen University, China. His research interest include, deep learning, panoramic video coding/streaming, holographic video, distributed video coding, multiview video coding, robust wireless video streaming and joint source-channel decoding etc.

Xu Wang received the B.S. degree from South China Normal University, Guangzhou, China, in 2007, the M.S. degree from Ningbo University, Ningbo, China, in 2010, and the Ph.D. degree from the Department of Computer Science, City University of Hong Kong, Hong Kong, in 2014. In 2015, he joined the College of Computer Science and Software Engineering, Shenzhen University as an Assistant Professor. His research interests are video coding and stereoscopic image/video quality assessment.

Peichang Zhang (Student member, IEEE) received the M.Sc. degree in computer science from the Budapest University of Technology, Budapest, Hungary in 2004. He is currently pursuing the Ph.D. degree in signal processing at Tampere University of technology, Tampere, Finland. From 2004 to 2006, he was a Software Engineer with Archi-Data. Since 2006, he has been Software Engineer, Lead Software Engineer, then CTO of Holografika. He has been a visiting researcher at Tampere University of Technology from 2013 to 2014. He is the author or co-author of three book chapters, three journal papers and more than 30 conference papers. His research interests include 3D displays, more specifically light-field displays, and the capture / compression / rendering of light fields. He has served as Program Committee member for numerous international IEEE conferences, Local Organizing Chair of 3DTV-Con 2014, and is a contributing member of the International 3D Society and the International Committee for Display Metrology (ICDM), where he contributed to the first IDMS standard. He was the Head of Delegation to MPEG for Hungary. He also served as the chair of the Working Group 5 ("3D End-User Devices") of 3D-ConTourNet COST Action.



Jianmin Jiang received PhD from the University of Nottingham, UK, in 1994. From 1997 to 2001, he worked as a full professor of Computing at the University of Glamorgan, Wales, UK. In 2002, he joined the University of Bradford, UK, as a Chair Professor of Digital Media, and Director of Digital Media & Systems Research Institute. He worked at the University of Surrey, UK, as a full professor during 2010-2014 and a distinguished chair professor (1000-plan) at Tianjin University, China, during 2010-2013. He is currently a Distinguished Chair Professor

and director of the Research Institute for Future Media Computing at the College of Computer Science & Software Engineering, Shenzhen University, China. He was a chartered engineer, fellow of IEE, fellow of RSA, member of EPSRC College in the UK, and EU FP-6/7 evaluator. His research interests include, image/video processing in compressed domain, digital video coding, medical imaging, computer graphics, machine learning and AI applications in digital media processing, retrieval and analysis. He has published around 400 refereed research papers.

Lajos Hanzo (F' 04), Fellow of EURASIP, DSc received his degree in electronics in 1976 and his doctorate in 1983. In 2009 he was awarded an honorary doctorate by the Technical University of Budapest and in 2015 by the University of Edinburgh. In 2016 he was admitted to the Hungarian Academy of Science. During his 40-year career in telecommunications he has held various research and academic posts in Hungary, Germany and the UK. Since 1986 he has been with the School of Electronics and Computer Science, University of Southampton, UK, where he holds the chair in telecommunications. He has successfully supervised 121 PhD students, co-authored 18 John Wiley/IEEE Press books on mobile radio communications totalling in excess of 10 000 pages, published 1800+ research contributions at IEEE Xplore, acted both as TPC and General Chair of IEEE conferences, presented keynote lectures and has been awarded a number of distinctions. Currently he is directing a 60-strong academic research team, working on a range of research projects in the field of wireless multimedia communications sponsored by industry, the Engineering and Physical Sciences Research Council (EPSRC) UK, the European Research Council's Advanced Fellow Grant and the Royal Society's Wolfson Research Merit Award. He is an enthusiastic supporter of industrial and academic liaison and he offers a range of industrial courses. He is also a Governor of the IEEE VTS. During 2008 - 2012 he was the Editor-in-Chief of the IEEE Press and a Chaired Professor also at Tsinghua University, Beijing. For further information on research in progress and associated publications please refer to http://www-mobile.ecs.soton.ac.uk