IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 15 April 2025, accepted 2 July 2025, date of publication 10 July 2025, date of current version 17 July 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3588095

RESEARCH ARTICLE

Efficient Region-Wise Packing of Stereoscopic ERP Videos Based on Information Loss Minimization

HOSSEIN PEJMAN[1], (Student Member, IEEE), STÉPHANE COULOMBE[1,2], (Senior Member, IEEE), CARLOS VÁZQUEZ[1], (Senior Member, IEEE), AND AHMAD VAKILI[3]

[1]Department of Software and IT Engineering, École de technologie supérieure, Montreal, QC H3C 1K3, Canada
[2]International Laboratory on Learning Systems (ILLS), McGill-ÉTS-Mila-CNRS-CentraleSupelec-Université Paris Saclay, Montreal, QC H3H 2T2, Canada
[3]Department of Research and Development, Summit Tech Multimedia, Montreal, QC H2N 1N2, Canada

Corresponding author: Hossein Pejman (hossein.pejman-tavallaee.1@ens.etsmtl.ca)

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada; in part by Mitacs; and in part by the Summit Tech Multimedia through the Mitacs Accelerate Program and the Alliance-Mitacs Accelerate Program under Grant IT19631, Grant ALLRP 585980-23, and Grant IT35906.

ABSTRACT Utilizing frame-compatible (FC) formats for packing stereoscopic videos often comes with challenges, as they require higher transmission bandwidth and larger memory buffers on the decoder compared to single-view videos. When it comes to stereoscopic 360° videos, as the primary content consumed by virtual reality (VR) applications, these requirements become even more challenging since they ask for ultra-high-resolution formats with high frame rates (e.g., 6K, 8K, or 12K at 100 frames per second). To address these challenges, sub-sampled versions of the left and right views are usually used to form the spatial FC format, leading to a loss of visual quality. In this paper, we propose an efficient region-wise packing method for equirectangular projection (ERP) videos with minimum information loss by exploiting the uneven sampling characteristic of ERP. Moreover, we propose a content-adaptive (CA) packing method for ERP videos, where the sizes of partitions, each with a particular horizontal downsampling factor, are adaptively determined based on spatial complexity. We then utilize a low-complexity frequency-domain approach to estimate the optimal partition sizes of the CA packing. We use these proposed methods to determine the optimal packing of the stereoscopic ERP videos in the FC format. Experimental results, using the VVenC Versatile Video Coding (VVC) encoder, show that compared with the standard side-by-side (SbS) format, with uniform horizontal half-downsampling (UHHDS), the proposed CA packing method provides an average 13.84% and 12.02% Bjøntegaard-Delta bitrate (BD-BR) reduction for Random Access (RA) and Low Delay B (LDB) configurations, respectively, with an average encoding time comparable to SbS. In addition, when the performance is measured based on user attention probability, using the Laplacian Distribution model, the coding performance of our proposed packing methods outperforms the state-of-the-art packing method with significantly lower computational complexity.

INDEX TERMS Region-wise packing, frame-compatible formats, stereoscopic 360° video, equirectangular projection, downsampling, discrete Fourier transform.

I. INTRODUCTION

Compression techniques for 3D videos have consistently been part of video coding standards. The multiview extension of High Efficiency Video Coding (HEVC) [1], MV-HEVC [2], enables 3D video encoding. Similarly, in the Versatile Video Coding (VVC) [3], the multi-layer profile enables 3D (multi-view) video coding [4]. Recently, the MPEG Immersive Video (MIV) coding standard [5] added support for immersive content captured by multiple cameras

The associate editor coordinating the review of this manuscript and approving it for publication was Jiachen Yang.

with six degrees of freedom (6DoF). Stereoscopic videos, as the simplest type of 3D video, can be encoded with the above-mentioned tools. However, there are limitations to these approaches for encoding stereoscopic videos, primarily the incompatibility of multi-view coding tools with most existing single-view video coding and transmission systems [6]. Another approach to encoding stereoscopic content is using frame-compatible (FC) formats [7]. In spatial FC formats, two views are packed into a single view by arranging them spatially, e.g., side-by-side (SbS), top-bottom (TB), or by dividing them into tiles (tile format) [7]. The main advantage of FC formats is full compatibility with existing codecs and delivery systems used for single-view videos [7]. However, this results in doubling the pixel count of the packed frame. Because of the limitations of existing video transmission systems in terms of bandwidth and limited resources on the display (user) side [8], such as the limited size of on-chip memory buffers on hardware decoders [9], the spatial resolution of videos in the spatial FC format is often reduced, resulting in degraded visual quality. This is particularly relevant for stereoscopic 360° applications usually dealing with ultra-high-resolution [10] and high frame-rate [11] content (e.g., 12K with 100 frames per second) requiring bandwidths of up to hundreds of megabits per second [10]. Solving this problem is important because stereoscopic 360° video is the preferred choice for producing content for virtual reality (VR) headsets. It offers a more realistic and immersive experience compared to monoscopic 360° video, due to its ability to provide depth perception. The VR technology is widely used in various medical [12], [13], training [14], [15], and entertainment [16], [17] applications, highlighting the importance of efficiently coding stereoscopic 360° videos as the primary content consumed by the VR applications.

In this paper, we focus on the equirectangular projection (ERP) format, because it is the most widely used projection format for 360° videos [18] and one of the two projection formats supported by Omnidirectional MediA Format (OMAF) [19]. It maps the 3D sphere to the 2D plane with uneven sampling density across latitudes, since all the sphere's circumferences are mapped to the same number of pixels [20]. As a result, uniform half-downsampling used for SbS and TB formats causes higher information loss (distortion) at the center (equator) of the frame where sampling density is the lowest, compared to the top and bottom (poles) where sampling density is the highest [18]. This suggests that the polar regions can be downsampled more aggressively without causing significant distortions. Moreover, because the horizontal borders of ERP videos wrap around, carefully packing views into the FC frame can potentially reduce discontinuity artifacts caused by packing.

In our previous work [21], by taking these ERP's characteristics into account, we proposed a low-complexity region-dependent downsampling packing (RDDP) method for stereoscopic ERP, using horizontal downsampling. In this paper, we refer to it as the Half Sampling Ratio Region-Wise Packing (HSR-RWP) method to highlight that it is a region-wise packing with a sampling ratio (SR) of 0.5. The term *SR* in this paper refers to the ratio of the number of pixels after resampling of the video, P_{rs}, to the number of pixels in the original video, P_{org}:

$$SR = \frac{P_{rs}}{P_{org}} \tag{1}$$

The HSR-RWP method efficiently reduces ERP's pixel redundancy while preserving the quality of the central areas, where human attention is the highest [22], [23]. It is tailored for packing stereoscopic ERP with $SR = 0.5$ (similar to SbS format with uniform half downsampling) to guarantee compatibility with video coding and transmission systems designed for single-view videos and can be considered as a type of OMAF region-wise packing format [24].

In this paper, we extend our previous work and propose a generalized, efficient region-wise packing for ERP frames based on the uneven sampling density of ERP's rows. It adopts the general approach used in the HSR-RWP method, which involves aggressively downsampling polar regions (top and bottom of the ERP frame) and preserving the original resolution in the central area of the ERP frame, as the regions with the highest user attention (*equator bias*) [22], [23]. However, in contrast to the HSR-RWP method, this approach can be employed to determine the optimized packing for an arbitrary *SR*. Moreover, we present a content-adaptive packing method, in which the height of partitions (with different horizontal downsampling factors) can be adaptively adjusted according to the spatial complexity of the ERP content. The main contributions of this paper can be summarized as follows:

- We propose a generalized, efficient region-wise packing method for ERP frames. It can be optimized based on the uneven sampling density of ERP's rows, determining the optimal size of regions for a desired SR. It is independent of the packing layout.
- We propose a content-adaptive packing method. The size of partitions in the packing method is adaptively adjusted according to the spatial complexity of the frame's rows. Moreover, we introduce a low-complexity approach in the frequency domain to estimate the optimal size of partitions.
- We use a special partition flipping in the layout of the proposed stereoscopic packing methods to alleviate seam artifacts [25] caused by vertical discontinuity borders. This technique also reduces the number of pixel rows needed for padding.

The remainder of this paper is organized as follows. Section II briefly reviews previous works related to stereoscopic and ERP frame packing methods. Section III describes the proposed methods. Section IV presents the experimental results. Finally, Section V concludes the paper.

II. RELATED WORK

In this section, we first review prior studies in the literature related to the FC format and stereoscopic packing methods. Then, we briefly review the HSR-RWP method introduced in our previous work [21].

A. REVIEW OF VIDEO PACKING METHODS

Prior research on stereoscopic FC formats has focused on non-360° videos [26], [27], [28]. Consequently, the unique characteristics of ERP videos, such as uneven sampling density or being horizontally borderless, were not considered in these approaches. The primary goal of these studies was to maintain the visual information of downsampled views by employing various techniques, such as incorporating FC formats with enhancement layers in multi-view coding (MVC) [27], or new interpolation methods using disparity information between views or temporal correlations between frames of individual views [28].

Regarding monoscopic 360° videos, improving the coding efficiency of 360° videos by taking the uneven sampling density characteristics of ERP into account has been a topic of interest in previous studies [20], [29], [30], [31]. This is usually done by applying different downsampling factors to different regions of the ERP frame. For instance, this is performed in [29], where the number of tiles (downsampled regions) and their sizes are optimized based on bit allocation and sample count constraints for each tile. However, solving this optimization problem requires encoding tiles with different quantization parameters (QPs), which adds significant computational complexity. In [20], the authors found that the user attention for areas with an absolute latitude greater than $\frac{\pi}{3}$ was considerably lower compared to other regions. Accordingly, they proposed a packing method in which these areas in ERP are horizontally downsampled by a factor of 2, while the remaining range of latitudes retains the original size of the ERP video. In [30], the authors proposed a packing scheme in which the polar regions of ERP are mapped to nested polygonal shapes with a sampling density similar to the equator of the sphere. A drawback of this representation is that the process of nested chain packing is computationally complex.

In [31], a tile-based segmentation method for ERP was proposed in which polar tiles can be mapped either to a circular or a square shape. The coding performance of this method for the all-intra mode is better than ERP. However, this method is better characterized as a projection format rather than a packing method, as it involves operations more complex than simple downsampling, particularly for mapping polar regions to circular shapes. Although most of the proposed methods for monoscopic 360° videos can be adapted for packing stereoscopic 360° videos with FC formats, they are usually computationally complex, as they typically require intricate remapping of some regions or an optimization process for packing.

In [32], two region-wise mixed-resolution packing schemes for 6K and 8K ERP contents are proposed to stream them as 4K content. They are specifically designed for applications limited to a maximum video decoding resolution of 4K. To generate the bitstream of these packing methods, three versions of the source content at different resolutions must be encoded and stored on the server side. Additionally, packing 8K ERP sources involves a temporal interleaving operation and requires twice the frame rate of the source content, imposing extra complexity on the decoder. In [33] and [34], a spatially adaptive QP adjustment method was proposed to mitigate the negative impact of uneven sampling density in ERP on coding efficiency. This method adjusts the QP at the Coding Tree Unit (CTU) level based on the latitude of the block. Similarly, [35] proposed an adaptive QP approach considering both the content complexity and the latitude of CTU. Although using adaptive QP improves the coding efficiency of ERP, it does not change the resolution. Therefore, applying this method to stereoscopic ERP video cannot address the limitations of memory buffers on the decoder.

B. OVERVIEW OF THE HSR-RWP METHOD

In our previous work [21], we observed that for ERP videos the amount of distortion caused by horizontal downsampling was significantly lower than that caused by vertical downsampling, while for conventional 2D videos downsampling in either direction causes similar distortions. This implies that horizontal downsampling is a better choice when the ERP video needs to be resized to a lower resolution. We thus proposed a region-adaptive downsampling method, called HSR-RWP, based on horizontal downsampling for each view of the stereoscopic ERP. As shown in Fig. 1 in the proposed HSR-RWP method we have

- N rows of pixels at the top and bottom (poles) are downsampled by a factor of 4 (red region in Fig. 1).
- N rows of pixels at the center (equator) are packed without downsampling (green region in Fig. 1).
- The remaining pixel rows between the poles and center regions (middle-top and middle-bottom regions, represented in orange in Fig. 1) are downsampled by a factor of 2.

Indeed, the HSR-RWP method is achieved by modifying the uniform horizontal half-downsampling (UHHDS) used in SbS format where downsampling factors of the center and polar regions are changed to 1 and 4, respectively. In the end, all regions of the left and right views (shown in Fig. 1) are packed as illustrated in Fig. 2. The conditions regarding downsampling factors and the size (height) of different regions in Fig. 1 can be defined using the constraints below:

$$\begin{aligned} &S_P = 4, \quad S_M = 2 \\ &h_c = h_p = N, \quad h_m = (H - 3N)/2 \end{aligned} \tag{2}$$

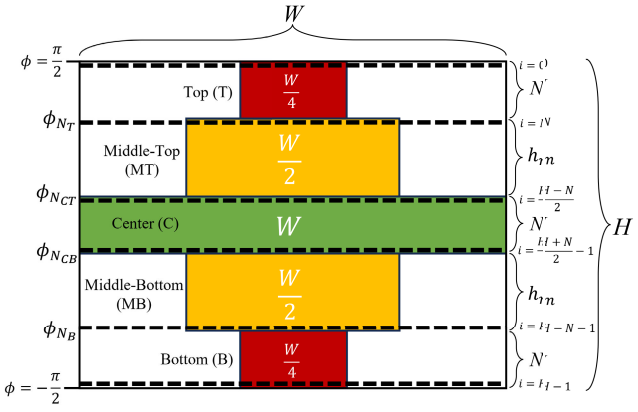

FIGURE 1. Region-adaptive downsampling of ERP, for each view, in the proposed HSR-RWP method. Dashed lines represent the rows associated with the highest absolute latitudes of each region.

