Recent Advances in Omnidirectional Video Coding for Virtual Reality: Projection and Evaluation

Zhenzhong Chen, Yiming Li, Yingxue Zhang

PII: S0165-1684(18)30005-7
DOI: 10.1016/j.sigpro.2018.01.004
Reference: SIGPRO 6697

To appear in: *Signal Processing*

Received date: 31 July 2017
Revised date: 17 November 2017
Accepted date: 2 January 2018

Please cite this article as: Zhenzhong Chen, Yiming Li, Yingxue Zhang, Recent Advances in Omnidirectional Video Coding for Virtual Reality: Projection and Evaluation, *Signal Processing* (2018), doi: 10.1016/j.sigpro.2018.01.004

Highlights

- Overview of the recent advances of 360 video coding, especially in projection and evaluation methods
- Projections benefiting for 360 video coding are classified and compared
- The current problems and future trends of omnidirectional video processing are discussed

Recent Advances in Omnidirectional Video Coding for Virtual Reality: Projection and Evaluation

Zhenzhong Chen[1] [2], Yiming Li[1], and Yingxue Zhang[2]

[1] *State Key Laboratory of Software Engineering, School of Computer Science, Wuhan University, Wuhan 430072, P.R. China.*
[2] *School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, P.R. China.*

Abstract

In this paper, we review the recent advances in the pipeline of omnidirectional video processing including projection and evaluation. Being distinct from the traditional video, the omnidirectional video, also called panoramic video or 360 degree video, is in the spherical domain, thus specialized tools are necessary. For this type of video, each picture should be projected to a 2-D plane for encoding and decoding, adapting to the input of existing video coding systems. Thus the coding influence of the projection and the accuracy of the evaluation method are very important in this pipeline. Recent advances, such as different projection methods benefiting video coding, specialized video quality evaluation metrics and optimized methods for transmission, are all presented and classified in this paper. In addition, the coding performances under different projection methods are specified. The future trends of omnidirectional video processing are also discussed.

Keywords: Virtual reality; Omnidirectional video; Video coding; Projection; Evaluation

1. Introduction

Pursuing the immersive experience to simulate the real world in the digital devices has been an increasingly hot topic. Many efforts are in the way to provide better user experience with high resolution/quality video, HDR video

content, large screen display, *etc.* Recently, with the availability of commercial Virtual Reality (VR) Head Mounted Displays (HMD) such as Oculus Rift or HTC Vive, VR video application attracts great attention. With these products, users can enjoy the omnidirectional video and can choose their desired viewport by moving heads as they do in the real world, thus the immersive experience can be provided. As the content of VR, the demand of omnidirectional video proliferates with the increasing attraction and popularity of VR applications, while it should be noted that there still exist many obstacles for omnidirectional video processing. For immersive visual experience, high resolution (6K or beyond) and high frame rate (*e.g.*, 90 fps) are expected, so that the bitstream tends to be very large, causing severe resource consuming on storage and bandwidth. Therefore, improving compression efficiency of omnidirectional video is in urgent demand. However, being different from the traditional 2-D video, omnidirectional video is in the spherical domain that is a bounding sphere containing the content of the whole surroundings. In other words, though many video coding standards have been developed by the International Telecommunications Union (ITU) and Motion Picture Expert Group (MPEG), *e.g.*, H.264/AVC [1], H.265/HEVC [2], there is no specialized video coding algorithm for spherical domain video coding. The lack of efficient compression method for omnidirectional video significantly affects the development of VR application.

To improve the omnidirectional video coding efficiency, Joint Video Exploration Team (JVET) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11 have established an Ad hoc group for this research recently. Considering the development and efficiency of current conventional video coding standards, it is suggested to project the original spherical information into a 2-D plane for encoding so that the current video coding framework can be used. The pipeline of omnidirectional video coding is shown in Fig. 1, from which the following challenging issues are illustrated:

(1) Projections: The transformation from sphere to 2-D plane will introduce artifacts, like the redundant samples, shape distortion and discontinuous boundary. The redundant samples cause many invalid pixels to be coded. The

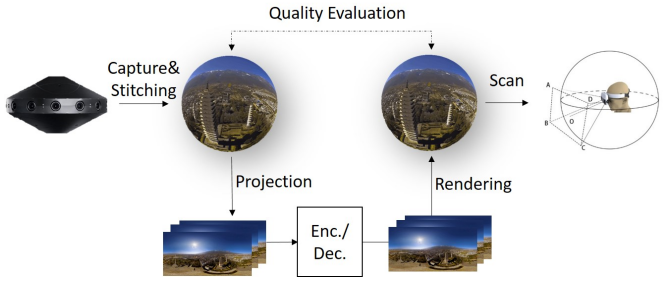


Figure 1: The pipeline of omnidirectional video coding.

discontinuous boundary affects the prediction performance and the shape distortion leads to inefficient Motion Estimation (ME) and Motion Compensation (MC) in video coding. In general, different projection methods result in different kinds of artifacts. For example, Equirectangular Projection (ERP) suffers from redundant samples and horizontal stretching problem near the pole area. For the research of high efficiency omnidirectional video coding, it is necessary to evaluate the coding performance of different projections and choose the best one, despite that there are infinite kinds of projections and each brings unique effect on the final 2-D plane [3].

(2) Evaluation criteria: Quality assessment is an important issue in video coding. In traditional 2-D video coding, a great number of objective quality metrics have been proposed, *e.g.*, mean squared error (MSE), PSNR, structural similarity index (SSIM) [4], and many other full reference (FR) image/video quality assessment methods baed on human visual system (HVS) [5][6][7][8][9]. It should be noted that omnidirectional video is commonly represented by projection planes. At the display side, an inverse projection is performed before viewport rendering. This non-linear transformation leads to the condition that the pixels in these two domains do not correspond to each other, which means the distortion calculated in 2-D plane cannot reflect the actual distortion in spherical domain. To measure the accurate quality, a new evaluation criterion should be proposed.

Besides, as mentioned before, the omnidirectional video coding and eval-

uation is different from that of the traditional video, thus some specialized optimization algorithms are also proposed to improve the coding and transmission efficiency. Generally, the research on omnidirectional video coding is on the rise. Many new schemes have been proposed, thus a detailed summarization is necessary. In this paper, we give a review of the recent advances in the omnidirectional video for those aforementioned challenges, especially in the projection process and evaluation metrics. Since the coding optimization tools are mostly designed for a specific projection map and currently the primary projection method has not been specified yet, it is not involved in this paper and some typical algorithms can be found in [10, 11, 12, 13, 14, 15]. The organization of the paper is as follows. We describe the background and the framework for omnidirectional video coding in next section. A review and discussion of different projection methods are given in Section 3. In Section 4, the recent advances in the omnidirectional video quality evaluation are presented. The performance of different projections and the accuracy of different evaluation metrics are discussed in Section 5. Finally, we summarize our paper in Section 6.

2. The Framework for Omnidirectional Video Coding

As a joint group of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET is the main international working group for the research and standardization of omnidirectional video coding. The goal of their research will mostly be involved in this paper, which can be summarized as:

- Study the effect on compression of different omnidirectional video projections.
- Discuss refinements of common test conditions, test sequences, and evaluation criteria, including subjective evaluation/objective evaluation.
- Study viewpoint generation methods and viewport-dependent video coding and streaming.

