Visual Distortions in 360-degree Videos

Roberto G. de A. Azevedo, Neil Birkbeck, Francesca De Simone, Ivan Janatra, Balu Adsumilli, Pascal Frossard

***Abstract*—Omnidirectional (or 360-degree) images and videos are emergent signals being used in many areas such as robotics and virtual/augmented reality. In particular, for virtual reality, they allow an immersive experience in which the user can interactively navigate with 3-degree of freedom in the scene wearing a head-mounted display. Current approaches for capturing, processing, delivering, and displaying 360-degree content, however, present many open technical challenges and introduce several types of distortions in these visual signals. Some of the distortions are specific to the nature of 360-degree images, and often different from those encountered in the classical image communication framework. This paper provides a first comprehensive review of the most common visual distortions that alter 360-degree signals undergoing state of the art processing in common applications. While their impact on viewers' visual perception and the immersive experience at large is still unknown —thus, it stays an open research topic— this review serves the purpose of identifying the main causes of visual distortions in the end-to-end 360-degree content distribution pipeline. It is essential as a basis for benchmarking different processing techniques, allowing the effective design of new algorithms and applications. It is also necessary to the deployment of proper psychovisual studies to characterise the human perception of these new images in interactive and immersive applications.**

***Index Terms*—Omnidirectional video, 360-degree video, visual distortions, artifacts, compression.**

I. Introduction

FROM Virtual Reality (VR) to robotics, innovative applications exploiting omnidirectional images and videos are expected to become widespread in the near future. Fully omnidirectional cameras, able to capture a 360-degree real-world scene, have recently started to appear as commercial products and professional tools. User-generated and professional 360-degree content is already being distributed using popular content sharing platforms, such as YouTube and Facebook.

While the popularity of 360-degree content and applications is rapidly increasing, many technical challenges at different steps of the 360-degree signal acquisition, processing, and distribution chain remain open. The current approaches to process and distribute omnidirectional visual signals rely on algorithms and technologies designed for classical image and video signals captured by perspective cameras. The 360-degree imaging pipeline, however, has some particularities that induce specific distortions if not handled properly. First, the geometry of the content capture system is spherical, rather than planar [1]. To reuse existing file formats and algorithms designed for perspective signals, the spherical signal is warped into a planar representation [2], which is not a classical natural visual signal [3]. Second, the 360-degree content rendering, eventually via a Head Mounted Display (HMD), is characterised by an interactive and immersive dimension that represents a significant novelty [4]. Therefore, existing algorithms need to be adapted and optimized to process these signals efficiently and satisfy the new requirements.

In addition, 360-degree content distribution is expected to push the current storage and network capacities to their limits. Most HMDs currently on the market provide up to full High Definition display resolution. Since these devices usually provide a 110-degree of field-of-view, 4K resolution is widely accepted as a minimum functional resolution for the full 360-degree planar signal. Nevertheless, HMDs with 4K or 8K display resolution are already appearing on the market, which means that 360-degree planar signals with resolution of 12K or higher will have to be efficiently stored and transmitted soon [5]. Thus, new coding and transmission schemes, able to cope with increasingly high data rates and to satisfy user's visual quality expectations, will be continuously needed.

To improve existing omnidirectional processing pipelines and design new perceptually-optimised omnidirectional visual communications, it becomes critical to design tools to detect the *visual distortions* (or *artifacts*) introduced by each processing step, and, ultimately, quantify their impact on the perceived quality of the signal presented to the user and on the immersive experience [6].

Visual distortions occurring in images and videos captured by perspective cameras and undergoing compression and transmission have been largely characterized and analyzed in the literature, both for standard 2D [7], [8], [9], [10] and stereoscopic 3D signals [11], [12], [13], [14]. Nevertheless, *new types of distortions* can occur in 360-degree visual signals dataflows. These distortions have not been characterized in the literature yet, to the best of the authors' knowledge. In [15], a classification of the distortions caused by 360-degree content capture is presented, but the analysis does not include other processing steps such as coding, transmission, rendering, as well as the impact of the display technology. Some works reporting upon subjective studies in which users are asked to assess the overall quality of a set of processed 360-degree images or videos have recently appeared in the literature [16], [17], [18], [19], [20], [21], [22]. These works focus on the methodology used to collect user feedback and provide valuable guidelines to deploy classical subjective quality assessment experiments, but none of them analyzes the perceptual impact of 360-degree specific distortions.