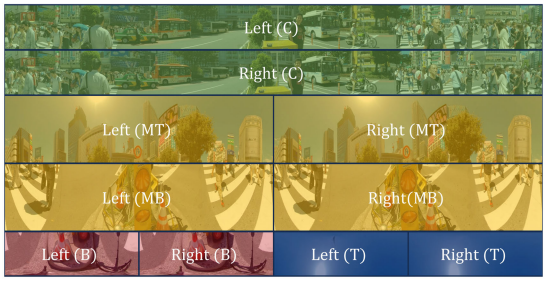

FIGURE 2. Illustration of the proposed HSR-RWP method ($S_M = 2$, $S_P = 4$) for stereoscopic ERP video.

where S_P is the downsampling factor at the poles, S_M is the downsampling factor of middle regions, and h_c, h_p, and h_m are the heights of the center, poles, and middle regions, respectively. By adhering to the constraints of Eq. (2), the SR with our HSR-RWP remains the same as that of UHHDS (SR= 0.5). This guarantees that the same transmission system used for monoscopic videos can be utilized for stereoscopic videos. Moreover, with respect to the constraints of Eq. (2), the value of N, corresponding to the latitude ϕ_{N_T} in Fig. 1, is the only factor affecting the overall downsampling distortion of HSR-RWP. It determines the number of rows that are kept intact (not downsampled) at the center, and that are downsampled by a factor of 2 or 4 at the middle and polar regions, respectively.

There are some points regarding the constraints of Eq. (2) that are worth further explanation:

- For simplicity, we assume that the middle regions (MT and MB regions in Fig. 2) have the same importance and are packed with the same downsampling factor. This is reasonable because they are equidistant from the equator, and thus have the same sampling density. This assumption is also true for the polar regions (T and B regions in Fig. 2).
- We only use three different horizontal downsampling factors, as using a higher number of downsampling factors increases the number of vertical discontinuities. Consequently, more pixels are required for padding to alleviate seam artifacts [31].

In this paper, we introduce two region-wise packing methods with SR=0.5 for ERP frames. One is optimized based on the uneven pixel density of ERP frames and the other uses the spatial frequency information of the video to minimize the information loss (distortion) caused by downsampling. The general approach used in these methods consists in keeping the region near the center of ERP intact and aggressively downsampling the poles. In contrast to projection formats, which usually involve computationally complex forward and backward mapping, both proposed packing methods only use simple horizontal downsampling operations.

III. PROPOSED PACKING METHODS

This section comprises two parts, each proposing an optimal region-wise packing approach aimed at minimizing information loss due to downsampling. The first method addresses this using general packing conditions (parameters) by considering the uneven sampling density characteristic of ERP. The second method is content-adaptive, determining the optimal height of the packing partitions to minimize energy loss (EL) in the frequency domain, using the Discrete Fourier Transform (DFT). The downsampling EL (distortion) varies according to the spatial complexity of the video content. Therefore, in contrast to the general packing method, the height of partitions in the content-adaptive method may not be the same for different video contents.

The optimization process used in the proposed packing methods can be utilized for packing with arbitrary layout and SR. However, in this paper, we focus on packing with SR=0.5 and a layout similar to HSR-RWP because it enables the transmission of stereoscopic videos in the same format (dimension) as monoscopic videos, and represents an application with practical interest for the proposed method.

Finally, we modify the layout of the proposed packing methods to alleviate seam artifacts at discontinuity borders caused by packing when subsequently coding with high QPs.

A. OPTIMIZING THE GENERAL REGION-WISE PACKING METHOD BY EXPLOITING THE UNEVEN SAMPLING DENSITY OF ERP

In this section, we propose a solution to determine the optimal parameters (size of partitions) of the general region-wise packing method for an arbitrary $SR > 0$, based on pixel information loss due to downsampling.

The general approach used in the packing method consists of using different downsampling factors, S_P and S_M, for the polar and middle regions, respectively, and maintaining the center region of ERP at its original resolution as the region receiving the highest user attention. In this packing method, S_M and S_P parameters can be any value as long as $S_P \geq S_M \geq 1$, while the N (ϕ_{N_T}) still adheres to the constraints of Eq. (2). This, in turn, means the optimal packing with a specific SR>0

can be identified by determining the optimal values of S_M, S_P, and ϕ_{N_T} among all possible cases. Obviously, we can optimize the packing method based on one parameter while keeping the other parameters fixed. This is investigated in Subsection III-A2 for the HSR-RWP method, as a particular case of the proposed general packing method, where, for the specific values of S_M and S_P, the optimal ϕ_{N_T} is determined. Moreover, the optimization process is independent of the layout used for packing, allowing partitions with optimal size to be arranged in any configuration to create a rectangular frame for single or multi-view ERP videos as required.

ERP maps the latitudes of the sphere with different circumference sizes to the same number of pixels in the 2D plane [20]. This implies that the sampling density on the 2D plane, associated with a given latitude of the sphere, increases from the equator toward the poles. This, in turn, causes *pixel redundancy* near the poles compared to the equator [36]. We define the (horizontal) pixel redundancy, $r(i)$, in ERP as follows:

$$\begin{aligned} r(i) &= 1 - \cos(\phi_i) \\ \phi_i &= \tfrac{\pi}{2}\left(1 - \tfrac{2i+1}{H}\right), \quad 0 \le i < H \end{aligned} \tag{3}$$

where i is the i-th row of pixels in the ERP video frame starting from the top, ϕ_i is the latitude on the sphere corresponding to the i-th row of pixels in ERP [37], $r(i)$ is the relative pixel redundancy in row i compared to the number of pixels at the equator, and H is the height of the ERP frame in pixels.

Downsampling row i of an ERP frame may introduce redundancy or result in information loss. However, using a downsampling factor of $\frac{1}{\cos(\phi_i)}$ for row i ensures a resized row without pixel redundancy or information loss. In this case, the width of the downsampled row i can be considered as the actual pixel information (API) of that row, and $\cos(\phi_i)$ represents the minimum *SR* required to theoretically preserve the API of row i. This means that applying a downsampling factor greater than $\frac{1}{\cos(\phi_i)}$ causes a loss in pixel information in row i. Let S_i denote the horizontal downsampling factor for latitude ϕ_i (row i). We define the pixel information loss ratio (PILR), $l(\phi_i)$, at latitude ϕ_i as:

$$l[i] = \begin{cases} \cos(\phi_i) - \dfrac{1}{S_i}, & \dfrac{1}{S_i} < \cos(\phi_i) \\ 0, & \dfrac{1}{S_i} \ge \cos(\phi_i) \end{cases}, \quad 0 \le i < H. \tag{4}$$

If $\frac{1}{S_i} \ge \cos(\phi_i)$, the width of the downsampled row is larger than the minimum theoretical width for preserving API, meaning the API of the row is effectively preserved. Conversely, $\frac{1}{S_i} < \cos(\phi_i)$ results in a loss of pixel information. In other words, information loss occurs at latitude ϕ_i if the value of $SR = \frac{1}{S_i}$ is lower than *SR* corresponding to API ($\cos(\phi_i)$) for that latitude.

1) PILR MINIMIZATION

Obviously, PILR varies depending on the height of the regions with different downsampling factors. The value of N determines the heights of these regions, as well as the height of the center packed with original resolution, implying that N affects the value of PILR. In Fig. 1, ϕ_{N_T} and ϕ_{N_B} denote the latitudes corresponding to the N-th row from the top and from the bottom of the frame, respectively, such that $\phi_{N_T} = -\phi_{N_B}$. For simplicity, we will use the notation ϕ_N instead of ϕ_{N_T} and $-\phi_{N_B}$, except when a distinction between the two is necessary. In [21], we showed that the range of all possible latitudes ϕ_N is $[\frac{\pi}{6}, \frac{\pi}{2}]$ and $[-\frac{\pi}{6}, -\frac{\pi}{2}]$ for the upper and lower half-parts of ERP, respectively.

Assuming that the number of rows is very large, as is the case in 360° contents with resolutions of 6K or more, we can consider ϕ_i as a continuous variable, denoted ϕ. Therefore, Eq. (4) can be expressed in the continuous domain as follows:

$$l(\phi) = \begin{cases} \cos(\phi) - \dfrac{1}{S_\phi}, & \dfrac{1}{S_\phi} < \cos(\phi) \\ 0, & \dfrac{1}{S_\phi} \ge \cos(\phi) \end{cases}, \quad \phi \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]. \tag{5}$$

where S_ϕ and $l(\phi)$ are horizontal downsampling factor and PILR for the continuous latitude ϕ, respectively.

As shown in Fig. 1, the height of regions with different downsampling factors for the upper and lower hemispheres in ERP is the same. This means the PILR due to downsampling is identical for both hemispheres. Moreover, because the center region between latitudes of $\phi_{N_{CB}}$ and $\phi_{N_{CT}}$ is not downsampled, the PILR for this region is zero. Therefore, to find the optimal value of N, finding the minimum PILR for the upper hemisphere suffices. From Eq. (5), the total PILR of the frame, $\mathcal{L}_I$, for a given N (which defines ϕ_{N_T}) in the continuous domain can be defined as follows:

$$\begin{aligned} \mathcal{L}_I(\phi_{N_T}) &= \int_{-\frac{\pi}{2}}^{\phi_{N_{CB}}} l(\phi)d\phi + \int_{\phi_{N_{CT}}}^{\frac{\pi}{2}} l(\phi)d\phi \\ \Rightarrow \mathcal{L}_I(\phi_{N_T}) &= 2\int_{\phi_{N_{CT}}}^{\frac{\pi}{2}} l(\phi)d\phi. \end{aligned} \tag{6}$$

Note that since 360° videos usually have $H > 1000$, the terms $\frac{\pi}{2H}$ could be neglected in the computations of Eq. (3). Therefore, based on Eq. (3), and from corresponding i values of ϕ_{N_T} and $\phi_{N_{CT}}$ in Fig. 1, $\phi_{N_{CT}}$ can be expressed in terms of ϕ_{N_T}:

$$\begin{aligned} \phi_{N_T} &\approx \frac{\pi}{2} - \frac{\pi \times N}{H} \Rightarrow N \approx \frac{H}{2}\left(1 - \frac{2\phi_{N_T}}{\pi}\right) \\ \text{and } \phi_{N_{CT}} &= \frac{\pi}{2}\left(1 - \frac{(H-N)}{H}\right) = \frac{\pi N}{2H} \\ \Rightarrow \phi_{N_{CT}} &= \frac{\pi}{4} - \frac{\phi_{N_T}}{2}. \end{aligned} \tag{7}$$

Considering the range of ϕ_{N_T} in the upper hemisphere, $\phi_{N_{CT}} = 0$ when $\phi_{N_T} = \pi/2$ and $\phi_{N_{CT}} = \pi/6$ when $\phi_{N_T} = \pi/6$. Therefore, the minimal value of $\phi_{N_{CT}}$ occurs

when ϕ_{N_T} is maximum and $\phi_{N_{CT}}$ increases as ϕ_{N_T} decreases until they meet at $\pi/6$. Thus, clearly, $\phi_{N_{CT}} \leq \phi_{N_T}$ for any value of ϕ_{N_T} as illustrated in Fig. 1.

To find the optimal latitude ϕ_{N_T}, we separately find the minimum PILR for the polar and middle regions of the upper hemisphere. In the general case, for a given latitude ϕ in the polar region ($\phi > \phi_{N_T}$) of the upper hemisphere with a downsampling factor of $S_\phi = S_P$, the PILR can be defined as follows:

$$l_P(\phi) = \begin{cases} \cos(\phi) - \frac{1}{S_P}, & \cos(\phi) > \frac{1}{S_P} \\ 0, & \cos(\phi) \leq \frac{1}{S_P} \end{cases}, \quad \phi \in \left[\phi_{N_T}, \frac{\pi}{2}\right]. \tag{8}$$

$$\mathcal{L}_P(\phi_{N_T}) = \int_{\phi_{N_T}}^{\frac{\pi}{2}} l_P(\phi)\, d\phi, \tag{9}$$

where $l_P(\phi)$ is the PILR in latitude ϕ and $\mathcal{L}_P(\phi_{N_T})$ is the total PILR of the polar region.