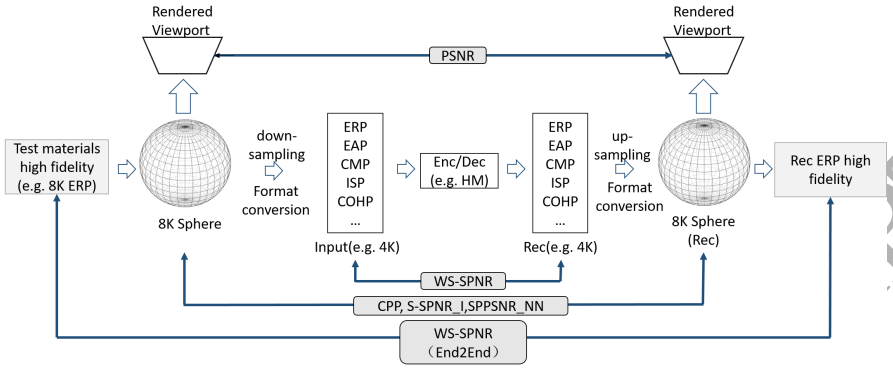


Figure 2: Omnidirectional video testing procedure recommended by JVET.

- Study coding tools dedicated to omnidirectional video, and their impact on compression.

Fig. 2 shows the omnidirectional video coding framework recommended by JVET. First, high fidelity test materials are provided in ERP and will be regarded as the ground truth for quality evaluation. Various projections are generated from these high fidelity test materials for coding efficiency comparison. Since the pixels in the projection plane mostly do not correspond to the integer pixels in the original sphere, interpolation operation is necessary in the projection process. He *et al.* pointed out that for luma pixel interpolation, 6 tap Lanczos is the best choise and for chroma, 4 tap Lanczos is accurate enough, which gives a good trade off between the accuracy and time consuming [16], and thus being suggested in the testing procedure. Besides, the down-sampling process is used to eliminate the unfair bias among ERP and other projections [17][18]. For quality evaluation, the uniform quality evaluation methods in spherical domain like CPP-PSNR, S-PSNR and WS-PSNR are selected and they will be introduced in details in Section 4. Traditional PSNR is used to evaluate the quality in rendered viewport. The encoder/decoder for coding process is agnostic while HM [19] or JEM [20] is preferred. This testing procedure is only used for research, especially for the exploration on the coding efficiency of different projections and the difference of several quality evaluation

Figure 3: The map projected by ERP.

metrics. It is noted if the projection method and evaluation criterion are specified, the testing procedure can be simplified and the down-smapling process will be unnecessary.

3. Omnidirectional video projections

Omnidirectional video projection can be classified into two categories: viewport-independent projection and viewport-dependent projection. The viewport-independent projection can be further classified into map-based projection, patch-based projection, tile-based projection and rotation-based projection. For the viewport-dependent projection, it is used for VR streaming, thus some generalized projection methods benefiting for streaming (*e.g.* tiling methods) are also included.

3.1. Viewport-independent projection

3.1.1. Map-based projection

Equirectangular Projection (ERP) and Cylindrical Equal-Area Projection (EAP) are the two most original projections, which derive from the map projection. Thus, the compression efficiency is not the key factor to be considered in ERP and EAP.

As shown in Fig. 3, the horizontal and vertical coordinates in ERP correspond to the longitude and the latitude in sphere. Since the longitude varies from 0 to 2π and latitude varies from 0 to π, the ERP is normally presented in a

Figure 4: The map projected by EAP.

2:1 ratio of width to height. Because ERP is intuitive and easy to generate, it is the most commonly used projection method. But its drawback is also obvious, *i.e.*, the oversampling near the pole resulting in large shape distortion and much bit wasting in encoding.

Similar to ERP, EAP is originally a map projection method, which is proposed to solve the oversampling problem in ERP. As shown in Fig. 4, EAP adaptively decreases the sampling rate in vertical coordinate by multiplying $cos(\theta)$, where θ is the latitude. In this way, EAP can avoid oversampling and guarantee that the area is equal to the original sphere, at the cost of a more serious shape distortion.

The Cube Map Projection (CMP) assumes that there is a circumscribed cubic box surrounding the sphere, the pixels on the sphere are projected to the cube firstly, then cube is unfolded into 6 surfaces and rearranged for compact expression. Different arrangements of the unfolded surfaces cause different compression efficiency [21]. Currently, the arrangement in Fig. 5 is suggested by JVET. Compared with ERP, CMP is more suitable for graphics library rendering, thus it is mostly used in VR games. And the shape is not distorted in CMP, bringing a better ME and MC efficiency. Whilst, the cubic projection still results in a suboptimal resolution distribution, with resolution increasing towards the corners of the cube and away from the forward viewing direction. In [22], the oversampling rate of CMP is given, which is up to 190% compared

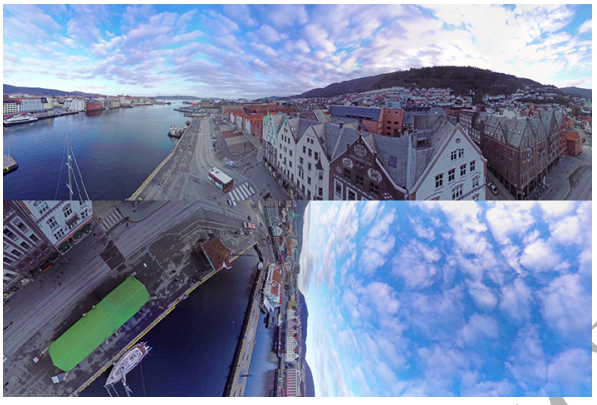


Figure 5: The map projected by CMP.

to the original sphere.

3.1.2. Patch-based projection

It must be pointed out that CMP is a kind of patch-based projection. To solve the oversampling problem of CMP, some newly proposed method try to use polyhedron with more faces to approach the ideal sampling rate. We name them patch-based projections.

As shown in Fig. 6, a dodecahedron-projection is proposed in [23]. The sphere is projected to a rhombus dodecahedron, then it is split and rearranged into a 3×4 rectangle. The principle of rearrangement depends on the location correlation, in order to reduce the discontinuous side as far as possible.

A similar octahedron-projection, *i.e.*, OHP is introduced in [24][25], which is illustrated as Fig. 7. The information in spherical domain is projected to the octahedron face and then unfolded. For patch-based projections, there are many kinds of arrangement methods for the rearrangement of patches. JVET recommends two rearrangement schemes for coding efficiency comparison namely Compact Layout 1 for the Octahedron Projection (COHP1) and Compact Layout 2 for the Octahedron Projection (COHP2).

Besides, another Icosahedron Projection (ISP) is proposed by Samsung [26], which can be rearranged into a compact format as well, called CISP. Fig. 8

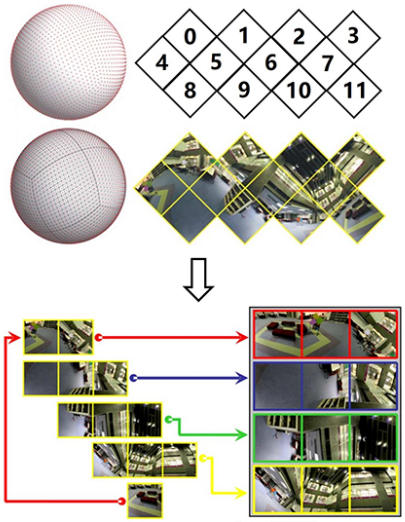


Figure 6: The rhombus dodecahedron projection. [23]

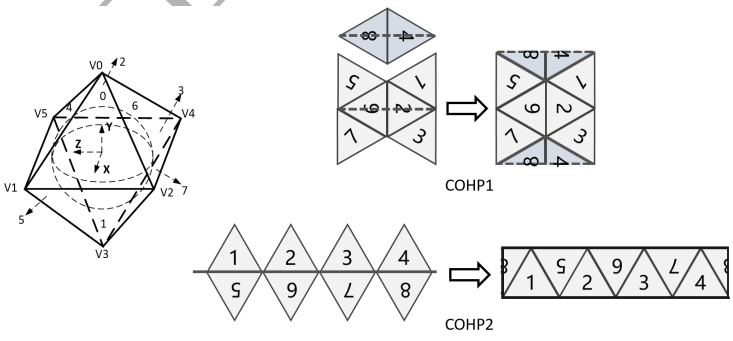


Figure 7: The process of OHP and COHP.