This paper reviews and characterizes the most common visual distortions found in 360-degree signals undergoing state of the art end-to-end processing, including acquisition, lossy compression, transmission, and visualization by the end user.

R. Azevedo and P. Frossard are with Signal Processing Lab. (LTS4), École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland.

F. de Simone is with Centrum Wiskunde & Informatica (CWI), Amsterdam, The Netherlands.

N. Birkbeck, I. Janatra, and B. Adsumilli are with Youtube, Mountain View, California, USA.

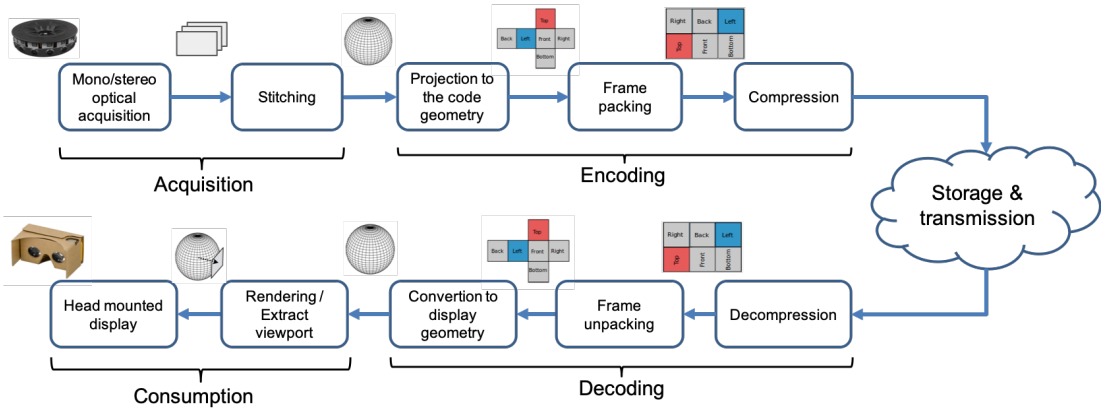


Fig. 1: End-to-end 360-degree video processing pipeline.

The goal is the isolation of individual distortions with the aim of obtaining a description of their visual manifestations, causes, and relationships. Visual examples are presented when possible[1]. This timely review serves as a tool for the benchmark of different processing algorithms and display devices, in terms of perceptual quality and will help in the deployment of psychovisual studies to characterize human perception of these new signals in new consumption scenarios. Moreover, being aware of the different visual artifacts and their causes is necessary for the development of more effective algorithms, that are able to properly cope with the specific nature of 360-degree images. Also, a brief overview of the existing tools used in state of the art to assess the quality of omnidirectional signals is presented, and perspective on future research directions are discussed.

The remainder of the paper is organized as follows. Section II presents the typical 360-degree video processing pipeline, used by most of the state-of-the-art approaches, and reviews some of the processing techniques currently in use in each step. Sections III–VI detail, for each step of the 360-degree pipeline, from acquisition to visualization, the most common artifacts and discuss their causes. Section VII discusses the current approaches, the open issues on the visual quality assessment of 360-degree videos, and how to use this comprehensive review to improve them. Finally, Section VIII brings our conclusions and points out future work.

II. 360-degree video processing pipeline

Fig. 1 depicts the end-to-end 360-degree signal processing pipeline considered in this paper, from acquisition to consumption via an HMD. Each step is briefly detailed hereafter.

A. Acquisition

Different optical systems have been proposed in the past to *capture* wide field of view signals [23], [24]. Nowadays, most of the commercial omnidirectional cameras with a full 360-degree field of view (e.g., the Ricoh Theta, the Gear360, and the Orah cameras) are multi-sensor systems, in which each sensor is a dioptric camera (sometimes with fish-eye lenses). These systems can be modeled as central cameras that project a point in the 3D space to a point on a spherical imaging surface, i.e., the viewing sphere [1]. In practice, the omnidirectional output signal is the result of a *mosaicking* (i.e., *stitching*) algorithm, specific to the acquisition systems, which merges the overlapping field of view signals acquired by all dioptric sensors to produce a wide-view *panorama* image [25].