Similarly, the PILR for the middle region ($l_M(\phi)$ and $\mathcal{L}_M(\phi_{N_T})$) with downsampling factor $S_\phi = S_M$ is defined as follows:

$$l_M(\phi) = \begin{cases} \cos(\phi) - \frac{1}{S_M}, & \cos(\phi) > \frac{1}{S_M} \\ 0, & \cos(\phi) \leq \frac{1}{S_M} \end{cases}, \quad \phi \in \left[\phi_{N_{CT}}, \phi_{N_T}\right]. \tag{10}$$

$$\mathcal{L}_M(\phi_{N_T}) = \int_{\phi_{N_{CT}}}^{\phi_{N_T}} l_M(\phi)\, d\phi. \tag{11}$$

For the moment, we ignore the values proposed in Subsection II-B and make no assumption regarding the values of S_M and S_P except that $S_M \leq S_P$ since the polar regions exhibit higher pixel redundancy than the middle regions. As a result, the best ϕ_{N_T} in terms of lowest PILR can be found by minimizing the total PILR of polar and middle parts:

$$\mathcal{L}(\phi_{N_T}) = \mathcal{L}_P(\phi_{N_T}) + \mathcal{L}_M(\phi_{N_T}). \tag{12}$$

$$\phi^*_{N_T} = \underset{\phi_{N_T} \in [\frac{\pi}{6}, \frac{\pi}{2}]}{\operatorname{argmin}} \mathcal{L}(\phi_{N_T}). \tag{13}$$

To do this, we define ϕ_P and ϕ_M as thresholds representing latitudes above which the downsampling factors of S_M and S_P result in PILR=0 for the middle and polar regions, respectively. We have:

$$\begin{aligned} \cos(\phi) \leq \frac{1}{S_P} \Rightarrow \phi \geq \arccos\left(\frac{1}{S_P}\right) \\ \Rightarrow \phi_P \triangleq \arccos\left(\frac{1}{S_P}\right) \text{ and } l_P(\phi) = 0, \ \forall \phi \geq \phi_P \end{aligned}$$

$$\begin{aligned} \cos(\phi) \leq \frac{1}{S_M} \Rightarrow \phi \geq \arccos\left(\frac{1}{S_M}\right) \\ \Rightarrow \phi_M \triangleq \arccos\left(\frac{1}{S_M}\right) \text{ and } l_M(\phi) = 0, \ \forall \phi \geq \phi_M \end{aligned} \tag{14}$$

Since we assume that $S_M \leq S_P$, we have $\phi_M \leq \phi_P$. As can be seen in Fig. 1, $\phi_{N_{CT}}$ and ϕ_{N_T} are the lower and upper bounds of the middle-top (MT), and from Eq. (7), the value of $\phi_{N_{CT}}$ depends on ϕ_{N_T}. As illustrated in Fig. 3, considering the possible values of ϕ_{N_T} in the range ($\frac{\pi}{6} \leq \phi_{N_T} \leq \frac{\pi}{2}$), ϕ_M and ϕ_P ($0 < \phi_M \leq \phi_P$), we have six possible range cases to consider depending on the position of ϕ_{N_T} and $\phi_{N_{CT}}$ with respect to ϕ_M and ϕ_P:

$$\begin{cases} R_1: \ \phi_{N_T} < \phi_P \text{ and } \phi_{N_T} \geq \phi_{N_{CT}} \geq \phi_M \\ R_2: \ \phi_{N_T} < \phi_P \text{ and } \phi_{N_T} \geq \phi_M \geq \phi_{N_{CT}} \\ R_3: \ \phi_{N_T} < \phi_P \text{ and } \phi_M \geq \phi_{N_T} \geq \phi_{N_{CT}} \\ R_4: \ \phi_{N_T} \geq \phi_P \text{ and } \phi_P \geq \phi_{N_{CT}} \geq \phi_M \\ R_5: \ \phi_{N_T} \geq \phi_P \text{ and } \phi_P \geq \phi_M \geq \phi_{N_{CT}} \\ R_6: \ \phi_{N_T} \geq \phi_P \text{ and } \phi_{N_T} \geq \phi_{N_{CT}} \geq \phi_P \end{cases} \tag{15}$$

We can express Eq. (13) based on these six range cases as follows:

$$\begin{aligned} \phi^*_{N_T} &= \underset{\phi_{N_T} \in \{\phi^*_{R_k}\}_{k=1\ldots6}}{\operatorname{argmin}} \mathcal{L}(\phi_{N_T}), \\ \phi^*_{R_k} &= \underset{\phi_{N_T} \in R_k}{\operatorname{argmin}} \mathcal{L}(\phi_{N_T}), \ k = 1, 2, 3, 4, 5, 6 \end{aligned} \tag{16}$$

where $\phi^*_{R_k}$ represents the ϕ_{N_T} value that satisfies the range condition R_k and minimizes the sum of PILRs for both the polar and middle regions.

To solve this optimization problem, we have to find the minimum of $\mathcal{L}(\phi_{N_T}) = \mathcal{L}_P(\phi_{N_T}) + \mathcal{L}_M(\phi_{N_T})$ for each range case R_k and identify the minimum among them.

a: POLAR REGION FOR RANGE CASES R_1, R_2, AND R_3

From Eq. (9), we have:

$$\mathcal{L}_P(\phi_{N_T}) = \int_{\phi_{N_T}}^{\phi_P} l_P(\phi)\, d\phi + \int_{\phi_P}^{\frac{\pi}{2}} l_P(\phi)\, d\phi. \tag{17}$$

But according to the definition of ϕ_P in Eq. (14), it follows that:

$$l_P(\phi) = 0, \forall \phi \in [\phi_P, \ \frac{\pi}{2}] \Rightarrow \int_{\phi_P}^{\frac{\pi}{2}} l_P(\phi)\, d\phi = 0. \tag{18}$$

Therefore,

$$\begin{aligned} \mathcal{L}_P(\phi_{N_T}) &= \int_{\phi_{N_T}}^{\phi_P} \left(\cos(\phi) - \frac{1}{S_P}\right) d\phi \\ &= \sin(\phi_P) - \frac{\phi_P}{S_P} - \sin(\phi_{N_T}) + \frac{\phi_{N_T}}{S_P}. \end{aligned} \tag{19}$$

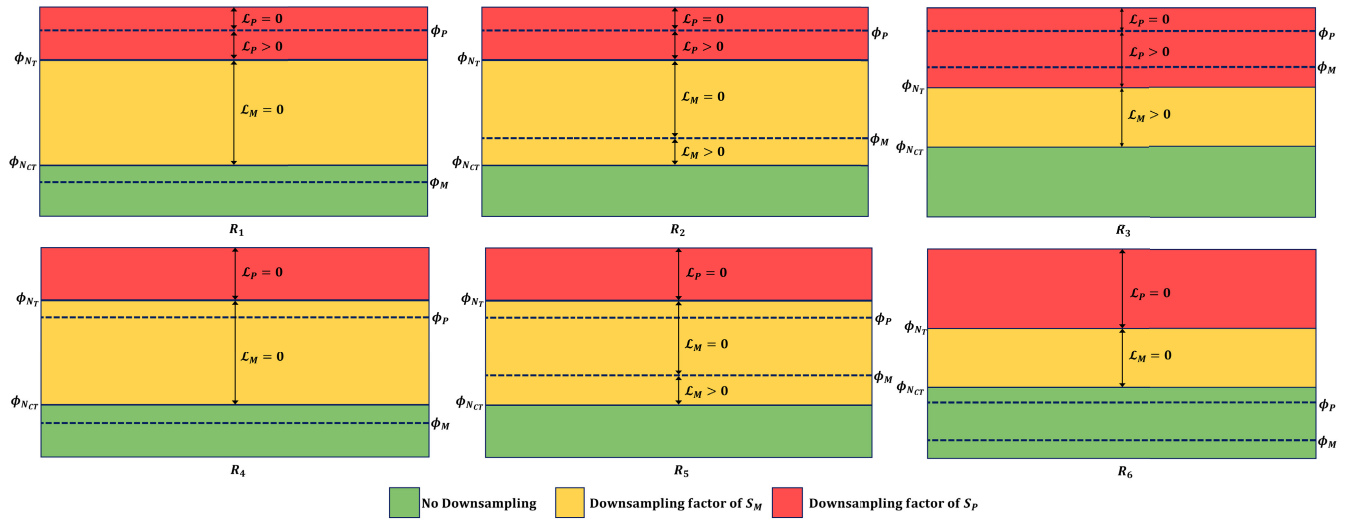

FIGURE 3. Illustrations of possible situations of ϕ_{N_T}, ϕ_M, and ϕ_P for ranges cases R_1 to R_6. For range cases with $\phi_{N_T} \geq \phi_P$ (R_4 to R_6), the PILR due to downsampling in the polar regions ($\mathcal{L}_P$) is zero. In the middle regions, $\mathcal{L}_M = 0$ if $\phi_{N_{CT}} \geq \phi_M$ (cases R_1, R_4, and R_6).

b: POLAR REGION FOR RANGE CASES R_4, R_5, AND R_6

In these range cases, because $\phi_{N_T} \geq \phi_P$, the entire polar area (the red region in Fig. 3) is always above ϕ_P. As a result, the PILR due to downsampling with factor S_P in the polar region is always zero. Thus, we can write:

$$\mathcal{L}_P(\phi_{N_T}) = \int_{\phi_{N_T}}^{\frac{\pi}{2}} l_P(\phi)\, d\phi = 0, \forall \phi_{N_T} \in \left[\phi_P, \frac{\pi}{2}\right]. \quad (20)$$

c: MIDDLE REGION FOR RANGE CASES R_1, R_2, AND R_3

Regarding $\mathcal{L}_M(\phi_{N_T})$, for R_1, if $\phi_{N_{CT}} \geq \phi_M$, the entire middle range is packed at latitudes above ϕ_M (yellow region in Fig. 3). Therefore, PILR caused by downsampling of the middle region with the factor of S_M is zero:

$$\mathcal{L}_M(\phi_{N_T}) = 0, \quad \forall \phi_{N_T} \text{ satisfying } R_1. \quad (21)$$

For R_2, $\phi_{N_T} \geq \phi_M \geq \phi_{N_{CT}}$ means ϕ_M is inside the MT area, such that, from Eq. (11), we have:

$$\mathcal{L}_M(\phi_{N_T}) = \int_{\phi_{N_{CT}}}^{\phi_M} l_M(\phi)\, d\phi + \int_{\phi_M}^{\phi_{N_T}} l_M(\phi)\, d\phi. \quad (22)$$

But, by definition of ϕ_M in Eq. (14), the second term is zero. Therefore, it follows, using Eq. (7), that:

$$\begin{aligned}\mathcal{L}_M(\phi_{N_T}) &= \int_{\frac{\pi}{4}-\frac{\phi_{N_T}}{2}}^{\phi_M} \left(\cos(\phi) - \frac{1}{S_M}\right) d\phi \\ &= \sin(\phi_M) + \frac{1}{S_M}\left(\frac{\pi}{4} - \phi_M\right) \\ &\quad - \sin\left(\frac{\pi}{4} - \frac{\phi_{N_T}}{2}\right) - \frac{\phi_{N_T}}{2S_M}. \end{aligned} \quad (23)$$

For R_3, the whole middle region is subject to PILR> 0, thus $\mathcal{L}_M(\phi_{N_T})$ is calculated as follows:

$$\begin{aligned}\mathcal{L}_M(\phi_{N_T}) &= \int_{\frac{\pi}{4}-\frac{\phi_{N_T}}{2}}^{\phi_{N_T}} \left(\cos(\phi) - \frac{1}{S_M}\right) d\phi \\ &= \sin(\phi_{N_T}) - \sin\left(\frac{\pi}{4} - \frac{\phi_{N_T}}{2}\right) \\ &\quad - \frac{3\phi_{N_T}}{2S_M} + \frac{\pi}{4S_M}. \end{aligned} \quad (24)$$

TABLE 1. Equations for computing the total PILR for all ranges R_k.

Ranges	$\mathcal{L}(\phi_{N_T}) = \mathcal{L}_P(\phi_{N_T}) + \mathcal{L}_M(\phi_{N_T})$
R_1	$\sin(\phi_P) - \frac{\phi_P}{S_P} - \sin(\phi_{N_T}) + \frac{\phi_{N_T}}{S_P}$
R_2	$c1 - \sin\left(\frac{\pi}{4} - \frac{\phi_{N_T}}{2}\right) - \sin(\phi_{N_T}) + \phi_{N_T}\left(\frac{1}{S_P} - \frac{1}{2S_M}\right)$ with $c1 = \sin(\phi_M) + \frac{1}{S_M}\left(\frac{\pi}{4} - \phi_M\right) + \sin(\phi_P) - \frac{\phi_P}{S_P}$
R_3	$c2 - \sin\left(\frac{\pi}{4} - \frac{\phi_{N_T}}{2}\right) + \phi_{N_T}\left(\frac{1}{S_P} - \frac{3}{2S_M}\right)$ with $c2 = \frac{\pi}{4S_M} + \sin(\phi_P) - \frac{\phi_P}{S_P}$
R_4	0
R_5	$c3 - \sin\left(\frac{\pi}{4} - \frac{\phi_{N_T}}{2}\right) - \frac{\phi_{N_T}}{2S_M}$ with $c3 = \sin(\phi_M) + \frac{1}{S_M}\left(\frac{\pi}{4} - \phi_M\right)$
R_6	0

d: MIDDLE REGION FOR RANGE CASES R_4, R_5, AND R_6

As shown in Fig. 3, similar to R_1, $\mathcal{L}_M(\phi_{N_T})$ for the range cases R_4 and R_6 is always zero. Finally, the equation for calculating $\mathcal{L}_M(\phi_{N_T})$ for R_5 is given by Eq. (23) since this case is similar to R_2.

e: FINAL EQUATIONS FOR ALL CASE RANGES

The equations to compute the total PILR for all range cases R_k are summarized in Table 1. We will see that, depending on the values of S_M and S_P, a subset of these range cases will apply, leading to the determination of the associated optimal value of ϕ_{N_T}.