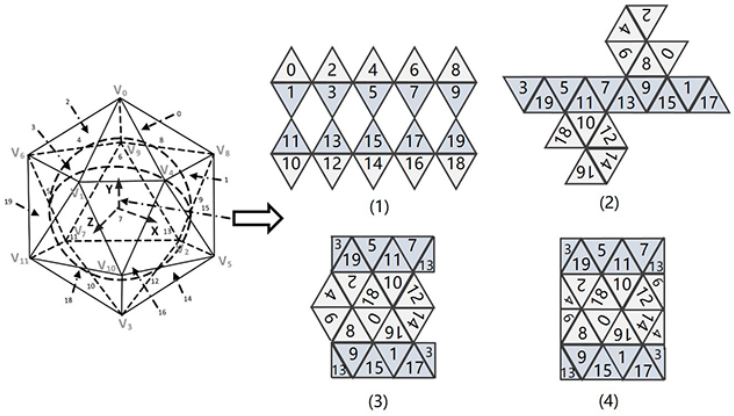


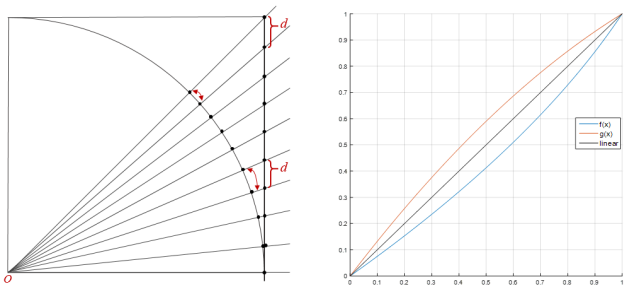
Figure 8: The process of CISP.

illustrate the projection process.

For these patch-based projections, more faces or more patches decrease the oversampling rate, at the cost of an increasing number of discontinuous boundaries. To solve the oversampling problem of CMP whilst avoiding the rise of discontinuous boundaries, Google and Qualcomm Inc. conducted two similar work, where Equi-Angular Cubemap (EAC) Projection and Adjusted Cubemap Projection (ACP) are provided [27][28]. As shown in Fig. 9(a), the oversampling problem of CMP is due to the nonuniform projection in different angles, thus a nonlinear transformation is proposed. As illustrated by the blue line in Fig. 9(b), after the first projection calculated by the original CMP, a sampling rate adjustment process will be added based on the location to compensate the oversampling problem introduced by the original CMP. The horizontal axis in Fig. 9(b) means the current angle to the center, where 0.5 corresponds to the diagonal in Fig. 9(a). The vertical axis is the second sampling rate in the adjustment process.

3.1.3. Tile-based projection

Yu *et al.* [29] proposes a tile-based projection, which splits the ERP plane into several tiles shown in Fig. 10. The sampling rate in horizontal direction is



(a) (b)

Figure 9: The ACP projection. (a) shows the problem in traditional CMP. (b) is the adjustment function of ACP projection.

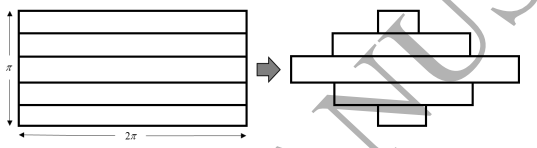


Figure 10: Illustration of tiled-based projection proposed by Yu *et al.* [29].

decreased according to the original sampling rate on the sphere surface, thus the oversampling can be avoided. As illustrated in [29], this scheme could promote up to 12.2% coding efficiency in intra coding.

As shown in Fig. 11, a similar tile-based projection is introduced by Li *et al.* [22]. Two poles in sphere are projected to circles instead of tiles, to eliminate distortion and improve the coding efficiency. In [22], the sampling rate of this scheme is calculated, which is about 113% while the sampling rate of Yu's method is 123%, thus it has better coding efficiency compared with Yu's work. They further optimizes their work in the JVET proposal [30], in which the vertical layout is suggested for the sake of a smaller line buffer [30]. The number of tiles is decreased to 3 for less discontinuous boundaries as well. This new scheme, also called Segmented Sphere Projection (SSP), is illustrated in Fig. 12.

Apart from the projections mentioned above, Sreedhar *et al.* [31] proposes to use nested polygonal chain packing to solve the distortion in the pole area of

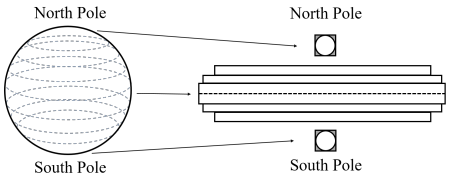


Figure 11: Illustration of tiled-based projection proposed by Li *et al.* [22].

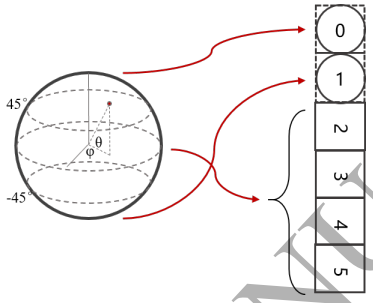


Figure 12: The process of SSP [30].

ERP. A similar method is introduced by Wang [32] from Peking University.

3.1.4. Rotation-based projection

Some work found that the coding efficiency could be further improved by rotating the original sphere surface before projection [34][35]. On this basis, a Rotated Sphere Projection (RSP) is proposed by Abbas [33] from Gopro Inc., this scheme unfolds the sphere under two different rotation angle and stitches them like a baseball surface. As shown in Fig. 13, in this projection, two faces are arranged in two rows in the final projection plane. The first face (or row) can be obtained by directly clipping the middle 270×90 degree region from the ERP image. If we rotate the sphere, ERP image can also be rotated so that pixels at the poles are brought to the equator and pixels corresponding to back face are brought to the front, then the second face (or row) is generated by clipping the same region as the first face. The dotted line in Fig. 13 illustrates the clipping region.

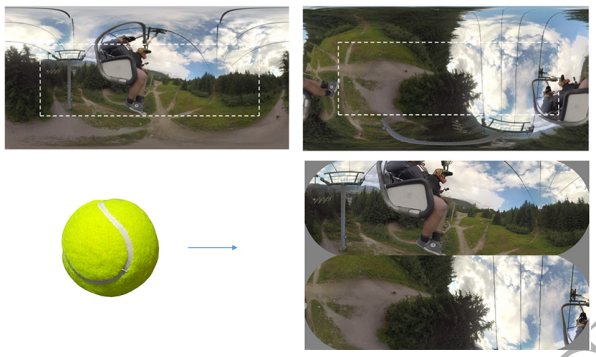


Figure 13: The process of RSP. [33]

3.2. Viewport-dependent projection

3.2.1. Typical viewport-dependent projection

Differing from the viewport-independent projection, this kind of projection takes the Field of View (FoV) into consideration. The main idea is to transmit only the visible information instead of alleviating projection distortion like the viewport-independent projection. A pyramid projection is proposed by Facebook Inc. [36], as shown in Fig. 14. The original spherical surface is projected to a pyramid, and then unfolded and rearranged to a rectangular. The area of user's current view will be projected to the bottom of the pyramid, which is the only face sampled in full resolution while others will be projected to the other faces of the pyramid. By this way much bit rate can be reduced by ignoring the irrelevant pixels, which is beneficial for streaming. However, this scheme needs to encode one video into multi representations for different viewport, and choose the corresponding faces according to the user's view. In their work, 30 unfolded pyramid maps with different viewports are pre-encoded, thus this scheme can save a lot of transmission bandwidth, but at the cost of storage resource. Besides, similar scheme is also proposed by Qualcomm, named as TSP [37]. The TSP uses quadrangular frustum pyramid instead of pyramid, which can alleviate the video quality degradation when users change their viewpoint.