In automatic *image stitching* processes, the overlapping regions between the cameras are *aligned* using different planar models (e.g., affine, perspective, or cubic transformation models); then, the views are *blended* and *warped* to the omnidirectional 3D surface, commonly a sphere surface [15]. For *video stitching*, additional *video synchronization* (if the individual sensors are not finely synchronized) and *video stabilization* (for moving cameras) may be needed [26], [27].

The output signal of the stitching process is usually stored, using standard file formats, as a rectangular array of samples (*planar representation*), resulting from the projection of the sphere to a plane (*map projection* or *spherical parametrization*) [28]. The planar representation allows re-using existing image and video content distribution chains, including encoders, packagers, and transmission protocols. Additionally, it is practical for rendering since hardware graphics systems need a simple arrangement of samples to access spherical images as a texture map. Most of the consumer-level omnidirectional cameras stitch to a planar representation referred to as *equirectangular panorama* (Fig. 2a). (Some professional-level cameras also allow to access the individual camera input, so that it is possible to use off-line software or manually fix possible stitching errors). The panorama uses an equirectangular projection (ERP) that maps a sphere to a plane by sampling the spherical signal on an equi-angular grid and using the longitude and latitude of each sample on the sphere as coordinates of the sample projected on the plane [28].

A few stereoscopic omnidirectional camera systems, able to capture the stereo views in all directions [29], have also been recently built as prototype [30] and professional capture systems —e.g., Facebook Surround 360, Jump [31], Obsidian [32]. They commonly output an Omni-Directional Stereo (ODS) representation that contains two modified ERP signals [33], corresponding to the left and right views for the

[1] The visual examples presented in this paper are mainly to illustrate the visual effect of the distortions. A true visual appreciation of the artifacts can only be provided by viewing the affected sequences on an HMD.

human eyes. Capturing stereoscopic omnidirectional dynamic scenes, however, is very challenging, since there is the inherent problem of self-occlusion among the cameras. A broad discussion on the different possibilities for the acquisition of both static and dynamic omnidirectional stereoscopic content is provided in [24].

B. Encoding

The goal of the encoding step is to reduce, in a lossless or lossy way, the redundancy in the signal, and thus the required space to store and transmit it. Most of the current 360-degree video systems re-use the same encoding tools as classical video solutions, such as H.264, H.265, VP9, or AV1 . One of the main challenges then resides in mapping the content into rectangular frames that are inputs for these video encoders.

A straightforward solution to encode 360-degree visual signals is to directly use the ERP (or ODS) signal output by an omnidirectional camera as input for any state-of-the-art encoder. Nevertheless, the equirectangular representation is not the most efficient representation for encoding. First, the regular sample distribution in the planar domain corresponds to a non-uniform sampling density on the sphere, with higher density towards the polar areas. Such a sample distribution is wasteful because, as have been demonstrated by subjective tests and head motion capturing study [34], [35], the content at the poles is usually not the most semantically interesting part of the scene being captured. Second, the ERP signal presents strong warping distortions towards the top and bottom image boundaries, which correspond to the polar areas in the spherical domain. Besides these geometric properties, the omnidirectional image signal has statistical characteristics which are not those of typical natural visual signals generated by perspective cameras, for which the encoding tools have been tuned for. By using the ERP representation in classical video encoders, the compression is therefore suboptimal.

Alternative planar representations that address both problems, by implying a more uniform sampling density in the spherical domain and being characterized by less strong warping distortions, have been proposed in the literature, such as *cube map* (CMP), *octahedron*, and tile-based projections [36]. Among those, CMP is the most common one. CMP is composed by the projection of the sphere in a circumscribed cube, resulting in six square cube faces (see Fig. 2b). Some studies have shown that using CMP can save up to 25% of the bitrate when compared to a similar user perceived quality in the ERP format [37]. Also, CMP is well-known in the computer graphics and gaming communities, and thus it is well-supported by graphics frameworks such as OpenGL [38].