2) APPLICATION TO THE HSR-RWP WITH SR=0.5

HSR-RWP is a particular case of the general packing method with $S_M = 2$ and $S_P = 4$. For the HSR-RWP method, ϕ_M and ϕ_P can be computed according to the following downsampling factors:

$$S_M = 2, \;\; \phi_M = \arccos\left(\frac{1}{S_M}\right) = \frac{\pi}{3} = 60^\circ$$
$$S_P = 4, \;\; \phi_P = \arccos\left(\frac{1}{S_P}\right) \approx 0.419\pi \approx 75.5^\circ \quad (25)$$

Moreover, considering the value of ϕ_M for HSR-RWP and the range of $\phi_{N_{CT}}$ we have (from Eq. (7)):

$$\phi_{N_{CT}} = \frac{\pi}{4} - \frac{\phi_{N_T}}{2}, \; \phi_{N_T} \in [\frac{\pi}{6}, \frac{\pi}{2}] \Rightarrow \phi_{N_{CT}} \in [0, \frac{\pi}{6}]$$
$$\phi_M = \frac{\pi}{3}, \; \phi_{N_{CT}} = [0, \frac{\pi}{6}] \Rightarrow \phi_M > \phi_{N_{CT}} \quad (26)$$

This means the possible range cases in Table 1 for the HSR-RWP method are R_2, R_3, and R_5. Therefore, to determine the ϕ_{N_T} that minimizes PILR, we need to minimize $\mathcal{L}(\phi_{N_T})$ for R_2, R_3, and R_5. It can be shown that the minimum occurs when the derivative of $\mathcal{L}(\phi_{N_T})$ for R_2 is zero:

$$\frac{d\mathcal{L}(\phi_{N_T})}{d\phi_{N_T}} = 0 \Rightarrow \phi_{N_T} = \frac{\pi}{2} - 2\arcsin(\frac{1}{4})$$
$$\Rightarrow \underset{\phi_{N_T} \in R_2}{\operatorname{argmin}} \mathcal{L}(\phi_{N_T}) = \frac{\pi}{2} - 2\arcsin(\frac{1}{4})$$
$$\rightarrow \mathcal{L}(\frac{\pi}{2} - 2\arcsin(\frac{1}{4})) = \mathcal{L}(61.04^\circ) = 0.24884. \quad (27)$$

The derivative of $\mathcal{L}(\phi_{N_T})$ relative to ϕ_{N_T} for R_3 is *strictly decreasing* and therefore, $\mathcal{L}(\phi_{N_T})$ is minimum when ϕ_{N_T} is equal to the upper bound of range ϕ_{N_T} for R_3:

$$\frac{d\mathcal{L}(\phi_{N_T})}{d\phi_{N_T}} < 0, \; \forall \phi_{N_T} \in R_3$$
$$\Rightarrow \underset{\phi_{N_T} \in R_3}{\operatorname{argmin}} \mathcal{L}(\phi_{N_T}) = \phi_M = \frac{\pi}{3}$$
$$\rightarrow \mathcal{L}(\frac{\pi}{3}) = \mathcal{L}(60^\circ) = 0.24899. \quad (28)$$

Finally, $\frac{d\mathcal{L}(\phi_{N_T})}{d\phi_{N_T}}$ for R_5 is *strictly increasing*, and therefore, $\mathcal{L}(\phi_{N_T})$ in the lower bound range ϕ_{N_T} R_5 is minimum:

$$\frac{d\mathcal{L}(\phi_{N_T})}{d\phi_{N_T}} > 0, \; \forall \phi_{N_T} \in R_5 \Rightarrow \underset{\phi_{N_T} \in R_5}{\operatorname{argmin}} \mathcal{L}(\phi_{N_T}) = \phi_P$$
$$\rightarrow \mathcal{L}(\phi_P) = \mathcal{L}(75.5^\circ) = 0.27949. \quad (29)$$

Therefore, according to Eq. (16) we have:

$$\phi^*_{N_T} = \underset{\phi_{N_T} \in \{60^\circ, 61.04^\circ, 75.5^\circ\}}{\operatorname{argmin}} \mathcal{L}(\phi_{N_T}) = 61.04^\circ \quad (30)$$

As mentioned earlier, in the general case, the equations of Table 1 can be utilized for finding ϕ_{N_T} minimizing PILR of downsampling factors of S_M ($S_M \geq 1$), S_P ($S_P \geq 1$) with $S_P \geq S_M$. However, some points are worth noting:

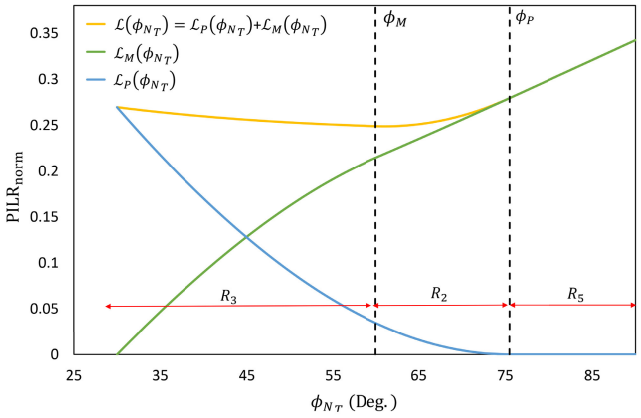

FIGURE 4. PILR of polar and middle parts for different ϕ_{N_T} values in R_k ranges for the HSR-RWP method.

- Since $S_M \geq 1$ and $S_P \geq 1$, the downsampling operations are optional in both regions. However, the constraint $S_P \geq S_M$ must be met.
- Without adhering to constraints of S_M and S_P in Eq. (2), it is not feasible to make a rectangular packing with SR=0.5 using the layout shown in Fig. 2, and without adding inactive pixels.
- The SR (resolution) of the HSR-RWP method, in which S_M and S_P follow the constraint Eq. (2), remains consistent across various ϕ_{N_T} values. However, for other values of S_M and S_P, the SR of packing varies according to the selected ϕ_{N_T}. In other words, a given SR can be attained through several sets of ϕ_{N_T}, S_M, and S_P values. Therefore, achieving optimal packing for a desired SR with minimal PILR requires considering both ϕ_{N_T} and downsampling factors of S_M and S_P. This may be considered further in future research.
- In scenarios where R_4 or R_6 is a possible range, there might exist multiple latitudes with $\mathcal{L}(\phi_{N_T})$=0. Each of these latitudes can be considered optimal for the corresponding SR resulting from it.

For digital images with discrete variable i representing the pixel rows of the image, the total PILR defined in Eq. (6) must be computed in the discrete domain as follows:

$$\mathcal{L}_I(\phi_{N_T}) = 2\int_{\frac{\pi}{2}-\frac{\phi_{N_T}}{2}}^{\frac{\pi}{2}} l(\phi)\, d\phi \approx \frac{2\pi}{H} \sum_{i=0}^{\frac{H-N}{2}-1} l[\phi_i]. \quad (31)$$

Fig. 4 depicts the PILR of polar and middle parts for different ϕ_{N_T} values in the various range cases R_k with S_M and S_P used for HSR-RWP (PILR is divided by $\frac{\pi}{H}$ for normalization). As can be seen, the minimum PILR for ϕ_{N_T} around 60° is almost constant. This can also be inferred from Eq. (28) and Eq. (27), where the PILR of $\phi_{N_T} = 61.04^\circ$ and $\phi_{N_T} = 60^\circ$ are almost the same. Therefore, the $\phi_{N_T} = 60^\circ$ used for HSR-RWP in our previous work [21] provides a near-optimal PILR.

Moreover, to verify that the value of $\phi_{N_T} = 60^\circ$ used for HSR-RWP yields the minimum downsampling distortion,

TABLE 2. WS-PSNR-Y of the proposed HSR-RWP method with different partitioning latitudes: for optimal with the highest WS-PSNR ($\phi_N = \phi_O$), $\phi_N = 60°$, and for uniform-downsampling (UD) ($\phi_N = \phi_{UD} = 90°$).

Sequence	Latitude (Deg.)	WS-PSNR-Y (dB)		
	ϕ_O	$\phi_N = \phi_O$	$\phi_N = 60°$	$\phi_N = \phi_{UD}$
Balboa	46	49.87	48.77	46.02
BranCastle2	60	41.37	41.37	39.75
Broadway	51	49.27	48.53	45.91
ChairliftRide	60	53.94	53.94	51.46
Gaslamp	53	53.14	52.96	49.87
Harbor	56	52.52	52.37	48.71
KiteFlite	62	50.13	50.06	47.71
Landing2	51	46.92	46.71	44.97
SkateboardInLot	54	57.87	57.45	49.88
Trolley	60	48.77	48.77	46.09
Average	**55.3**	**50.38**	**50.09**	**47.03**

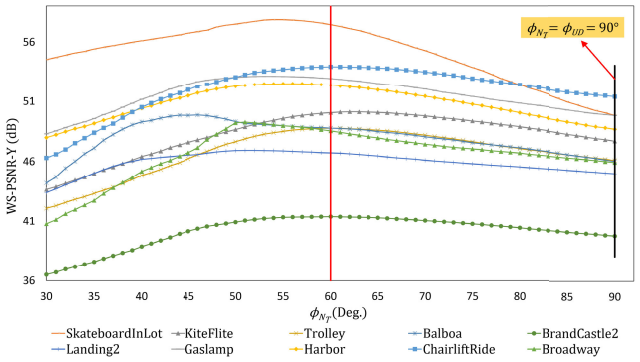

FIGURE 5. WS-PSNR-Y of HSR-RWP method for different latitudes in the first frame of CTC videos.

we computed weighted-to-spherically uniform PSNR (WS-PSNR) [38] of the luma (Y) component for the first frame (as the representative of sequence) of ten 6K/8K 360° videos used in common test conditions (CTC) [39], for all possible values of N, the discrete equivalent of ϕ_{N_T}, in a packing method with the same S_M and S_P used for HSR-RWP ($S_M = 2$, $S_P = 4$). Fig. 5 shows the downsampling distortion represented by the WS-PSNR-Y of the packing method, for various latitudes ϕ_{N_T} of CTC videos. As shown, for most of the videos, the highest WS-PSNR-Y is achieved around $\phi_N = 60°$ ($\phi_N = \phi_{N_T}$). This is demonstrated more clearly in Table 2, where the *average* of optimal latitudes of all tested sequences, $\phi_O = 55.4°$, is almost the same as the average distortion for $\phi_N = 60°$.

B. CONTENT-ADAPTIVE PACKING BASED ON INFORMATION ENERGY LOSS

The PILR is based on the uneven sampling density mapping of the sphere to the 2D plane using ERP. This feature causes the center region to have a higher spatial complexity than the polar areas, consequently making it more susceptible to downsampling distortion (information loss). However, in addition to the uneven sampling density inherent to ERP, the video content itself is another important factor determining the spatial complexity across latitudes. Specifically, when selecting a ϕ_N to minimize PILR in the HSR-RWP method, we implicitly assume that the video content fully utilizes the frequency spectrum provided by the ERP format. Since this assumption does not always hold, the approach may not always be the most efficient. This can be seen in Table 2, where for the *Balboa* and *Broadway* videos, the WS-PSNR of the HSR-RWP method with ϕ_N associated with the minimum PILR ($\phi_N = 60°$) has a notable difference compared to the WS-PSNR of the best latitude ($\phi_N = \phi_O$). Moreover, the region packed without downsampling in the HSR-RWP method is always selected at the center of the ERP frame. In this case, if the center of ERP has a lower spatial complexity compared to the middle regions, then maintaining the original resolution in the center is inefficient. For more efficient packing of video with such spatial complexity distribution and with the same downsampling factors used for HSR-RWP, we introduce a content-adaptive (CA) packing method with the following features:

- *Adaptive center and polar regions sizes*: in contrast to the HSR-RWP method having a fixed $\phi_N = 60°$, we determine ϕ_N, based on the specific video content to minimize downsampling EL. This implies that the size of the polar and the center regions is adaptive.
- *Adaptive latitude range for the center region*: in HSR-RWP, the center region, which is packed without downsampling, is always symmetric with respect to the equator (latitude zero). In the proposed CA packing method, we eliminate this restriction, such that the center region can be positioned lower or higher to pack a latitude range with the highest spatial complexity. In this case, the height of the MT and MB regions will not be equal.

As a result, the proposed CA packing method provides great flexibility in adjusting both the size and position of the central region. We now determine the optimal way to select these parameters by considering the spectral characteristics of each video to process.

1) CA PACKING PARAMETERS OPTIMIZATION USING WS-PSNR

The optimal CA packing is achieved by jointly finding ϕ_N and the offset latitude of the center region, denoted by λ, for which the downsampling distortion of packing is minimum. Note that λ represents the middle of the center region, and therefore is computed relative to the equator [38]. To find the optimal CA packing, we proceed as follows:

i) For each $\phi_N \in \mathcal{S}$, $\mathcal{S} \triangleq [\frac{\pi}{6}, \frac{\pi}{2}]$, move the center region (with the height of N rows) in the range $[\phi_{N_B}, \phi_{N_T}]$. This means the middle of the center region, denoted by λ, can move in the range:

$$\lambda \in \left[\phi_{N_B} + \frac{\frac{\pi}{2} + \phi_{N_B}}{2}, \phi_{N_T} - \frac{\frac{\pi}{2} - \phi_{N_T}}{2}\right]. \quad (32)$$

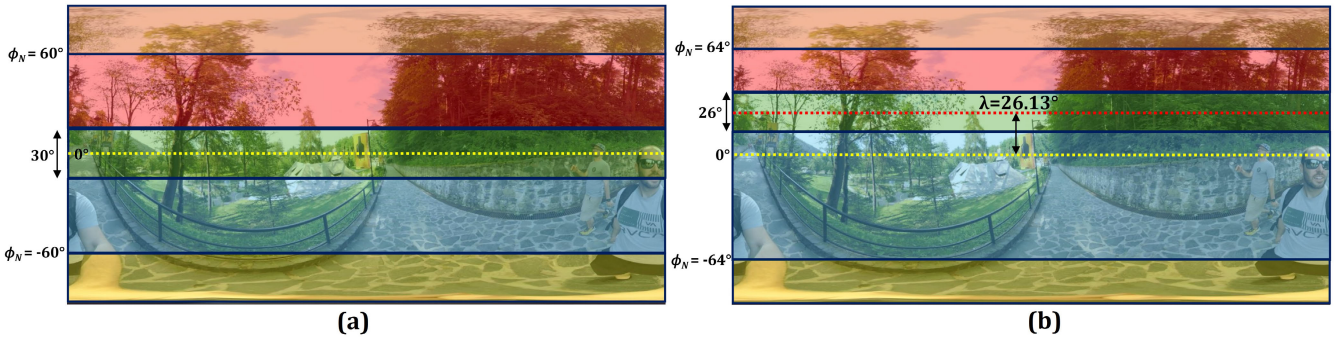

FIGURE 6. Illustrations of (a) the HSR-RWP with $\phi_N = 60°$ ($\phi_N = \phi_{N_T} = -\phi_{N_B}$) and fixed center (λ=0) and (b) CA packing with $\phi_N = 64°$ (the height of the center and polar region is 26°) and $\lambda = 26.13°$ for *BranCastle2* sequences. The center of CA packing moves 26.13° from the middle of the frame. Therefore, the middle-top (MT) and middle-bottom (MB) regions do not have the same height. The height of MT (red region) is smaller than MB (blue region).