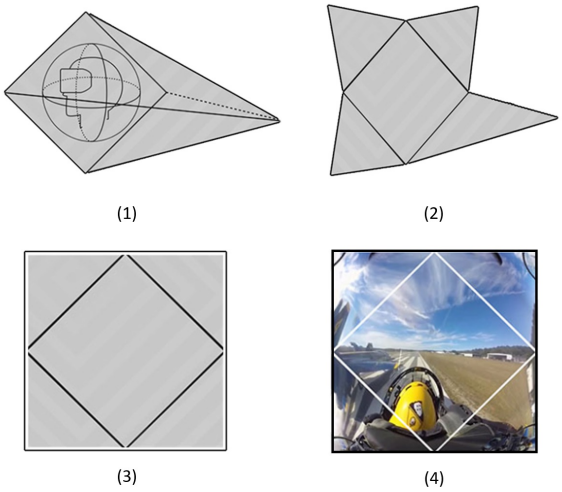


Figure 14: The process of the pyramid projection. [36]

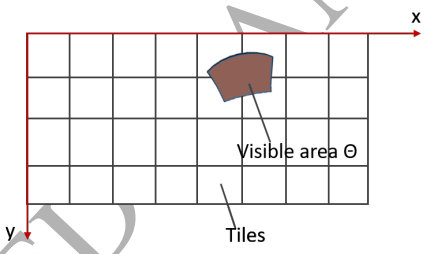


Figure 15: Visualization of the tiling approach. [38]

3.2.2. Geometric-layout-agnostic approach

As mentioned, typical viewport-dependent projections are designed for streaming extremely high bit rate omnidirectional video in the such limited bandwidth. Actually, there are some other methods, we name them geometric-layout-agnostic approach, to address the same issue. Although it is hard to identify whether this approach belongs to projection or streaming technology, considering the intrinsic similarity between these methods (streaming corresponding video signal according to user's view), we still decide to introduce them in this Section.

As a typical projection-agnostic approach, tiling has been well studied in some earlier work [39][40][41], which aim to split the large-resolution video into several tiles and transmit them independently. Specifically, only the visible tiles corresponding to user's view are transmitted (or in higher quality/resolution). Fig. 15 visualizes a typical tiling approach where all the tiles are pre-encoded in a set of quality levels. At the streaming process, only the tiles containing visible area are transmitted in high quality or high resolution and the other tiles are streamed in basic low-quality level. In these tiling methods, the projection layout is agnostic. ERP is used for tiling in [42][43][38]. CMP is used in [44], and in [45][46] the so called hexaface sphere is used as the geometric layout. MPEG-DASH is mostly used to support their work, where Media Presentation Description (MPD) [47] stores all the tiles with different video quality/resolution (*e.g.*, [45][46][48]). When the client requests for a video with viewport, the server will transmit the corresponding tiles by the MPD file. Besides, utilizing the scalable video coding (SVC) base and enhancement layers [44] is another choice to support the tiles with a set of quality levels.

In [49], it it pointed out that there are several drawbacks for the tiling methods, *e.g.*, the client has to reconstruct the video from independent tiles so that the latency may be increased. Thus, in the proposed work, a viewport-adaptive quality method is assumed to be better. As illustrated in Fig. 16, the server pre-defines several ROI areas named quality emphasized region (QER), which means the quality of ROI area is much better than the others. Given different QER, each video is encoded into multi-representation for storing. At the transmission side, the corresponding representation is sent. Compared to the tiling methods, this approach can solve the tile independence problem, remove the additional reconstruction process and reduce the file number, while the storage redundancy for non-ROI area (many duplications exist in each representation) cannot be neglected, neither do the problems of Facebook pyramid projection and TSP.

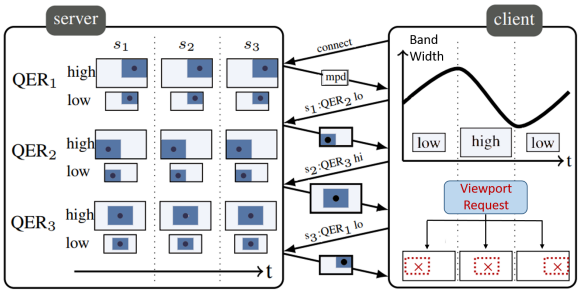


Figure 16: Viewport-adaptive streaming system: the server offers 6 representations (3 QERs at 2 bit-rates). [49]

4. Omnidirectional Video Quality Evaluation

To evaluate the coding efficiency of the large amount of projection methods proposed in the previous section, an accurate omnidirectional video quality evaluation criterion must be adopted. Since omnidirectional video will be rendered to sphere or viewport after decoding for human viewing, the problem that PSNR does not reflect the actual omnidirectional video quality needs to be addressed. In this section, the objective quality evaluation indicators recommended by JVET are reviewed.

Considering that the videos will be rendered to the sphere, Yu *et al.*[17] proposed an evaluation framework upon the spherical domain. As shown in Fig. 17. The pixels in the original map and in the encoded 2-D plane were first mapped to the same sphere. Then a large number (in the implementation, 655362 was applied) of sampling points uniformly distributing on the sphere were used to calculate the mean error between the original and encoded signals as a replacement of PSNR calculated in 2-D plane. In this work, two indicators, *i.e.*, S-PSNR and L-PSNR, were given. L-PSNR assigns larger weight to the pixels near equator, while S-PSNR does not. S-PSNR is adopted by JVET committee as one of the indicators, which can approximate the average quality on all of the possible views presented to the observers.

In order to conduct an equal comparison of videos across multiple projection

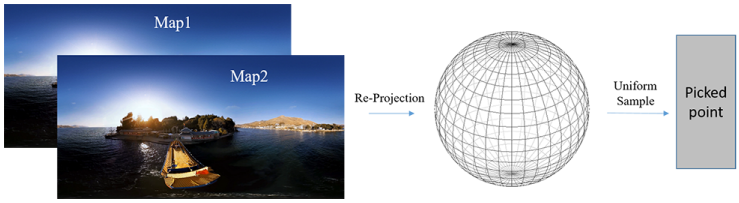


Figure 17: S-PSNR: Comparison of signals on a same sphere. [13]

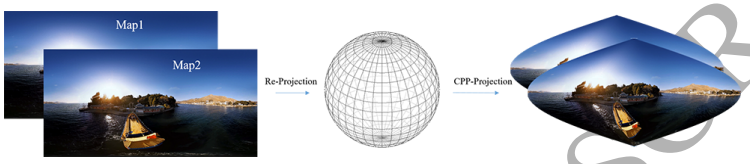


Figure 18: CPP-PSNR: Project to sphere First and calculate upon CPP projection. [13]

schemes, another indicator is proposed by Zakharchenko *et al.* [50, 51, 52], which is called CPP-PSNR. Like S-PSNR, CPP-PSNR also maps the pixels from 2-D plane to sphere first. While S-PSNR uses limited number of sampling points, CPP-PSNR chooses a craster parabolic projection [53] (CPP) method to project the sphere without spatial resolution change, which is shown in Fig. 18. All the test videos were transformed into the CPP format so that the distortion could then be calculated in the CPP plane, which enables quality comparisons between different projection schemes under equal condition and eliminates the influence of projection variations on the quality.