As exemplified by the CMP format, some of the current projection methods result in different sets of faces, which then need to be packed together into one planar image (*frame packing* step of Fig. 1). For instance, a common packing method for CMP is the cubemap 3x2 arrangement, shown in the sample content of Fig. 2 (right image). Different frame packing methods may result in different discontinuities between the faces. A primary goal of the frame packing is to minimize the number of discontinuities in the planar representation.

Once the arrangement of faces has been completed, rectangular frames are constructed, possibly with additional padding, and eventually fed into classical video compression engines.

In the case of ODS content, the individual omnidirectional images for each eye are usually packed together in a frame-compatible stereo interleaving approach [39], e.g., through a top/bottom or side-by-side frame representation. In theory, other approaches that have been explored for standard stereoscopic 3D video —such as, simulcast, asymmetric coding, and multiview coding [39]— can also be adapted to the omnidirectional stereoscopic case. However, since such approaches are still underexplored for 360-degree content, they are not considered in the rest of the paper.

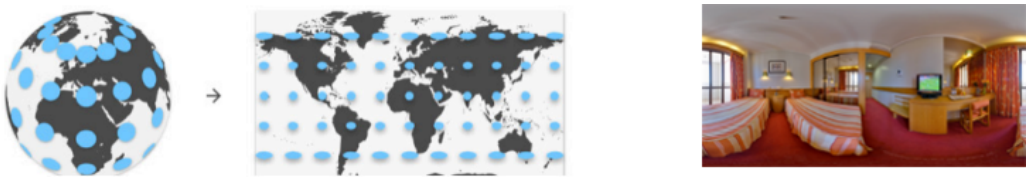
C. Transmission

In principle, since the encoding process results in traditionally compressed 2D (or ODS) frames, 360-degree delivery can use the same video streaming algorithms as classical image communication systems. Nevertheless, 360-degree content implies new technical challenges on content distribution due to the high data rate of omnidirectional signals and the low latency requirements of immersive communication. Also, unlike conventional video, the user does not look at the entire scene at once and can navigate around the content.

Nowadays, to reuse existing delivery architectures for video on demand and live streaming services, content delivery solutions relying on Dynamic Adaptive Streaming over HTTP (DASH) [40] are the most prominent ones to 360-degree video [41], [42], [43], [44], [45], [46], [47], [48], [49], [50], [51], [52], [53]. In these approaches, the server stores an adaptation set, i.e., a set of multiple versions (representations) of the same content, encoded at different bit-rates and resolutions. Each representation is temporally divided into consecutive segments of fixed duration, commonly ranging from 1s to 5s.

Since only a portion of the whole spherical video (the *viewport*) is displayed by the HMD at any given time instant, a number of *viewport-aware* streaming schemes have been recently devised to exploit this fact (in contrast to the *viewport-agnostic ones*, which handle the omnidirectional video as a conventional 2D video). In such an approach, a client is provided with different video representations, each one favoring a different viewport. The client should select and download some of the representations by taking into account not only the prediction of the available bandwidth but also a prediction of the user's navigation pattern. Nowadays, two main variations of viewport-aware dynamic adaptive streaming are being explored: *viewport-dependent projection* and *tile-based streaming*.

In the *viewport-dependent projection* approach [43], [45], [50], [53] a certain area may be favored in the planar representation using different projections. It is possible to have different viewport-dependent quality representations, each one favoring a specific viewport of the content. Thus, the client can choose the optimal viewing quality approach by selecting the projection representation which provides the best representation of the current user's viewport.



(a) Equirectangular projection (ERP). ERP maps meridians to equally spaced vertical straight lines, and parallels to equally spaced horizontal lines. In this projection, poles regions are stretched compared to the equator region.

(b) Cubemap projection (CMP). CMP is performed by projecting the sphere signal to a circumscribed cube. It results in six faces that need to be rearranged to form a rectangular frame.

Fig. 2: Examples of map projections.