Since $\phi_N = \phi_{N_T} = -\phi_{N_B}$ (see Fig. 1), the possible range of λ is only specified by ϕ_N:

$$\lambda \in \mathcal{S}_{\phi_N} \triangleq \left[-\phi_N + \tfrac{\frac{\pi}{2}-\phi_N}{2}, \phi_N - \tfrac{\frac{\pi}{2}-\phi_N}{2}\right]. \quad (33)$$

ii) Then, for each $\phi_N \in \mathcal{S}$, compute WS-PSNR for all possible λ values and find the one that provides maximum WS-PSNR for the frame. We denote it by λ_{ϕ_N}:

$$\lambda_{\phi_N} = \underset{\lambda \in \mathcal{S}_{\phi_N}}{\text{argmax}}\ \text{WS-PSNR}(\phi_N, \lambda),\ \forall \phi_N \in \mathcal{S}. \quad (34)$$

iii) Among all λ_{ϕ_N} values, identify the one with maximum WS-PSNR. That yields optimal λ and ϕ_N values for the CA packing:

$$\phi_N^* = \underset{\phi_N \in \mathcal{S}}{\text{argmax}}\ \text{WS-PSNR}(\phi_N, \lambda_{\phi_N}), \quad (35)$$

$$\lambda^* = \lambda_{\phi_N^*}. \quad (36)$$

Table 3 shows the optimal parameter values and associated WS-PSNR-Y for the proposed CA packing method for the first frame of each CTC video. As can be seen, the CA method with adaptive λ provides an average WS-PSNR gain of 0.9 dB and 0.62 dB compared to the HSR-RWP method with $\lambda = 0$, when $\phi_N = 60°$ and $\phi_N = \phi_O$, respectively (see Table 2). Fig. 6 shows a comparison between the proposed CA packing method and HSR-RWP with $\phi_N = 60°$. As can be seen, the center region of the CA packing is shifted to the top of the frame, and its height decreases to 26° because of choosing $\phi_N = 64°$.

2) LOW-COMPLEXITY DISTORTION ESTIMATION USING DFT

In the previous subsection, to find the optimal CA packing parameter values, the WS-PSNR of a frame for all possible pair values of λ and ϕ_N was computed. As shown in Table 3, this is a computationally intensive process with an average processing time of 811.9 seconds per frame on the computer presented in Section IV. This is because it requires downsampling and reconstructing the frame, and then computing WS-PSNR for each pair of λ and ϕ_N values. Clearly, this approach is not efficient. Instead, we propose a low-complexity method to estimate optimal parameters for the CA packing in the frequency domain, using the DFT.

TABLE 3. Results of optimal CA packing using WS-PSNR-based method for the first frame of each CTC video.

Sequence	ϕ_N^* (Deg.)	λ^* (Deg.)	WS-PSNR-Y (dB)	Time (Sec.)
Balboa	47	10.72	51.31	590.3
BranCastle2	57	16.52	42.05	564.2
Broadway	51	14.71	52.07	530.0
ChairliftRide	58	13.01	54.67	1071.1
Gaslamp	53	4.66	53.21	958.5
Harbor	56	5.27	52.78	951.3
KiteFlite	62	4.57	50.21	951.9
Landing2	53	4.57	47.04	515.1
SkateboardInLot	55	0.79	57.88	1035.3
Trolley	60	-2.46	48.81	951.5
Average	**55.2**	**7.24**	**51.00**	**811.9**

In what follows, we are using the theory and notations of discrete-domain images as presented in [40] and [41] and applied them to the 1D case. According to the Nyquist theorem [42], the minimum sampling frequency, f_s, is twice the highest frequency f_{max} present in the signal. Therefore:

$$f_{max} = \frac{f_s}{2} \quad (37)$$

For a downsampling factor α, the highest preserved frequency in the reconstructed image, after downsampling and upsampling, becomes[1]:

$$f_{max}^{\alpha} = \frac{f_{max}}{\alpha} = \frac{f_s}{2\alpha} \quad (38)$$

We define the total EL caused by downsampling as the sum of the energies of the frequencies lost by the downsampling operation (i.e. frequencies higher than f_{max}^{α}).

We propose using EL caused by horizontal downsampling to determine the optimal λ and ϕ_N values for the CA packing. To do this, we only need to compute the 1D DFT for individual rows. Since the DFT of a digital image is conjugate

[1] We have the same result with classic digital signal processing theory [42]; that downsampling by a factor α results in the maximum normalized frequency of ω_{max} to become ω_{max}/α.

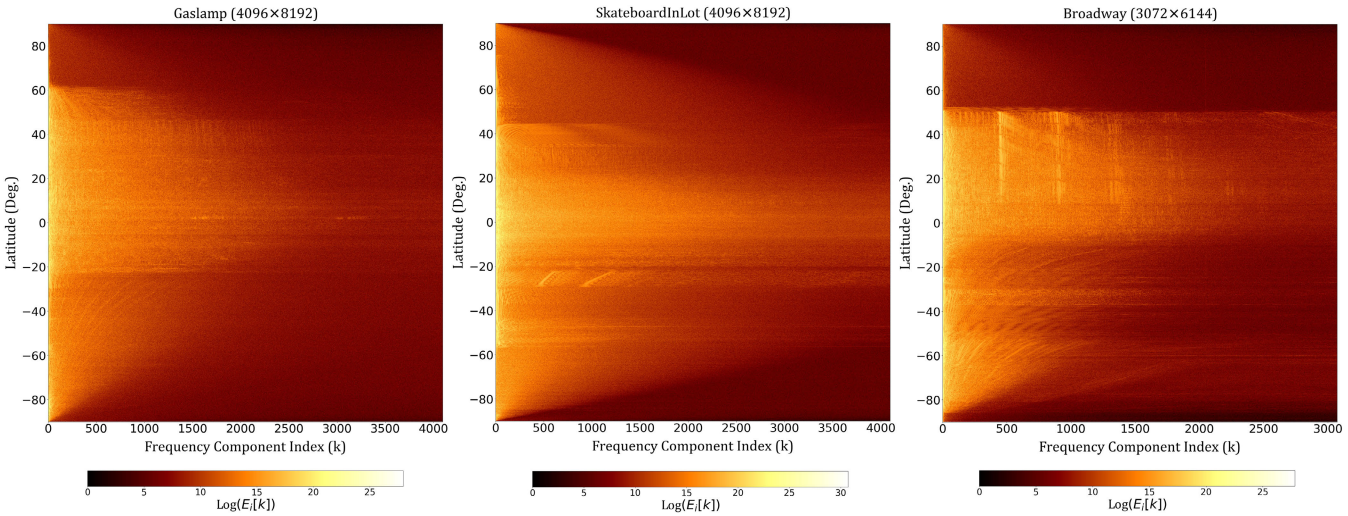

FIGURE 7. Examples of the energy spectrum of frequency components obtained by 1D DFT on each row for the luma (Y) channel of the first frame of videos (for better visualization, the logarithm of energy is used). The energy of high frequencies in polar areas of ERP (darker regions) is lower than the central regions (brighter regions), indicating that these areas have lower spatial complexity compared to the central regions.

symmetric [43], the number of unique (positive) frequency components of DFT, for each row with the width of W, is $\frac{W}{2}$. For unique frequency components, we assume that the larger indices represent higher spatial frequencies, such that:

$$f_k \triangleq \frac{k}{M-1}, \quad 0 \le k \le M-1, \quad \text{with } M = \frac{W}{2} \tag{39}$$

where f_k is the normalized spatial frequency corresponding to the DFT frequency component with index k.

The energy of the k-th frequency component, $X[k]$, in the 1D DFT is:

$$E[k] = X_{re}^2[k] + X_{im}^2[k] \tag{40}$$

where $X_{re}[k]$ and $X_{im}[k]$ are the real and imaginary parts of the k-th DFT frequency component, f_k.

Fig. 7 illustrates examples of the energy spectrum of frequency components obtained by applying 1D DFT to each row. As can be seen, the energy of the DFT frequency components can serve as an indicator of spatial complexity variations across latitudes of ERP. In polar areas, the energy of high-frequency components is lower than in the central regions. This is because polar regions in ERP frame have higher pixel uniformity with lower spatial complexity compared to the central regions. Conversely, there are typically significant pixel intensity fluctuations in the center of ERP, leading to strong energy in high-frequency components.

These observations suggest that the energy of frequency components can potentially be used as a reliable tool to estimate the optimal parameters for the CA packing method.

Let S_i represent the downsampling factor of row i, u_i the index corresponding to the lowest spatial frequency higher than $f_{max}^{S_i}$, and $E_i[k]$, the energy of DFT frequency component k in row i. We have:

$$E_i[k] \leftarrow 0, \ \forall\, k \ge u_i \triangleq \left\lceil \frac{M}{S_i} \right\rceil. \tag{41}$$

In other words, to account for energy loss due to downsampling, we set the energy corresponding to the lost frequencies to zero. The CA packing has the same three horizontal downsampling factors used for HSR-RWP. Therefore, the possible values for S_i are 1, 2, and 4 when row i is in the center, middle, and polar regions, respectively. For row i with the downsampling factor of S_i, we define its EL as:

$$EL_i = \sum_{k=u_i}^{M-1} E_i[k], \quad 0 \le i < H. \tag{42}$$

Note that the value of EL_i for the center region with $S_i = 1$ is zero. The total EL of the frame is then calculated by summing up the EL_i for all H rows of the frame.

$$EL_{frame} = \sum_{i=0}^{H-1} \sum_{k=u_i}^{M-1} E_i[k] \tag{43}$$

Although not explicitly stated to simplify the notation, EL_i and EL_{frame} of the CA packing are clearly functions of λ and ϕ_N. The optimal parameters of CA packing (λ^* and ϕ_N^*) using EL_{frame} are determined through the following steps:

i) Compute the horizontal DFT of the frame. This is done by computing the 1D DFT of each row.

ii) For each $\phi_N \in \mathcal{S}$, compute EL_{frame} for all possible λ values and find the one with minimum EL_{frame}, denoted λ_{ϕ_N}. Formally, we have:

$$\lambda_{\phi_N} = \underset{\lambda \in \mathcal{S}_{\phi_N}}{\operatorname{argmin}} EL_{frame}(\phi_N, \lambda), \ \forall \phi_N \in \mathcal{S} \tag{44}$$

iii) After computing all λ_{ϕ_N} values, select the one with minimum EL_{frame} value:

$$\phi_N^* = \underset{\phi_N \in \mathcal{S}}{\operatorname{argmin}} EL_{frame}(\phi_N, \lambda_{\phi_N}) \tag{45}$$

$$\lambda^* = \lambda_{\phi_N^*} \tag{46}$$

TABLE 4. Results of optimal CA packing using DFT-based method for the first frame of the CTC videos.

Sequence	ϕ_N^* (Deg.)	λ^* (Deg.)	WS-PSNR-Y (dB)	Time (Sec.)
Balboa	53	7.62	51.19	0.43
BranCastle2	64	26.13	41.76	0.44
Broadway	51	14.77	52.07	0.43
ChairliftRide	60	13.97	54.63	0.65
Gaslamp	58	4.31	53.13	0.64
Harbor	58	6.24	52.73	0.64
KiteFlite	66	3.60	50.10	0.66
Landing2	58	-13.95	46.88	0.42
SkateboardInLot	57	0.79	57.82	0.65
Trolley	64	1.67	48.65	0.64
Average	**58.9**	**6.52**	**50.90**	**0.56**

Table 4 summarizes the results of using the DFT-based method for computing CA packing optimal parameters (ϕ_N^* and λ^*).

As shown, the average WS-PSNR corresponding to optimal parameters of DFT-based method is only 0.1 dB lower than that of the WS-PSNR-based method (see Table 3). However, the optimal parameter values may vary significantly between the two methods. This discrepancy arises because the WS-PSNR may not be highly sensitive to parameter variation in certain regions. For example, as illustrated in Fig. 5, the WS-PSNR of some videos remains nearly identical when ϕ_{N_T} changes within a specific range.

Moreover, the average processing time of DFT-based method is only 0.56 seconds, significantly lower than that of the WS-PSNR-based method (811.9 seconds, approximately 1450× faster). The reason is that computing the 1D DFT of rows only needs to be done once. Then, according to Eq. (42) and Eq. (43), the EL of rows and EL_{frame} for all possible cases (different values of λ and ϕ_N) are computed through simple summation of DFT component values. Therefore, the computational complexity of this method is considerably lower than that of the WS-PSNR-based method.

Fig. 8 depicts examples of optimal λ using DFT-based method for $\phi_N = 60°$ ($\lambda_{60°}$). As shown, inside the range of latitudes with the downsampling factor of $S_i = 2$, the region with the highest EL (between green lines) is selected to be packed without downsampling to minimize overall EL of the CA packing.