Apart from S-PSNR and CPP-PSNR, WS-PSNR proposed by Sun [54, 55] is also an important indicator. Compared with the aforementioned two metrics, this scheme does not need to remap the plane. Instead, it evaluates the distortion with the weights given offline. For example, the weights of distortion in ERP is shown in (1):

$$W(i,j) = \frac{w(i,j)}{\sum_{i=0}^{width-1}\sum_{j=0}^{height-1} w(i,j)} \tag{1}$$

where $width$ and $height$ are the size of images. $w(i,j)$ is the scaling factor of area from equirectangular to sphere, which can be represented as:

$$w(i,j) = cos\left(\left(j - \frac{height}{2} + \frac{1}{2}\right)\cdot\frac{\pi}{height}\right) \tag{2}$$

Like the formula of PSNR ,WS-PSNR is obtained using:

$$WSPSNR = 10\log\left(\frac{MAX^2}{WMSE}\right) \tag{3}$$

$$WMSE = \sum_{i=0}^{width-1}\sum_{j=0}^{height-1}\left(y\left(i,j\right)-y'\left(i,j\right)\right)^2\cdot W\left(i,j\right) \tag{4}$$

where W is calculated in (1), $y(i,j)$, $y'(i,j)$ are the original pixel value and reconstruct pixel value and MAX is the max pixel value in the image. It should be noted that different projections lead to different weights. As for cubic projection with $a \times a$ resolution, the scaling factor is [54]:

$$w(i,j) = \left(3+\frac{(i+1)^2+(j+1)^2-(i+j)\cdot a}{a^2/4}\right)^{-3/2} \tag{5}$$

where (i,j) denotes the position in a face instead of the whole frame. The formula is the same for each face.

Without resampling, WS-PSNR shows more robust performance than S-PSNR and CPP-PSNR given high frequency distortions in that the latter two methods conduct interpolation in resampling process and thus eliminating some noises, especially in high frequency domain, which results in inaccurate performance under this circumstance.

In general, all the metrics mentioned above fix the flaws of the conventional PSNR on omnidirectional video quality evaluation from different aspects and are recommended as the quality indicators of omnidirectional videos in the common test condition (CTC) by JVET [18].

5. The Comparison of Projections and Evaluation Criteria

To evaluate the influence of different projections and validate the accuracy of different evaluation criteria, a simulative experiment is conducted. The experiment is conducted under the testing procedure shown in Fig. 2, which is recommended by JVET for the evaluation of different projections.

In accordance with the common test condition (CTC), 10 test sequences recommended by JVET are chosen, as shown in Fig. 19, the resolution of

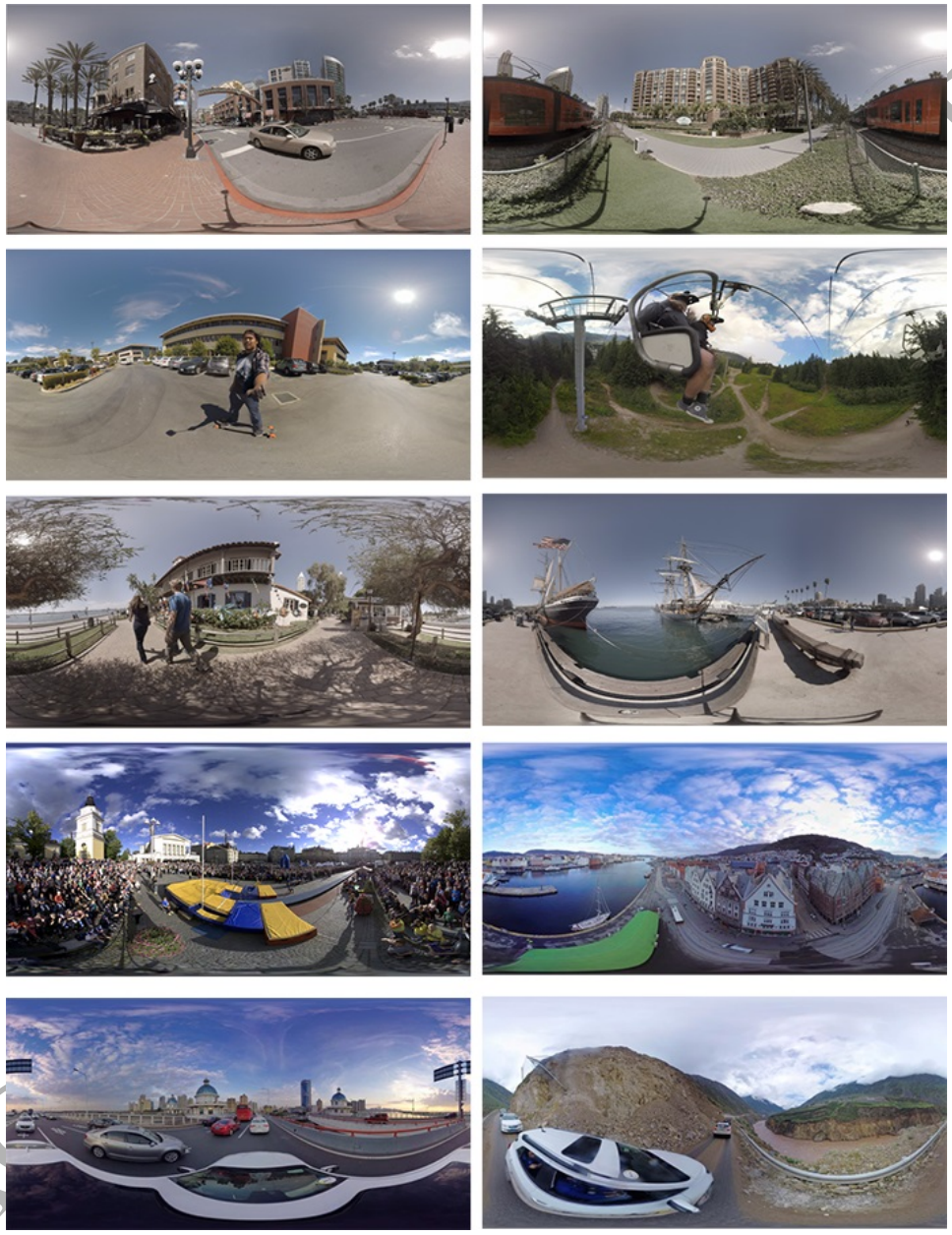


Figure 19: The thumbnails of the test sequences.