In the *tile-based* approach [41], [44], [46], [47], [48], [49], [51], [52], [54] the planar omnidirectional video is decomposed into independently decodable rectangular parts, i.e., tiles, so that each tile is encoded at different quality levels. The client can then choose to download only the tiles contained in the current user's viewport with high quality while downloading the non-visible ones with lower quality, or even ignoring them.

D. Consumption

At the client side, the inverse steps —*decoding*, *unpacking*, *conversion to display geometry* and *viewport extraction* (or *rendering*)— need to be performed, so that the user can visualize and interact with the 360-degree video content. When the content is rendered, the inverse mapping from the plane to the sphere is performed. The viewer is centered on the sphere and is able to navigate the content by changing their viewing direction: the visible portion of the sphere surface at each instant is projected to the *viewport*, depending on the user's viewing direction. This rendering is typically implemented on a Head-Mounted Display (HMD), or on more classical devices such as a computer or smartphone. With HMDs, the users can easily navigate the scene by turning their head freely. In desktops and smartphones, the users can consume the 360-degree content through a "magic window", in which they can interact with the content using a mouse (or another device) in a desktop, or by moving the position of the smartphone in the physical space. Given the new immersive features and challenging of HMD-based approaches (i.e., it can introduce new distortions still not fully understood), this paper focuses mainly on the HMD-based consumption approaches.

III. Artifacts caused by Acquisition

As previously mentioned, capturing 360-degree content is usually composed of two main steps: acquiring the visual content through a multicamera optical system and then stitching the multiple images into one global signal, generally in the form of a spherical image. Each of these steps may add visual distortions, which are discussed in what follows.

A. Sensor limitations

Each of the cameras of a 360-degree multicamera rig is subject to common *optical distortion* —e.g., *barrel*, *pincushion distortions*, and *chromatic aberrations*— *moiré effect*, *noise*, and *motion blur* [12]. In particular, wide-angle fisheye cameras, commonly used in the multicamera rigs, are prone to *chromatic aberrations*, more than regular perspective cameras. In addition, wide aperture angles on fish-eye cameras are only possible with large amounts of *barrel distortion*.

Additional artifacts may also occur due to inconsistencies between the cameras. For instance, *exposure artifacts* —i.e., very different brightness between adjacent cameras— may appear, and the lack of synchronization among the cameras may result in *motion discontinuities*. If those issues are not handled properly by the video stitching step, they will ultimately impact the overall pipeline and be perceived by the end user.

Omnidirectional stereoscopic 3D content capture is subject to typical distortions of standard stereoscopic 3D content, such as *keystone distortion*, *depth field curvature*, and *cardboard effect* [12], [14]. *Keystone distortion* is the result of the position of the two cameras (for left and right eyes) converging to slightly different planes, which causes a vertical parallax, i.e., a vertical difference between homologous points [14]. The same principle in the horizontal direction leads to the *depth plane curvature* artifact. The *cardboard effect* refers to an unnatural flattening of objects in stereoscopic images–affected objects appear as if they were cardboard cut outs [12].

Moreover, compared to capturing stereoscopic 3D content for cinema and TV, capturing omnidirectional stereoscopic 3D content adds up their own set of challenges [32]. For instance, the optical centers of individual cameras do not share the same center of projection. However, applying planar transformation models to synthesize multiple views together on a common virtual surface is only valid if the captured scene is a planar surface itself, or if the cameras share the same center of projection [15]. For off-centered cameras, transformation error increases with the off-center distance and the amount of depth

within the captured scene. The warping and stitching process can finally worsen *keystone* and *depth field curvature* issues.

B. Stitching issues

The unreliable information due to optical distortions and motion discontinuities between the different cameras usually makes the stitching process very challenging. (Indeed, given its complexity image and video stitching has been an active research area [25].) Besides combining and warping the individual images to create the spherical signal, the video stitching process may also have to compensate for some of the sensor limitations and inconsistencies among the cameras in the multicamera rig. Due to these challenges, most approaches are still affected by visually annoying artifacts, which may appear as *blurring*, *visible seams* (due to different exposures to color and brightness discontinuities), *ghosting*, *misaligned/broken edges and image structures*, *missing information* (e.g., objects with missing parts), and *geometrical distortions* (e.g., visible deformation of objects). Fig. 3 exemplifies some of these artifacts. Since some stitching algorithms estimate depth information by multiview geometry estimation, objects close to the cameras tend to be more affected by stitching errors, because the depth errors are more significant in these areas.