The values of λ^* and ϕ_N^* must be sent to the decoder for the unpacking process. Supplemental Enhancement Information (SEI) can contain information for video, picture (frame), and slice levels [44]. Therefore, CA packing parameters can be sent for each frame to the decoder via SEI without changing the syntax of the video coding standards (e.g., VVC or HEVC). For Intra-only configuration, CA packing parameters can be independently calculated and transmitted for each frame. For Random Access (RA) configuration with the closed group of pictures (GOP) option, they can be calculated for the first frame of each GOP and used for the whole GOP without any negative effect on the performance of inter-prediction. But when using RA configuration with the open GOP option, or Low Delay (LD) configuration, the frames of each GOP are used as reference frames for the following GOP. Consequently, changing the CA packing parameters, even if done at the boundaries of GOPs, may negatively impact the performance of inter-prediction. One possible solution is to update the CA packing parameters only when a significant change in the content occurs (e.g., a change in the video scene), ensuring that updating the CA packing parameters significantly reduces the downsampling distortion, thereby improving coding performance to a greater extent than the negative impact of CA packing on inter-prediction performance. It can be achieved, for example, by detecting changes in video scenes or consecutive frames. Proposing an efficient algorithm to reduce the interval of updating CA packing parameters is beyond the scope of this paper and is left for future work.

In this study, we used the DFT-based method to optimize the CA packing with the same downsampling factors as those used for HSR-RWP. However, this method is versatile and can be utilized for any downsampling factor in both horizontal and vertical directions.

C. SEAM ARTIFACTS ALLEVIATION

There are four vertical discontinuities in the proposed HSR-RWP method that cause seam artifacts in the reconstructed video, specifically when the video is encoded with a high QP (low bit-rate). In our previous work [21], these four discontinuities are alleviated by pixel padding. To reduce the required pixel padding for seam artifacts alleviation, we applied the following modifications to the HSR-RWP method proposed in our previous work:

- The center of the right view is vertically flipped, such that the lower border of the center regions of both views, which have similar pixel values, are packed using a common border (white line in Fig. 9). This significantly diminishes seam artifacts due to discontinuity border after encoding, making padding unnecessary for this border. Therefore, only three vertical discontinuity borders (highlighted by red lines in Fig. 9) need to be alleviated by padding.
- In addition, the upper borders of the bottom regions (Left(B) and Right(B) parts in Fig. 9) are closer to the center of ERP and as a result, have higher importance than the lower borders in terms of human attention [22], [23]. To avoid seam artifacts at the upper border of bottom regions, the Left(B) and Right(B) parts in Fig. 9 are vertically flipped. This aligns the upper borders of these regions with the frame's border, where pixels remain unaffected by the seam artifacts caused by discontinuity.

Moreover, as mentioned in [21], since the corresponding columns with similar pixel values in the left and right views are packed using a common border, there is no significant seam artifact due to horizontal discontinuity, except for the common border between the Right(B) and Left(T) parts in Fig. 9. Compared to the height of the frame, the size of

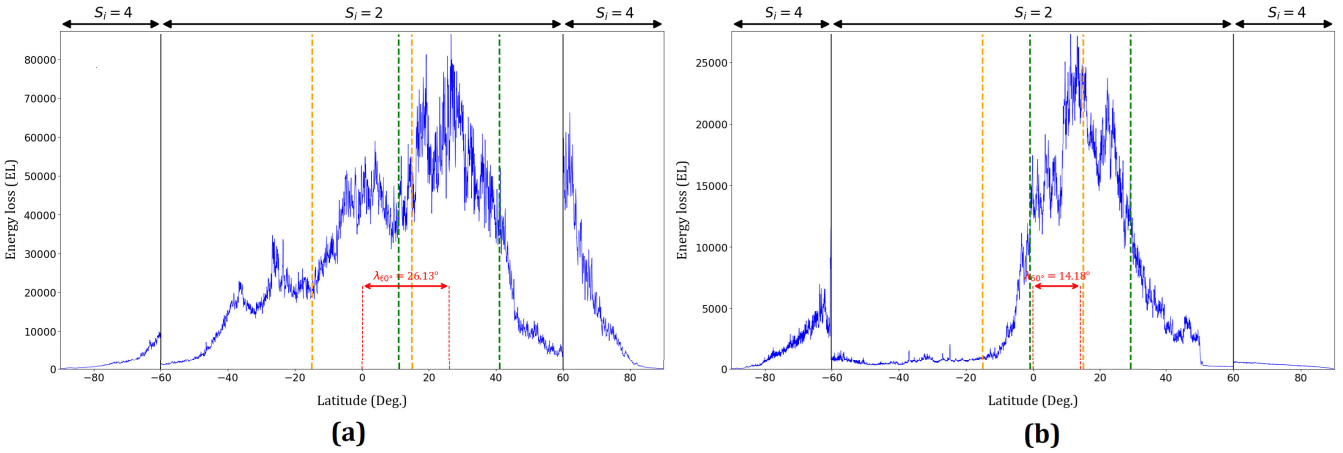

FIGURE 8. Examples of EL for different λ values for the luma (Y) channel of the first frames of (a) *BranCastle2* and (b) *Broadway* videos when $\phi_N = 60°$. The EL of each row (latitude) is normalized by dividing it by the number of unique frequency components (M). Polar latitudes ($\phi_N > |60°|$) are downsampled with the factor of $S_i = 4$, while other latitudes represent the EL for downsampling with the factor $S_i = 2$. The center region with a height of 30° (for $\phi_N = 60°$) is moved inside the range of latitudes with downsampling factor of $S_i = 2$ ($\phi_N \leq |60°|$) to find the region with the highest EL caused by the downsampling factor of $S_i = 2$ (the region between green lines with offsets of 26.13° and 14.18° relative to the latitude of zero for (a) and (b), respectively). Then, this region is packed without downsampling to minimize overall downsampling EL in the CA packing. For HSR-RWP with the fixed center packing, the region is selected in the middle of the frame (between orange lines) with $\lambda = 0$.

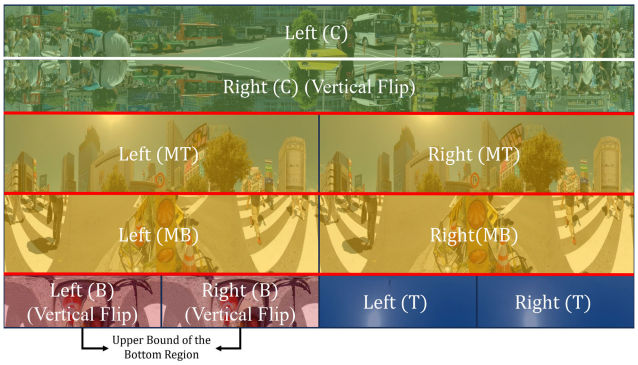

FIGURE 9. Illustration of the the modified HSR-RWP method for the stereoscopic ERP video. To reduce seam artifacts, the center of the right view and the bottom regions (highlighted by the white color) are vertically flipped. The discontinuity borders highlighted by red lines need to be alleviated by padding when a video is encoded by a high QP. After vertical flipping of the bottom regions, the upper bound of the bottom regions is aligned with the frame's border.

this discontinuity is small; however, adding pixel padding requires additional columns across all rows of the frame. This results in a significant number of unnecessary pixels being added to the frame. Moreover, this discontinuity occurs between the polar regions, which are of lower importance in terms of human attention compared to other regions of ERP. Consequently, no padding is applied to this horizontal discontinuity.

Since the arrangement of partitions in the CA packing is the same as that of HSR-RWP, all the aforementioned modifications can be applied to the CA packing.

IV. EXPERIMENTAL RESULTS

To evaluate the performance of our proposed packing methods, we tested them with eight 8K stereoscopic 360° videos (7680 × 3840, I420, 8-bit depth) available from [45]. This dataset includes video content with different levels of motion and texture complexities. The sequences used in simulations are listed in Table 5. For all simulations, we used VVenC 1.9.1 [46] with medium presets of RA and Low Delay B (LDB) configurations and four different QP values (22, 27, 32, and 37) for encoding the first 240 frames of each sequence. We used the OpenCV library [47] with Lanczos4 interpolation for all downsampling and upsampling operations. For performance comparison, we utilized the Bjøntegaard-Delta bitrate (BD-BR) measurement method [48] with end-to-end (E2E) WS-PSNR [39]. We used the average WS-PSNR of the left and right views to compute BD-BR, as they have highly similar content with only a slight horizontal disparity between them. Moreover, no view is preferred to another in the proposed packing methods, such that the downsamplings of the left and right views are identical. We ran our simulations on a computer equipped with an Intel® Core™ i9-10900 CPU@2.80GHz, 64GB of RAM, and 64-bit Windows 11.

The DFT-based CA parameters for the videos used in simulations are listed in Table 6. For simulations, we used the DFT-based CA parameters of the luma component for

TABLE 5. The sequences used in simulations.

#	Sequence
1	K2_Mountain_Peak
2	Key_West_Public_Square
3	Sample5_FlowState_Running_and_Aerial
4	Sample7_Shanghai
5	Tokyo_Day_and_Night_1
6	Tokyo_Day_and_Night_2
7	Tokyo_Day_and_Night_3
8	Tokyo_Day_and_Night_4

TABLE 6. The optimal parameters of different components for CA packing computed for the first frame of the left view of simulated videos using DFT-based method.

Sequence	Y		U		V	
	ϕ_N^* (Deg.)	λ^* (Deg.)	ϕ_N^* (Deg.)	λ^* (Deg.)	ϕ_N^* (Deg.)	λ^* (Deg.)
1	66	-18.84	82	-55.88	82	-9.00
2	67	1.31	62	10.31	62	10.31
3	61	3.09	55	-15.75	65	12.94
4	63	-16.41	80	-40.50	84	-36.19
5	62	8.72	57	7.69	54	11.62
6	59	2.25	47	4.69	47	4.69
7	57	12.94	54	11.62	54	11.62
8	66	-7.97	57	-11.06	57	-12.94

both the luma and chroma components. Moreover, since the content complexity of the first 240 frames of videos utilized for simulation does not significantly vary, we computed the CA parameters of the luma component only for the first frame of the left view and used these parameters for packing all video frames.

We used the SbS format with UHHDS as the anchor. We compared our proposed packing methods (HSR-RWP and CA packing) with Preserved Aspect Ratio (PAR) format for which the height and width of ERP frame are downsampled with the same factors. For generating the PAR format with SR=0.5, the width and height of ERP video are downsampled by the factor of 1.42 ($S_W = S_H = 1.42$), which results in a downsampled ERP with SR=0.49. Using SR=0.49 for PAR format is allowed because, based on the CTC [39], the variation of coded samples between the anchor and tested methods can be within ±3%. We also implemented the Nested Polygonal Chain Mapping (NPCM)-Full method (with SR=0.5), as the best method from the literature proposed in [30], to compare it with our proposed methods.

We used $\phi_N = 60°$ to generate sequences with our HSR-RWP method. As shown in Subsection III-A2, this latitude, on average, provides near-minimum downsampling distortion. Moreover, for the tested videos with the size of 7680 × 3840, using $\phi_N = 60°$ as the partitioning latitude can reduce the number of misalignments between discontinuity borders and CTU borders in the HSR-RWP packing method.

In the HSR-RWP and CA packing methods, to reduce the impact of vertical discontinuities, the partitions (tiles) are extended (overlapped) [49] by four rows in the direction of the borders producing the discontinuities (eight rows of pixels in total between partitions). The primary layout proposed in our previous work for the HSR-RWP method generates four rows with vertical discontinuities that need to be padded. Hence, 32 rows of padding are added to the height of the packing method for coding. By flipping partitions (as explained in Subsection III-C), the rows with vertical discontinuities in the HSR-RWP method that require padding, are reduced to three, decreasing the total rows of padding to 24. To evaluate the effectiveness of the flipping partitions in the HSR-RWP method on coding performance, we simulated both layouts of the HSR-RWP method with and without flipping. For the CA packing, the partitions are flipped in the same way as the HSR-RWP method.

Table 7 shows the coding performance of different packing methods for RA configuration. HSR-RWP-NF and HSR-RWP-F indicate the HSR-RWP proposed in our previous work [21] without flipping and the modified HSR-RWP method with flipping, respectively. As can be seen, the overall performance of the HSR-RWP-F for all three components, outperforms the HSR-RWP-NF. For the luma (Y) component, PAR format, on average, provides a better coding performance compared to the HSR-RWP-F method. However, this is mainly because the BD-BR-Y of sequences 1 and 4 for the PAR method are significantly better compared to those of HSR-RWP-F. As shown in Table 6, the λ^* values for these videos differ significantly from the $\lambda = 0$, indicating that the HSR-RWP-F packing (with $\lambda = 0$) fails to retain the region with the highest energy loss due to downsampling (factor of 2) in the central region at its original resolution. This means using the CA packing method can potentially improve the coding performance of luma components for these videos. This can be seen in Table 7, where the CA method significantly improves the coding performance of the luma component for videos 1 and 4, providing better overall BD-BR compared to the PAR format. In contrast, if spatial complexity is almost the same over various latitudes in a video, using CA packing cannot significantly enhance the coding performance. In such a case, which is observed in videos 5, 6, and 8, adjusting ϕ_N and λ based on the spatial complexity of the video may increase the misalignment between CTU and discontinuity borders, reducing the coding performance compared to the HSR-RWP-F method. This issue may easily be addressed using a threshold, such that the region with the original resolution only is moved from the center of the frame (with $\lambda = 0$) if the distortion reduction compared to the case with $\lambda = 0$ exceeds a specific threshold. Future work will investigate this.

In contrast to the luma component, both HSR-RWP-F and CA packing methods, on average, provide a better BD-BR compared to PAR for chroma components. However, as can be seen in Table 7, for these videos, BD-BR of chroma components are considerably high, and even CA packing does not improve the coding performance compared to HSR-RWP-F. This is because, for videos 1 and 4, the latitude range with the highest spatial complexity for chroma components differs from that of the luma component. Since in the CA method we used λ and ϕ_N values of the luma for chroma components, it may not provide better performance than HSR-RWP-F for chroma components. This is evident in Table 6 for CA packing, where for videos 1 and 4, λ^* and ϕ_N^* values for luma components are notably different from those for chroma components. This situation also exists for these videos in HSR-RWP-NF and HSR-RWP-F methods (with $\lambda = 0$ and $\phi_N = 60°$).