Table 1: Comparison in random access main10 configuration

Projections	S-PSNR_NN			S-PSNR_I			WS-PSNR			CPP-PSNR		
	Y	U	V	Y	U	V	Y	U	V	Y	U	V
EAP vs. ERP	11.7%	-2.3%	-3.1%	11.9%	-2.4%	-3.2%	11.5%	-2.3%	-3.2%	11.6%	-2.4%	-3.3%
CMP vs. ERP	-3.9%	-2.6%	-2.5%	-3.8%	-2.7%	-2.7%	-3.8%	-2.7%	-2.6%	-3.8%	-2.8%	-2.8%
COHP1 vs. ERP	-1.3%	3.3%	2.3%	-1.3%	3.0%	2.0%	-1.4%	3.3%	2.3%	-1.4%	3.0%	2.0%
COHP2 vs. ERP	2.5%	12.2%	11.1%	2.5%	11.4%	10.5%	2.5%	12.1%	11.1%	2.5%	11.5%	10.6%
CISP vs. ERP	-5.0%	-0.4%	-1.2%	-4.9%	-0.6%	-1.3%	-5.2%	-0.5%	-1.3%	-5.0%	-0.6%	-1.4%
SSP vs. ERP	-9.8%	-3.2%	-3.8%	-9.4%	-3.3%	-3.9%	-9.8%	-3.1%	-3.7%	-9.5%	-3.5%	-4.1%
ACP vs. ERP	-11.3%	-6.3%	-6.4%	-11.2%	-6.4%	-6.6%	-11.3%	-6.3%	-6.4%	-11.2%	-6.4%	-6.6%
RSP vs. ERP	-10.5%	-4.4%	-5.1%	-10.3%	-4.7%	-5.3%	-10.5%	-4.5%	-5.2%	-10.4%	-4.8%	-5.4%

these sequences is 8192×4096 or 3840×1920. The down-sampling operation is conducted in the projection process and the down-sampled projection map is used as the input of the codec. In this experiment, typical projection methods are selected, including ERP, EAP, CMP, COHP1, COHP2, CISP, Vertical SSP, ACP, and RSP, where ERP is selected as the anchor. Since the objective quality evaluation criteria are all designed for the evaluation of the whole map and cannot handle the viewport-dependent projections, they are not involved in this experiment. This test is under the the latest available JVET CTC [18]. 360Lib 2.1 and HM 16.15 are selected as the software. The QP values are 22, 27, 32 and 37 respectively. According to the CTC, our experiment is only conducted in the random access (RA) configuration. The results are shown in Table 1 and Fig. 20.

In this table, BD-Rate is calculated according to different evaluation methods, where S-PSNR_I and S-PSNR_NN are all based on S-PSNR, the difference is that S-PSNR_I uses the Lanczos interpolation to calculate the fractional pixels values, whilst S-PSNR_NN only uses the nearest neighbor algorithm for the calculation of the values in fractional pixels. From this table, we can see that ACP outperforms others in all evaluation criteria and different evaluation criteria have the similar BD-rate value. Besides, some noticeable phenomenon can also be found: Sampling rate does not necessarily correlate with coding performance. As we can see, EAP solves the oversampling problem of ERP, and the oversampling problem of CMP is much worse than ERP, while the coding per-

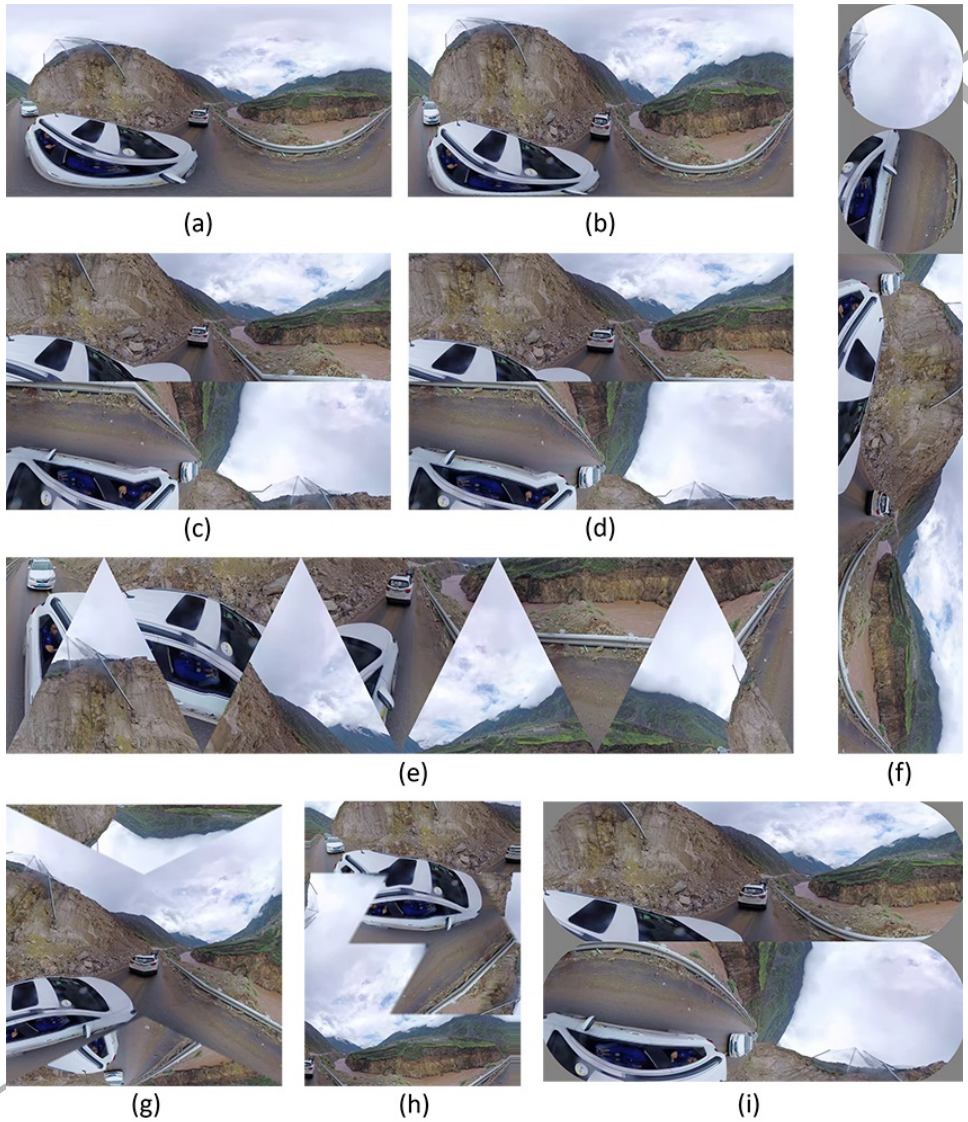


Figure 20: The coding results of the same sequence in different types of projection (QP = 37, Sequence: DrivingInCountry): (a) ERP, (b) EAP, (c) CMP, (d) ACP, (e) COHP1, (f) vertical SSP, (g) COHP2, (h) CISP, and (i) RSP.

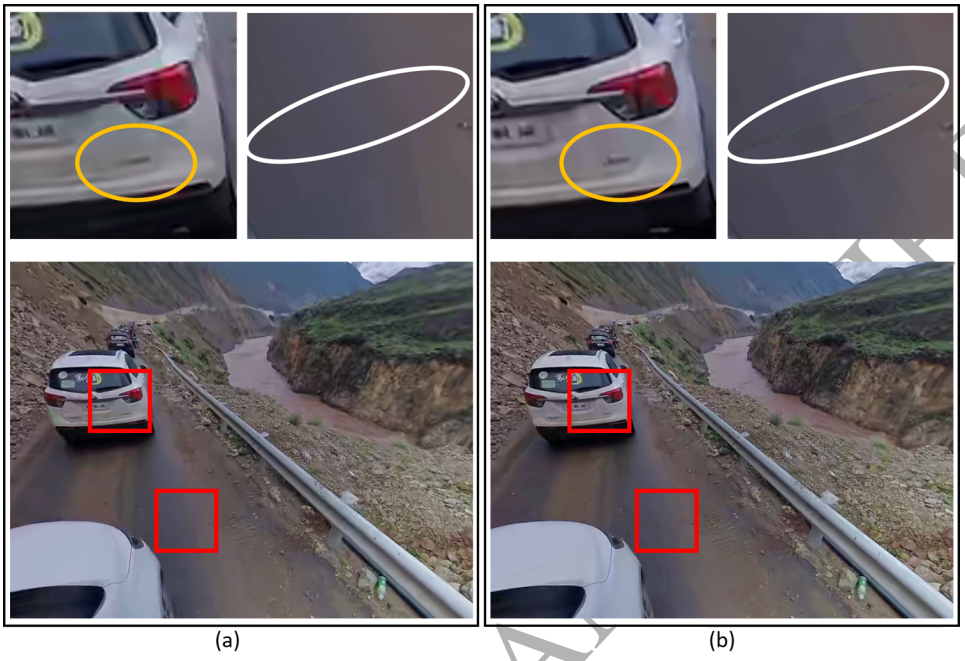


Figure 21: The coding results shown by viewport. (QP = 37, Sequence: DrivingInCountry) (a) shows the viewport of ERP and magnifying local areas from the viewport. (b) shows the viewport of ACP and magnifying local areas from the viewport.