Another artifact that may be present in 360-degree content is the *black circle* or *blurred circular areas* on poles (see Fig. 4). Such an artifact is mainly due to: the multicamera rig being unable to capture the full 360-degree field of view; or post-processing to remove the camera stand of the scene through inpainting techniques. Note that it is also quite common to replace this pole area with the name of the camera brand or alternative information.

In video settings, it may further be possible to see *motion discontinuities*, such as *object parts appearing and disappearing* abruptly and *temporal geometrical distortions*, when moving objects come across the stitching areas. Moreover, for some camera rigs, if the video stitching algorithm does not successfully treat the lack of synchronization among individual cameras, *synchronization issues* may be perceived by the end user. For instance, temporal inconsistent stitching of the camera views can result in unsteady scene appearance over time on the stitching areas, resulting in *wobbling artifacts*.

Interestingly, the regions affected by stitching artifacts are generally known for a given camera rig. Thus, from a cinematic point of view, it is usually a good practice to have less action and useful information in this area.

Finally, compared to monoscopic 360-degree video, capturing omnidirectional stereoscopic video usually increases the amount of stitching and blending errors. On the one hand, to reduce stitching errors, the baseline between the cameras should be minimized. On the other hand, the baseline between the cameras of different views needs to be increased in S3D content creation as parallax is required for generating a 3D effect [15]. In addition, minor flaws in the footage or stitching errors are usually magnified when viewed in stereoscopic 3D. Given that such errors can occur in different places in each view, this can result in *binocular rivalry* and *discomfort* when watching stereoscopic 360-degree content.

IV. Artifacts caused by Encoding

A. Projection to the coding geometry

Projecting a sphere to a plane is a common problem in map projections [28], and it is impossible to do so without adding some *geometrical distortions* and *discontinuities* on the planar presentation —i.e., neighboring regions on the spherical domain may end up not being neighbors on the planar domain, or vice-versa. Different projections may imply different *geometrical distortions* and *discontinuities* regions. Fig. 5 shows examples of geometrical distortions and discontinuities resulted from ERP and CMP. Even though these distortions are not meant to be directly viewed by the end user, their interaction with the lossy compression processing may result in visible artifacts.

Moreover, since the projection from the spherical to the planar representation (and the back-projection to the spherical domain at the client side) involves some resampling and interpolation, different map projections may result in *aliasing*, *blurring*, and *ringing* distortions (see Fig. 6a).

Also, if sampling and interpolation are not treated correctly additional distortions may happen, such as: *visible poles* due to oversampling on the poles areas, may appear when using the ERP representation (see Fig. 6b); and *visible seams* in the discontinuities regions (see Fig. 6c). Methods like graph-based techniques [55] that are well adapted to the specific geometry of images could reduce such artifacts by processing the data in their native geometry. However, current 360-degree systems exclusively rely on sampling and interpolation techniques in the classical rectangular geometry.

B. Compression

Since the current approaches use conventional 2D video compression schemes for the planar representation, they are also subject to the same artifacts thoroughly studied in 2D video, which are briefly presented in Table I. For a more in-depth discussion on 2D video artifacts, we refer the reader to [10], [9], [8], [7]. Mostly, the origins of the artifacts in lossy block-based transform video coding are (directly or indirectly) due to quantization errors in the transform domain [9].

With its particular geometry, the omnidirectional video is generally affected by a complex combination of the compression artifacts that affect the rectangular frames, as well as the frame packing and the warping due to the map projection used. For example, the blocking artifacts produced by the compression in the planar domain will also be warped due to the omnnidirectional geometry. Thus, they might be perceived as different *warped blocking patterns*, which depend on the underlying geometry of the map projection. For instance, in the ERP representation, blockiness close to the poles might be perceived as a *blocking radial pattern* (see Fig.7b). Similarly, for the CMP representation, it may be possible to see the *perspective projection of the blocking artifacts*, and eventually *identify the underlying cube faces*.