TABLE 7. Coding performance of different packing methods for RA configuration using VVC (VVENC 1.9.1).

Seq.	PAR			HSR-RWP-NF [21]			HSR-RWP-F			CA Packing			NPCM-Full [30]		
	BD-BR(%)			BD-BR(%)			BD-BR(%)			BD-BR(%)			BD-BR(%)		
	Y	U	V	Y	U	V	Y	U	V	Y	U	V	Y	U	V
1	-13.24	4.52	5.16	-5.37	10.33	7.82	-6.12	9.53	7.11	-15.39	8.34	8.72	-19.91	8.49	8.90
2	-15.51	-0.61	-1.00	-14.12	-1.87	-5.84	-14.32	-2.05	-6.11	-13.91	-1.11	-5.45	-21.39	-2.02	-3.16
3	-5.76	4.31	5.67	-8.87	-4.23	5.04	-9.97	-5.68	3.86	-10.75	-6.77	-2.28	-17.29	-10.90	-1.57
4	-17.94	1.73	1.59	-10.77	4.53	6.12	-11.44	4.03	5.05	-18.28	12.59	13.87	-24.99	6.92	7.86
5	-7.00	0.39	0.01	-6.19	2.03	0.67	-6.95	0.96	-0.69	-6.12	2.70	2.01	-10.75	-0.08	-2.19
6	-10.41	1.27	0.92	-8.82	0.13	-1.85	-9.51	-1.11	-2.72	-9.15	0.15	-1.24	-16.79	-3.18	-4.26
7	-24.10	-1.49	-1.17	-23.99	-0.53	-1.13	-24.34	-0.98	-1.32	-26.78	-2.24	-2.88	-30.88	-2.72	-3.11
8	-8.56	1.76	1.37	-10.59	-9.68	-9.39	-10.99	-9.80	-9.79	-10.32	-10.26	-10.07	-18.09	-10.52	-10.20
Avg	**-12.81**	**1.49**	**1.57**	**-11.09**	**0.09**	**0.18**	**-11.71**	**-0.64**	**-0.57**	**-13.84**	**0.42**	**0.34**	**-20.01**	**-1.75**	**-0.97**

TABLE 8. Packing (P) and unpacking (U) times of different methods and their ratios related to the UHHDS (anchor) for 240 frames. For the CA packing method, it is supposed that the CA parameters (λ and ϕ_N) are updated with an interval of 32 frames, sharing $\frac{1}{32}$ of the total parameters' computation time to each frame, on average. The total time is calculated by multiplying the frame time by the total number of frames (240 frames). The average packing and unpacking times are independent of encoding/decoding configurations; therefore, they are identical for all QPs and coding configurations (RA and LDB).

Method	Frame Time		Total Time (240 Frames)		Ratio	
	P (Sec.)	U (Sec.)	P (Sec.)	U (Sec.)	P	U
UHHDS	0.050	0.077	11.93	18.55	1.00	1.00
PAR	0.056	0.069	13.37	16.59	1.12	0.89
HSR-RWP-F	0.078	0.144	18.63	34.65	1.56	1.87
CA	0.096	0.138	22.97	33.09	1.92	1.78
NPCM-Full [30]	1.323	2.943	317.40	706.37	26.60	38.07

TABLE 9. Average total encoding and decoding times of tested videos (240 frames) encoded by RA configuration, for different packing methods and QPs.

Method	Enc Time (Sec.)				Dec Time (Sec.)			
	QP22	QP27	QP32	QP37	QP22	QP27	QP32	QP37
UHHDS	9002.2	5881.9	4168.4	3117.6	91.7	66.0	54.8	47.8
PAR	8692.4	5672.5	4046.3	3032.9	88.9	64.6	53.8	46.6
HSR-RWP-F	8342.3	5521.4	3995.3	3046.1	93.5	66.6	55.7	48.6
CA	8455.1	5701.0	4121.3	3136.6	96.8	68.9	57.4	50.3
NPCM-Full	8742.0	5841.4	4218.3	3197.5	97.1	69.0	58.1	50.5

The NPCM-Full method provides better performance compared to our proposed methods (approximately 6% lower BD-BR for the luma channel compared to CA packing). Note that both methods, the proposed approach and the NPCM, are based on the same principle of downsampling and repositioning of pixels. The NPCM method, however, requires performing the downsampling and repositioning operations with a different sampling factor for each row of the polar regions. It is, therefore, significantly more complex than our proposed method, which uses only three distinct downsampling factors for the entire frame. Indeed, as can be seen in Table 8, the average packing and unpacking times of NPCM method are 26.6× and 38× higher than the UHHDS methods, respectively, while these values for our proposed methods are comparable with UHHDS. Moreover, as shown in Table 9, the total unpacking (reconstruction) time is significantly higher than decoding time. Therefore, despite its better coding performance, NPCM is not a desirable solution for packing 360° video in real-time applications.

As shown in Table 8, the packing/unpacking time of the PAR method is approximately 2× faster than the proposed methods, with only 1% higher BD-BR. However, the main approach of our proposed methods (HSR-RWP and CA) is keeping the quality of the center region of 360° because we want to exploit the *equator bias* in terms of user attention. Indeed, in [22], the authors found that human subjects tend to fixate their eyes around the equator of 360° videos. They used a Laplacian Distribution to model the probability of eye-fixation across the latitudes. They observed that the latitude range $\phi \approx \pm 18°$ had the highest fixation rate, which approximately corresponds to a Laplacian Distribution model with $\mu = 0$ and $\beta = 0.2$. We used this Laplacian model to weight the WS-PSNR value across the pixel rows (latitudes) of ERP videos. By doing so, for computing the coding performance, we take both the user attention factor and the uneven pixel density characteristic of ERP (by *cosine weights* of WS-PSNR metric) into account.

Table 10 shows that based on the Laplacian model, the coding performance of our proposed methods (HSR-RWP-F and CA) significantly outperforms PAR. In addition, our methods also provide better performance compared to NPCM even though they are considerably faster in packing and unpacking operations. Moreover, the result shows that the performance of HSR-RWP-F is better than CA. The reason is that in contrast to the CA method, the region with original resolution in HSR-RWP-F is always entirely within the latitude range with the highest user attention (the highest Laplacian model weight). Nevertheless, the CA method can still be advantageous for packing the individual video content,

TABLE 10. Coding performance of different packing of methods measured by the Laplacian Distribution model for RA configuration using VVC (VVENC 1.9.1).

Seq.	PAR			HSR-RWP-F			CA Packing			NPCM-Full [30]		
	BD-BR(%)			BD-BR(%)			BD-BR(%)			BD-BR(%)		
	Y	U	V	Y	U	V	Y	U	V	Y	U	V
1	-17.28	3.68	4.35	-25.52	-10.72	-12.18	-23.59	-3.34	-1.93	-29.53	-4.51	-2.90
2	-28.73	-1.93	-4.07	-44.69	-18.72	-24.49	-40.49	-15.48	-21.18	-39.30	-7.49	-10.78
3	-11.50	3.29	4.88	-33.01	-22.26	-16.40	-32.31	-22.07	-21.12	-29.24	-17.97	-13.65
4	-29.12	1.18	1.70	-40.53	-17.25	-17.15	-39.70	-3.68	-2.99	-41.61	-7.00	-7.81
5	-8.49	-0.40	-0.82	-21.95	-13.31	-15.23	-17.75	-7.59	-9.05	-14.45	-4.01	-6.03
6	-12.76	2.24	1.65	-28.74	-16.11	-18.21	-28.49	-14.81	-16.69	-23.52	-7.97	-9.86
7	-24.79	-1.50	-1.08	-39.85	-15.10	-15.11	-37.02	-13.02	-13.28	-34.45	-6.14	-6.54
8	-12.92	1.59	0.68	-33.70	-24.24	-24.26	-27.67	-21.41	-21.21	-28.49	-14.21	-14.96
Avg	**-18.20**	**1.02**	**0.91**	**-33.50**	**-17.21**	**-17.88**	**-30.88**	**-12.68**	**-13.43**	**-30.07**	**-8.66**	**-9.06**

TABLE 11. Coding performance of different packing methods for LDB configuration using VVC (VVENC 1.9.1).

Seq.	PAR			HSR-RWP-NF [21]			HSR-RWP-F			CA Packing			NPCM-Full [30]		
	BD-BR(%)			BD-BR(%)			BD-BR(%)			BD-BR(%)			BD-BR(%)		
	Y	U	V	Y	U	V	Y	U	V	Y	U	V	Y	U	V
1	-9.90	1.27	1.21	-5.51	18.24	12.25	-7.62	15.46	8.64	-17.82	9.86	11.02	-26.48	-0.36	-1.93
2	-12.41	2.11	0.11	-13.20	-5.27	-9.55	-13.29	-4.95	-9.68	-13.33	-1.82	-8.62	-18.43	-3.63	-5.63
3	-6.67	3.10	4.66	-7.44	-3.28	6.13	-8.48	-5.33	4.12	-9.50	-7.42	-1.49	-15.13	-10.51	-1.41
4	-14.59	3.09	3.79	-12.38	4.85	7.77	-13.39	3.52	7.45	-18.56	6.78	11.58	-21.81	4.96	9.87
5	-6.29	0.25	0.62	1.79	1.02	0.73	0.61	-0.67	-0.86	-1.46	-0.35	0.43	-8.78	-1.83	-1.69
6	-10.41	-0.72	-2.15	-8.13	-4.99	-6.46	-9.01	-6.75	-7.85	-9.17	-5.80	-6.65	-16.31	-7.88	-9.35
7	-21.94	-2.26	-1.99	-12.75	2.82	2.22	-13.74	2.07	1.52	-18.73	-3.69	-4.95	-26.80	-6.81	-7.26
8	-7.02	-0.15	-0.04	-8.00	-9.30	-9.66	-8.74	-9.97	-10.66	-7.56	-9.50	-9.48	-15.01	-12.61	-12.90
Avg	**-11.15**	**0.84**	**0.78**	**-8.20**	**0.51**	**0.43**	**-9.21**	**-0.83**	**-0.91**	**-12.02**	**-1.49**	**-1.02**	**-18.59**	**-4.83**	**-3.79**

TABLE 12. Coding performance of different packing methods measured by the Laplacian Distribution model for LDB configuration using VVC (VVENC 1.9.1).

Seq.	PAR			HSR-RWP-F			CA Packing			NPCM-Full [30]		
	BD-BR(%)			BD-BR(%)			BD-BR(%)			BD-BR(%)		
	Y	U	V	Y	U	V	Y	U	V	Y	U	V
1	-12.52	-0.49	-0.28	-30.91	-20.28	-24.39	-26.79	-10.21	-7.82	-35.39	-22.21	-22.46
2	-23.93	0.32	-2.87	-45.01	-27.58	-32.60	-40.66	-22.89	-28.88	-35.51	-11.37	-15.36
3	-10.93	2.94	5.09	-30.83	-22.56	-15.73	-30.51	-23.11	-20.27	-25.86	-15.59	-12.67
4	-20.66	4.60	5.35	-38.09	-22.82	-23.45	-35.55	-10.48	-8.55	-32.25	-9.38	-8.41
5	-7.85	-0.21	0.03	-18.39	-16.90	-18.02	-16.34	-12.52	-13.13	-14.25	-6.36	-6.83
6	-12.07	0.23	-1.27	-27.42	-21.88	-24.17	-27.63	-20.86	-23.13	-22.11	-12.48	-14.87
7	-23.70	-1.64	-1.06	-33.90	-17.22	-17.18	-32.93	-17.97	-17.91	-32.54	-11.73	-11.65
8	-9.86	0.29	-0.31	-30.60	-26.87	-27.72	-23.92	-23.14	-22.67	-24.11	-16.43	-17.60
Avg	**-15.19**	**0.75**	**0.58**	**-31.89**	**-22.01**	**-22.91**	**-29.29**	**-17.65**	**-17.79**	**-27.75**	**-13.19**	**-13.73**

as the distribution of eye fixations may not necessarily be equator-biased for different types of video content [50].

Table 11 reports the simulation results of LDB configuration for tested packing methods. Similar to RA configuration:

- For all three components the overall performance of the HSR-RWP-F method is better than HSR-RWP-NF.
- The PAR format, on average, shows a better performance for the luma components.
- The CA packing can improve the performance of luma components for videos 1, 4, and 7, consequently providing better overall performance.
- Regarding chroma components, our proposed methods (HSR-RWP-NF, HSR-RWP-F, and CA packing) show better performance compared to PAR.

Moreover, as can be seen in Table 12, for the LDB configuration, similar to the RA configuration, the coding performance of our proposed methods (HSR-RWP-F and CA) is better than that of the PAR and NPCM methods, and the performance of HSR-RWP-F is slightly better than that of the CA method.

V. CONCLUSION

This paper presented two region-wise packing methods for ERP videos with a sampling ratio (SR) of 0.5, leveraging

computationally efficient downsampling operations. The first method, HSR-RWP, minimizes pixel information loss by exploiting the uneven sampling density of the ERP format. The second method, CA packing, optimizes packing by prioritizing regions of high spatial complexity in the frequency domain using DFT. Both methods are competitive in terms of coding performance while maintaining low computational complexity, making them suitable for real-time 360° video applications. Future work focuses on exploring packing methods with SR>0.5 and enhancing the coding efficiency of CA packing method.