formance is opposite. The coding efficiency and shape distortion have a strong correlation, so that we can see all the projections except EAP achieves great improvement compared with ERP.

From Fig. 20, it can be noted that most of the projections clip the pictures and rearrange them, generating discontinuous boundaries in the projection maps. Although some projections like CISP [26] and SSP [30] have tried to alleviate the discontinuity, the improvement is quite limited currently. The discontinuous boundaries may greatly influence intra prediction and motion estimation in video coding and consequently introduce some artifacts. To illustrate the problem, we zoom in some local areas of the viewport (a) and (b) shown in Fig. 21. (a) is the viewport from ERP and (b) is that from ACP. Since ACP achieves better coding performance than ERP, it generally retains more details

as shown in the yellow circle area. Concerning the discontinuity marked out with the white circle, however, inevitably affects the subjective perception of the quality. Padding/Overlapping method [29][56] in the discontinuous area can be seen as a compromise between the coding efficiency and perceptual quality. But due to the unavoidable coding performance decrease, further researches are still expected to address this dilemma.

6. Discussions and Conclusions

As a trend of future, it is obvious that VR technology has a broad space of application, but the large bitstream of omnidirectional video and the consumption of bandwidth pose a great challenge to the existing technologies. In this paper, we review most of the projection and quality evaluation methods, where various orientation of omnidirectional video research is opened:

(1) Compared with the traditional ERP projection, some of the new viewport-independent projection have achieved a good improvement, but it is still hard to meet the demand for real time transmission of extremely high-resolution. Oppositely, viewport-dependent approach can save much bandwidth when transmitting, while the duplication of bit-stream is the severe restriction for its application in storage area. Thus, a better representation approach which has the ability to facilitate the coding efficiency further or to combine the advantage of both sides is still in need.

(2) Since the streaming VR can be applied in many applications, the demand of viewport-dependent and corresponding evaluation criteria is going on the rise. For the objective quality evaluation criteria reviewed in this paper, although the metrics compared in the experiments are very similar and accurate, they are all designed for the evaluation of the whole map, which cannot handle the viewport-dependent projection. Evaluation criteria designed for viewport-dependent methods should not only take the PSNR calculated in viewport. The quality of viewport-surrounding area is also required since the viewport prediction error or transmission delay in the system is uncontrollable. The weights

of the non-viewport area should be defined corresponding to the accuracy of the viewport between which the client requested and received. Besides, Human visual system (HVS) is well studied in 2-D image/video objective quality evaluation, more research efforts are still expected for omnidirectional video.

In general, efforts are still on the way for omnidirectional video coding. As the future video coding standard (FVC) will begin to call for proposal, which also includes VR video coding, We believe that, in the near future, VR video compression can be perfectly handled and the immersive experience can be throughly enjoyed.

Acknowledgment

This work was supported by National Natural Science Foundation of China (No. 61471273, No. 61771348), Wuhan Morning Light Plan of Youth Science and Technology, LIESMARS Special Research Funding, and national key research & development (R&D) plan (No. 2017YFB1002202).

References

References

[1] T. Wiegand, G. J. Sullivan, G. Bjontegaard, A. Luthra, Overview of the H. 264/AVC video coding standard, IEEE Trans. Circuits Syst. Video Technol. 13 (7) (2003) 560–576.

[2] G. J. Sullivan, J.-R. Ohm, W. J. Han, T. Wiegand, Overview of the high efficiency video coding (HEVC) standard, IEEE Trans. Circuits Syst. Video Technol. 22 (12) (2012) 1649–1668.

[3] L. Zelnik-Manor, G. Peters, P. Perona, Squaring the circle in panoramas, in: Proc. IEEE Int. Conf. Computer Vision (ICCV), 2005, pp. 1292–1299.

[4] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.

[5] Z. Wang, Q. Li, Information content weighting for perceptual image quality assessment, IEEE Trans. Image Process. 20 (5) (2011) 1185–1198.

[6] L. Zhang, L. Zhang, X. Mou, D. Zhang, FSIM: A feature similarity index for image quality assessment, IEEE Trans. Image Process. 20 (8) (2011) 2378–2386.

[7] S. Wang, C. Deng, W. Lin, G.-B. Huang, B. Zhao, NMF-based image quality assessment using extreme learning machine, IEEE Trans. Cybern. 47 (1) (2017) 232–243.

[8] M. H. Pinson, S. Wolf, A new standardized method for objectively measuring video quality, IEEE Trans. Broadcast. 50 (3) (2004) 312–322.

[9] K. Seshadrinathan, A. C. Bovik, Motion tuned spatio-temporal quality assessment of natural videos, IEEE Trans. Image Process. 19 (2) (2010) 335–350.

[10] R. G. Youvalari, A. Aminlou, M. M. Hannuksela, M. Gabbouj, Efficient coding of 360-degree pseudo-cylindrical panoramic video for virtual reality applications, in: Proc. IEEE Int. Symp. Multimedia (ISM), 2016, pp. 525–528.

[11] Y. Liu, M. Xu, C. Li, S. Li, Z. Wang, A novel rate control scheme for panoramic video coding, in: Proc. IEEE Int. Conf. Multimedia and Expo (ICME), 2017.

[12] L. Li, Z. Li, X. Ma, H. Yang, H. Li, Co-projection-plane based 3-D padding for polyhedron projection for 360-degree video, in: Proc. IEEE Int. Conf. Multimedia and Expo (ICME), 2017.

[13] Y. Li, J. Xu, Z. Chen, Spherical domain rate-distortion optimization for 360-degree video, in: Proc. IEEE Int. Conf. Multimedia and Expo (ICME), 2017.

[14] J. Sauer, M. Wien, AHG8: Results for geometry correction for motion compensation of planar-projected 360VR video with JEM4.1 and 360Lib, in: Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET-E0026, Geneva, 2017.

[15] M. Budagavi, J. Furton, G. Jin, A. Saxena, J. Wilkinson, A. Dickerson, 360 degrees video coding using region adaptive smoothing, in: Proc. IEEE Int. Conf. Image Process. (ICIP), 2015, pp. 750–754.

[16] Y. He, Y. Ye, B. Vishwanath, AHG8: Interpolation filters for 360 video geometry conversion and coding, in: Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET-D0073, Chengdu, 2016.

[17] M. Yu, H. Lakshman, B. Girod, A framework to evaluate omnidirectional video coding schemes, in: IEEE Int. Symp. Mixed and Augmented Reality (ISMAR), 2015.

[18] J. Boyce, E. Alshina, A. Abbas, Y. Ye, JVET common test conditions and evaluation procedures for 360 video, in: Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET-E1030, Geneva, 2017.