As previously mentioned, inevitably, when using a 2D rectangular image to represent the full 360-degree spherical signal, some neighboring regions in the spherical domain are not neighboring in the planar representation. Thus, when

(a) (b) (c) (d) (e)

Fig. 3: Examples of *stitching* artifacts: (a) broken edges and (b) missing information; *blending* artifacts: (c) ghosting and (d) exposure; and *warping* artifacts: (e) geometrical distortions / object deformations.

(a) (b)

Fig. 4: Examples of post-processing on the poles due to missing areas: (a) a black circle and (b) a (inpainted) blurred circle.

such a region is coded in the planar domain (without taking into account the original neighbors) and projected back to the spherical domain, *discontinuities* or *visible seams* can appear. For instance, when using ERP, a seam can appear in the region closing the sphere, whereas unnatural seams can appear on some cube edges when a CMP representation is used instead (see Fig. 5). The origins of those visible seams due to compression can also be traced back to: (1) transform blocks falling between two faces in the planar representation; (2) color bleeding or ringing artifacts from one face to the other; and (3) deblocking filter algorithms that mismatch the faces discontinuities as blocking artifacts, and thus may smear content from one face to another [56]. In both cases, some data from one face bleeds to the neighboring one in the planar domain. Since those two faces are not neighbors on the spherical domain, seams may become visible. Due to changes in the properties of the *visible seams* during consecutive frames, it is also possible to end up with *flickering seams* in the temporal domain.

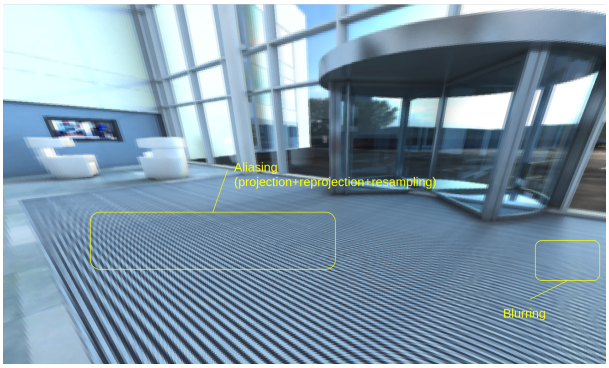
On a coarse quantized lossy compressed video, the appearance of *blocking*, *blurring*, *staircase* and *basis pattern* in combination with the warping and frame arrangement on the planar representation may also result in visible *spatial pattern transitions* on the viewports. This is mainly the result of the different compression distortions being applied in different directions in the planar domain, when compared to the viewports. This is the case, for instance, in CMP, where each different face may undergo different geometrical distortions and rotations, which may cause visually noticeable texture area changing its underlying "pattern" across adjacent CMP

(a) (b)

Fig. 5: Examples of discontinuities implied by the (a) ERP and (b) CMP representations. In (a) the vertical borders of the ERP representation are not neighbors on the planar representation, but they are neighbors on the spherical domain. In (b) discontinuities also happen on the borders of the frame, but, in addition, due to *frame packing* some faces which are not neighbors on the spherical domain, became neighbors on the planar representation.

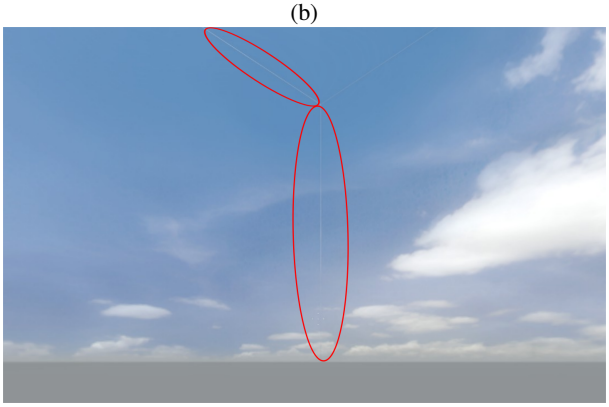
faces (see Fig. 7a). In the temporal dimension, if an object is crossing from one face to the other, it may also be possible to see dynamic changes on its underlying "pattern".