REFERENCES

[1] *High Efficiency Video Coding*, ISO/IEC Standard 23008-2, 2021.
[2] G. Tech, Y. Chen, K. Müller, J.-R. Ohm, A. Vetro, and Y.-K. Wang, "Overview of the multiview and 3D extensions of high efficiency video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 1, pp. 35–49, Jan. 2016.
[3] *Versatile Video Coding*, ISO/IEC Standard 23090-3, 2022.
[4] B. Bross, Y.-K. Wang, Y. Ye, S. Liu, J. Chen, G. J. Sullivan, and J.-R. Ohm, "Overview of the versatile video coding (VVC) standard and its applications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 10, pp. 3736–3764, Oct. 2021.
[5] J. M. Boyce, R. Doré, A. Dziembowski, J. Fleureau, J. Jung, B. Kroon, B. Salahieh, V. K. M. Vadakital, and L. Yu, "MPEG immersive video coding standard," *Proc. IEEE*, vol. 109, no. 9, pp. 1521–1536, Sep. 2021.
[6] J. Samelak, J. Stankowski, and M. Domanski, "Efficient frame-compatible stereoscopic video coding using HEVC screen content coding," in *Proc. Int. Conf. Syst., Signals Image Process. (IWSSIP)*, May 2017, pp. 1–5.
[7] A. Kondoz and T. Dagiuklas, *Novel 3D Media Technologies*. Cham, Switzerland: Springer, 2015.
[8] J. Bi, L. Wang, and G. Guo, "8K ultra HD TV broadcast system: Challenge, architecture and implementation," *Digit. Commun. Netw.*, vol. 11, no. 1, pp. 172–181, Feb. 2025.
[9] D. Zhou, S. Wang, H. Sun, J. Zhou, J. Zhu, Y. Zhao, J. Zhou, S. Zhang, S. Kimura, T. Yoshimura, and S. Goto, "An 8K H.265/HEVC video decoder chip with a new system pipeline design," *IEEE J. Solid-State Circuits*, vol. 52, no. 1, pp. 113–126, Jan. 2017.
[10] X. Corbillon, G. Simon, A. Devlic, and J. Chakareski, "Viewport-adaptive navigable 360-degree video delivery," 2016, *arXiv:1609.08042*.
[11] J. Wang, R. Shi, W. Zheng, W. Xie, D. Kao, and H.-N. Liang, "Effect of frame rate on user experience, performance, and simulator sickness in virtual reality," *IEEE Trans. Vis. Comput. Graphics*, vol. 29, no. 5, pp. 2478–2488, May 2023.
[12] N. Al-Hiyari and S. Jusoh, "The current trends of virtual reality applications in medical education," in *Proc. 12th Int. Conf. Electron., Comput. Artif. Intell. (ECAI)*, Jun. 2020, pp. 1–6.
[13] A. Gruenewald, R. Schmidt, L. Sayn, C. Gießer, T. J. Eiler, V. Schmuecker, V. Braun, and R. Brueck, "Virtual reality training application to prepare medical Student's for their first operating room experience," in *Proc. IEEE Int. Conf. Artif. Intell. Virtual Reality (AIVR)*, Nov. 2021, pp. 201–204.
[14] M. Shippee and J. Lubinsky, "Training and learning in virtual reality: Designing for consistent, replicable, and scalable solutions," in *Proc. Int. Conf. Electr., Comput. Energy Technol. (ICECET)*, Dec. 2021, pp. 1–7.
[15] P. Rodrigues, H. Coelho, M. Melo, and M. Bessa, "Virtual reality for training: A computer assembly application," in *Proc. Int. Conf. Graph. Interact. (ICGI)*, Nov. 2022, pp. 1–6.
[16] J. Song, K. Yang, and J. Yang, "The application of virtual reality in games," in *Proc. IEEE 2nd Int. Conf. Data Sci. Comput. Appl. (ICDSCA)*, Oct. 2022, pp. 1086–1090.
[17] F. Li, "VR interactive game design based on unity3D engine," in *Proc. Int. Conf. Robots Intell. Syst. (ICRIS)*, Nov. 2020, pp. 142–145.
[18] Y. Ye, J. M. Boyce, and P. Hanhart, "Omnidirectional 360° video coding technology in responses to the joint call for proposals on video compression with capability beyond HEVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 5, pp. 1241–1252, May 2020.
[19] M. M. Hannuksela and Y.-K. Wang, "An overview of omnidirectional MediA format (OMAF)," *Proc. IEEE*, vol. 109, no. 9, pp. 1590–1606, Sep. 2021.
[20] M. Yu, H. Lakshman, and B. Girod, "A framework to evaluate omnidirectional video coding schemes," in *Proc. IEEE Int. Symp. Mixed Augmented Reality*, Sep. 2015, pp. 31–36.
[21] H. Pejman, S. Coulombe, C. Vazquez, M. Jamali, and A. Vakili, "A novel region-dependent packing method for stereoscopic 360° videos using horizontal downsampling of equirectangular projection," in *Proc. Picture Coding Symp. (PCS)*, Jun. 2024, pp. 1–5.
[22] V. Sitzmann, A. Serrano, A. Pavel, M. Agrawala, D. Gutierrez, B. Masia, and G. Wetzstein, "Saliency in VR: How do people explore virtual environments?" *IEEE Trans. Vis. Comput. Graphics*, vol. 24, no. 4, pp. 1633–1642, Apr. 2018.
[23] M. Xu, C. Li, S. Zhang, and P. L. Callet, "State-of-the-art in 360° video/image processing: Perception, assessment and compression," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 1, pp. 5–26, Jan. 2020.
[24] *Coded Representation of Immersive Media—Part 2: Omnidirectional Media Format*, ISO/IEC, Geneva, Switzerland, 2023.
[25] K. Jafari, A. Aminlou, and M. M. Hannuksela, "Comparison of boundary artifact removal methods in coding of generalized cubemap projection using VVC," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2022, pp. 1625–1629.
[26] P. Blanchfield and D. Wang, "Improved tile format of stereoscopic video for 3-D TV broadcasting," *IEEE Trans. Broadcast.*, vol. 60, no. 1, pp. 134–140, Mar. 2014.
[27] T. Lu, H. Ganapathy, G. Lakshminarayanan, T. Chen, W. Husak, and P. Yin, "Orthogonal muxing frame compatible full resolution technology for multi-resolution frame-compatible stereo coding," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2013, pp. 1–6.
[28] P. Van Duc, P. T. Tin, A. V. Le, N. H. K. Nhan, and M. R. Elara, "Inter-frame based interpolation for top–bottom packed frame of 3D video," *Symmetry*, vol. 13, no. 4, p. 702, Apr. 2021.
[29] M. Yu, H. Lakshman, and B. Girod, "Content adaptive representations of omnidirectional videos for cinematic virtual reality," in *Proc. 3rd Int. Workshop Immersive Media Experiences*, New York, NY, USA, Oct. 2015, pp. 1–6.
[30] K. Kammachi-Sreedhar and M. M. Hannuksela, "Nested polygonal chain mapping of omnidirectional video," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 2169–2173.
[31] J. Li, Z. Wen, S. Li, Y. Zhao, B. Guo, and J. Wen, "Novel tile segmentation scheme for omnidirectional video," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 370–374.
[32] A. Zare, M. Homayouni, A. Aminlou, M. M. Hannuksela, and M. Gabbouj, "6K and 8K effective resolution with 4K HEVC decoding capability for 360 video streaming," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 15, no. 2s, pp. 1–22, Jul. 2019.
[33] F. Racapé, F. Galpin, G. Rath, and E. François, *AHG8: Adaptive QP for 360 Video Coding*, document JVET-F0038, 2017.
[34] P. Bordes, Y. Chen, C. Chevance, E. FranÃ§ois, F. Galpin, M. Kerdranvat, F. Hiron, P. de Lagrange, F. Le Léannec, K. Naser, T. Poirier, F. Racapé, G. Rath, A. Robert, F. Urban, T. Viellard, Y. Chen, W.-J. Chien, H.-C. Chuang, M. Coban, J. Dong, H. E. Egilmez, N. Hu, M. Karczewicz, A. Ramasubramonian, D. Rusanovskyy, A. Said, V. Seregin, G. Van Der Auwera, K. Zhang, L. Zhang, *Description of SDR, HDR and 360 Video Coding Technology Proposal By Qualcomm and Technicolor–medium Complexity Version*, document JVET-J0022, 2018.
[35] Z. Liu, K. Yang, X. Fu, M. Zhang, Z. Wang, and F. Mao, "Adaptive QP offset selection algorithm for virtual reality 360-degree video based on CTU complexity," *Multimedia Tools Appl.*, vol. 80, no. 3, pp. 3951–3967, Jan. 2021.
[36] R. G. Youvalari, A. Aminlou, and M. M. Hannuksela, "Analysis of regional down-sampling methods for coding of omnidirectional video," in *Proc. Picture Coding Symp. (PCS)*, Dec. 2016, pp. 1–5.
[37] Y. Ye, E. Alshina, and J. Boyce, *Algorithm Descriptions of Projection Format Conversion and Video Quality Metrics in 360Lib*, document JVET-H1004-v2, Oct. 2017.
[38] Y. Sun, A. Lu, and L. Yu, *AHG8: WS-PSNR for 360 Video Objective Quality Evaluation*, document JVET-D0040, Oct. 2016.
[39] P. Hanhart, J. Boyce, K. Choi, and K. Kin, *JVET Common Test Conditions and Evaluation Procedures for 360 Video*, document JVET-L1012-v1, Oct. 2018.
[40] E. Dubois, "The sampling and reconstruction of time-varying imagery with application in video systems," *Proc. IEEE*, vol. 73, no. 4, pp. 502–522, 1985.
[41] Y. Wang, J. Ostermann, and Y.-Q. Zhang, *Video Processing and Communications*. Englewood Cliffs, NY, USA: Prentice-Hall, 2002.

[42] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-Time Signal Processing*, 2nd ed., Englewood Cliffs, NJ, USA: Prentice-hall, 1999.
[43] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Upper Saddle River, NJ, USA: Prentice-Hall, 2018.
[44] Y.-K. Wang, R. Skupin, M. M. Hannuksela, S. Deshpande, Hendry, V. Drugeon, R. Sjöberg, B. Choi, V. Seregin, Y. Sanchez, J. M. Boyce, W. Wan, and G. J. Sullivan, "The high-level syntax of the versatile video coding (VVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 10, pp. 3779–3800, Oct. 2021.
[45] *Insta360 Pro 2 Camera 3D-360 Sample Videos*. Accessed: Jan. 21, 2023. [Online]. Available: https://www.insta360.com/product/insta360-pro2
[46] *Fraunhofer Versatile Video Encoder (VvenC-1.9.1) Source Code*. Accessed: Sep. 19, 2023. [Online]. Available: https://github.com/fraunhoferhhi/vvenc
[47] *OpenCV Library*. Accessed: Feb. 9, 2023. [Online]. Available: https://opencv.org
[48] G. Bjøntegaard, *Calculation of Average PSNR Differences Between RD-curves*, document VCEG-M33, Apr. 2001.
[49] A. M. Ahrar and H. Roodaki, "A new tile boundary artifact removal method for tile-based viewport-adaptive streaming in 360° videos," *Multimedia Tools Appl.*, vol. 80, no. 19, pp. 29785–29803, Aug. 2021.
[50] Z. Li, S. Qin, and L. Itti, "Visual attention guided bit allocation in video compression," *Image Vis. Comput.*, vol. 29, no. 1, pp. 1–14, Jan. 2011.

HOSSEIN PEJMAN (Student Member, IEEE) received the B.Sc. degree in computer engineering from Islamic Azad University, Central Tehran Branch, Tehran, Iran, in 2009, and the M.Sc. degree in computer engineering from Islamic Azad University, Science and Research Branch, Tehran, in 2012. He is currently pursuing the Ph.D. degree in computer engineering with the École de technologie supérieure (ÉTS), Montreal, Canada. His research interests include image and video processing, compression, computer vision, and machine learning with a recent focus on 360° video applications.

STÉPHANE COULOMBE (Senior Member, IEEE) received the B.Eng. degree in electrical engineering from the École Polytechnique de Montréal, Canada, in 1991, and the Ph.D. degree in telecommunications (image processing) from the INRS-Télécommunications, Montréal, in 1996.

From 1997 to 1999, he was with the Nortel Wireless Network Group, Montréal. From 1999 to 2004, he was a Research Engineer with the Nokia Research Center, Dallas, TX, USA, and later as the Program Manager of the Audiovisual Systems Laboratory. He joined the École de technologie supérieure (ÉTS-a constituent of the Université du Québec network), in 2004, where he is currently a Professor with the Department of Software and IT Engineering. From 2009 to 2018, he held the Vantrix Industrial Research Chair in video optimization. His research interests include video processing, compression, communications (transport), and systems, with a recent focus on immersive video and artificial intelligence for video and visual data applications.

CARLOS VÁZQUEZ (Senior Member, IEEE) received the B.Eng. and M.Sc. degrees from the Technical University of Havana (ISPJAE), in 1992 and 1997, respectively, and the Ph.D. degree from the INRS-EMT, Montréal, Canada, in 2003. From 2005 to 2012, he was a Research Scientist with the Communications Research Centre Canada (CRC). He has been an Associate Professor with the Software and IT Engineering Department, École de Technologie Supérieure (ÉTS-Montréal), since 2013. His research interests include image and video processing and computer vision, more specifically he is interested in 3D reconstruction from images, medical images processing, processing and coding of immersive visual content, augmented reality and 3D-360 video processing, and analysis and coding.

AHMAD VAKILI received the B.Sc. and M.Sc. degrees from Tehran Polytechnic and the Ph.D. degree from INRS, Montreal, in 2012. Since 2013, he has been with Summit Tech, where he is currently the Chief Research and Development Officer. His work focuses on advancing multimedia quality, real-time communications, and 360Â° video processing and streaming, as well as pioneering developments in edge computing, MEC, artificial intelligence, and machine learning.