[19] HEVC test software, https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware, accessed July 30, 2017.

[20] JEM test software, https://jvet.hhi.fraunhofer.de/svn/svn_HMJEMSoftware, accessed July 30, 2017.

[21] M. Zhou, AHG8: A study on compression efficiency of cube projection, in: Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET-D0022, Chengdu, 2016.

[22] J. Li, Z. Wen, S. Li, Y. Zhao, B. Guo, J. Wen, Novel tile segmentation scheme for omnidirectional video, in: Proc. IEEE Int. Conf. Image Process. (ICIP), 2016, pp. 370–374.

[23] C.-W. Fu, L. Wan, T.-T. Wong, C.-S. Leung, The rhombic dodecahedron map: An efficient scheme for encoding panoramic video, IEEE Trans. Multimedia 11 (4) (2009) 634–644.

[24] H.-C. Lin, C.-Y. Li, J.-L. Lin, S.-K. Chang, C.-C. Ju, AHG8: An efficient compact layout for octahedron format, in: Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET-D0142, Chengdu, 2016.

[25] H.-C. Lin, C.-C. Huang, C.-Y. Li, Y.-H. Lee, J.-L. Lin, S.-K. Chang, AHG8: An improvement on the compact OHP layout, in: Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET-E0056, Geneva, 2017.

[26] S. N. Akula, A. Singh, A. Dsouza, R. K. K, C. Pujara, R. N. Gadde, V. Zakharchenko, E. Alshina, K. P. Choi, AHG8 : Efficient frame packing for icosahedral projection, in: Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET-E0029, Geneva, 2017.

[27] Bringing pixels front and center in VR video, https://blog.google/products/google-vr/bringing-pixels-front-and-center-vr-video, accessed March 3, 2017.

[28] M. Coban, G. V. der Auwera, M. Karczewicz, AHG8: Adjusted cubemap projection for 360-degree video, in: Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET-F0025, Hobart, 2017.

[29] M. Yu, H. Lakshman, B. Girod, Content adaptive representations of omnidirectional videos for cinematic virtual reality, in: Proc. ACM Int. Workshop on Immersive Media Experiences, 2015, pp. 1–6.

[30] C. Zhang, Y. Lu, J. Li, Z. Wen, AHG8: Segmented sphere projection for 360-degree video, in: Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET-E0025, Geneva, 2017.

[31] K. Kammachi-Sreedhar, M. M. Hannuksela, AHG8: Additional test results of JVET-E0090 on nested polygonal chain packing of 360-degree ERP pictures, in: Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET-F0035, Hobart, 2017.

[32] Y. Wang, R. Wang, Z. Wang, K. Fan, A new panoramic video projection scheme., in: IEEE 1857.9 the 4th Meeting on Immersive Visual Content Coding, 1857.9-04-M0028, Guiyang, 2016.

[33] A. Abbas, D. Newman, AHG8: Rotated sphere projection for 360 video, in: Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET-F0036, Hobart, 2017.

[34] V. Zakharchenko, E. Alshina, K. Choi, C. Pujara, A. Dsouza, AHG8: Coding performance impact of omnidirectional projection rotation, in: Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET-E0050, Geneva, 2017.

[35] J. Boyce, Q. Xu, AHG8: Spherical rotation orientation SEI for coding of 360 video, in: Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET-E0075, Geneva, 2017.

[36] Next-generation video encoding techniques for 360 video and VR, https://code.facebook.com/posts/1126354007399553/next-generation-video-encoding, accessed March 3, 2017.

[37] G. V. der Auwera, M. Coban, Hendry, M. Karczewicz, AHG8: TSP evaluation with viewport-aware quality metric for 360 video, in: Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET-E0070, Geneva, 2017.

[38] F. Qian, L. Ji, B. Han, V. Gopalakrishnan, Optimizing 360 video delivery over cellular networks, in: Proc. Workshop on All Things Cellular: Operations, Applications and Challenges, 2016, pp. 1–6.

[39] K.-T. Ng, S.-C. Chan, H.-Y. Shum, Data compression and transmission aspects of panoramic videos, IEEE Trans. Circuits Syst. Video Technol. 15 (1) (2005) 82–95.

[40] H. Kimata, M. Isogai, H. Noto, M. Inoue, K. Fukazawa, N. Matsuura, Interactive panorama video distribution system, in: Proc. Technical Symposium at ITU Telecom World (ITU WT), 2011, pp. 45–50.

[41] P. R. Alface, J.-F. Macq, N. Verzijp, Interactive omnidirectional video delivery: A bandwidth-effective approach, Bell Labs Technical Journal 16 (4) (2012) 135–147.

[42] A. Zare, A. Aminlou, M. Hannuksela, M. Gabbouj, HEVC-compliant tile-based streaming of panoramic video for virtual reality applications, in: Proc. ACM Multimedia, 2016, pp. 601–605.

[43] D. Ochi, Y. Kunita, A. Kameda, A. Kojima, S. Iwaki, Live streaming system for omnidirectional video, in: Proc. IEEE Virtual Reality (VR), 2015.

[44] A. T. Nasrabadi, A. Mahzari, J. D. Beshay, R. Prakash, Adaptive 360-degree video streaming using layered video coding, in: Proc. IEEE Virtual Reality (VR), 2017, pp. 347–348.

[45] M. Hosseini, V. Swaminathan, Adaptive 360 VR video streaming: Divide and conquer, in: Proc. IEEE Int. Symp. Multimedia (ISM), 2016, pp. 107–110.

[46] M. Hosseini, V. Swaminathan, Adaptive 360 VR video streaming based on MPEG-DASH SRD, in: Proc. IEEE Int. Symp. Multimedia (ISM), 2016, pp. 407–408.

[47] T. Stockhammer, Dynamic adaptive streaming over HTTP–: standards and design principles, in: Proc. ACM Conf. Multimedia Syst., 2011, pp. 133–144.

[48] M. Graf, C. Timmerer, C. Mueller, Towards bandwidth efficient adaptive streaming of omnidirectional video over HTTP: Design, implementation, and evaluation, in: Proc. ACM Multimedia Syst. Conf., 2017, pp. 261–271.

[49] X. Corbillon, G. Simon, A. Devlic, J. Chakareski, Viewport-adaptive navigable 360-degree video delivery, in: Proc. IEEE Int. Conf. Communications (ICC), 2017, pp. 1–7.

[50] V. Zakharchenko, E. Alshina, A. Singh, A. Dsouza, AhG8: suggested testing procedure for 360-degree video, in: Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET-D0027, Chengdu, 2016.

[51] V. Zakharchenko, K. P. Choi, J. H. Park, Quality metric for spherical panoramic video, Proc. SPIE 9970 (2016) 99700C–99700C–9.

[52] V. Zakharchenko, K. P. Choi, E. Alshina, J. H. Park, Omnidirectional video quality metrics and evaluation process, in: Proc. Data Compression Conference (DCC), 2017, pp. 472–472.

[53] J. P. Snyder, Map projections–A working manual, US Government Printing Office, 1987.

[54] Y. Sun, A. Lu, L. Yu, AHG8: WS-PSNR for 360 video objective quality evaluation, in: Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET-D0040, Chengdu, 2016.

[55] Y. Sun, A. Lu, L. Yu, Weighted-to-spherically-uniform quality evaluation for omnidirectional video, IEEE Signal Processing Letters 24 (9) (2017) 1408–1412.

[56] C. Zhang, Y. Lu, J. Li, Z. Wen, AHG8: Padding method for segmented sphere projection, in: Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET-F0037, Hobart, 2017.