The use of compressors unaware of the geometry of omnidirectional videos also results in the content being more prone to motion compensation and flickering issues than the classical video counterparts. Most modern video codecs use block-based motion estimation for inter-frame compression —i.e., a block of pixels is matched to neighboring frames (and usually from blocks on neighboring areas, to speed things up), and if



(a)

(b)



(c)

Fig. 6: Examples of (a) aliasing and blurring, (b) visible poles; and (c) visible seams due to projection and resampling.

there is a good match, a 2D offset vector (smaller than a block) is calculated and stored instead of the original block. Indeed, if blocks are small enough, block vectors can represent general planar motion with perspective cameras rotating in all 3 axes. However, the planar representation of the 360-degree content implies that at some parts the motion is no longer planar and vectors cannot be predicted so well from neighbours. Thus, the motion model and intra-prediction are not optimal in regions such as the poles on ERP and in discontinuities in CMP, which may result in higher bitrates and compression artifacts, such as motion flickering, in these areas. [57]

Finally, in stereoscopic settings, the compressed ODS video is also subject to the same artifacts that have been studied in the context of stereoscopic 3D video [12], [14]. One of the leading sources of compression-related stereoscopic artifacts is the possibility of the compression algorithm to introduce different distortions on the left and right frames, resulting in *cross distortion*, which may affect depth perception and cause binocular rivalry. Visible seams artifacts, which are specific of 360-degree content, may also be affected by cross-distortions, and may result in a volumetric perception of the seams. When using a frame-compatible approach for the ODS content, the discontinuities between the left and right content may also result in the appearance of new seams Asymmetric stereoscopic 3D spatial resolution and compression [58] is also another potential source for cross distortions. The *cardboard effect* may also be introduced by compression.

TABLE I: Summary of conventional 2D video compression artifacts.

Artifact	Characteristics
Blocking	is related to the appearance of the division of the macroblocks; it is caused by coarse quantization of low-detail regions.
Blurring	is the result of loss of spatial details in moderate-high-detail regions; it occurs when high-frequency components in the transform domain are quantized to zero or due to strong deblocking filters.
Color bleeding	is the smearing of colors between areas of strongly contrasting luminance; it happens due to inconsistent image rendering on separately compressed color channels or due to interpolation on chroma-subsampled images/videos.
Ringing	appears as "halos" (artificial wave-like or ripple structure) around sharp edges, e.g., strong edges and lines.
Stair case and *basis pattern*	incapability of horizontal and vertical basis functions (as building blocking of the DCT and its variations) to accurately represent diagonal edges (similar to steep edges).
Flickering	refers to frequent changes in luminance or chrominance along the temporal dimension that do not appear in uncompressed video, and can be divided into *mosquito noise* (when it occurs at the borders of moving objects), *coarse-granularity flickering* (when it suddenly occurs in large spatial areas) and *fine-granularity flickering* (when it appears to be flashing on a frame-by-frame basis) [10].
Jerkiness	occurs when the temporal resolution is not high enough to catch up with the speed of moving objects, and thus the object motion appears to be discontinuous.
Floating	is the appearance of illusive motion in certain regions as opposed to their surrounding background; the illusive motion is erroneous because these regions are supposed to stay or move together with the background.

V. Transmission-related artifacts

Transmission delays and communication losses affect the streaming of omnidirectional video sequences, similarly to how they affect traditional videos. As the recent streaming systems are based on adaptive streaming algorithms, we focus on their specific artifacts in the following.

Depending on the streaming scheme (*viewport-agnostic*, *viewport-dependent projection*, or *tile-based*) in use and on the implemented adaptation logic for the 360-degree content, different distortions can appear and impact the user experience.

In the *viewport-agnostic* adaptive streaming, the typical DASH distortions, such as *delay*, *rebuffering events*, and *quality fluctuation*, may be perceived in the user's field of view. These distortions have been widely studied and characterized for conventional 2D video content and displays [59], [60]

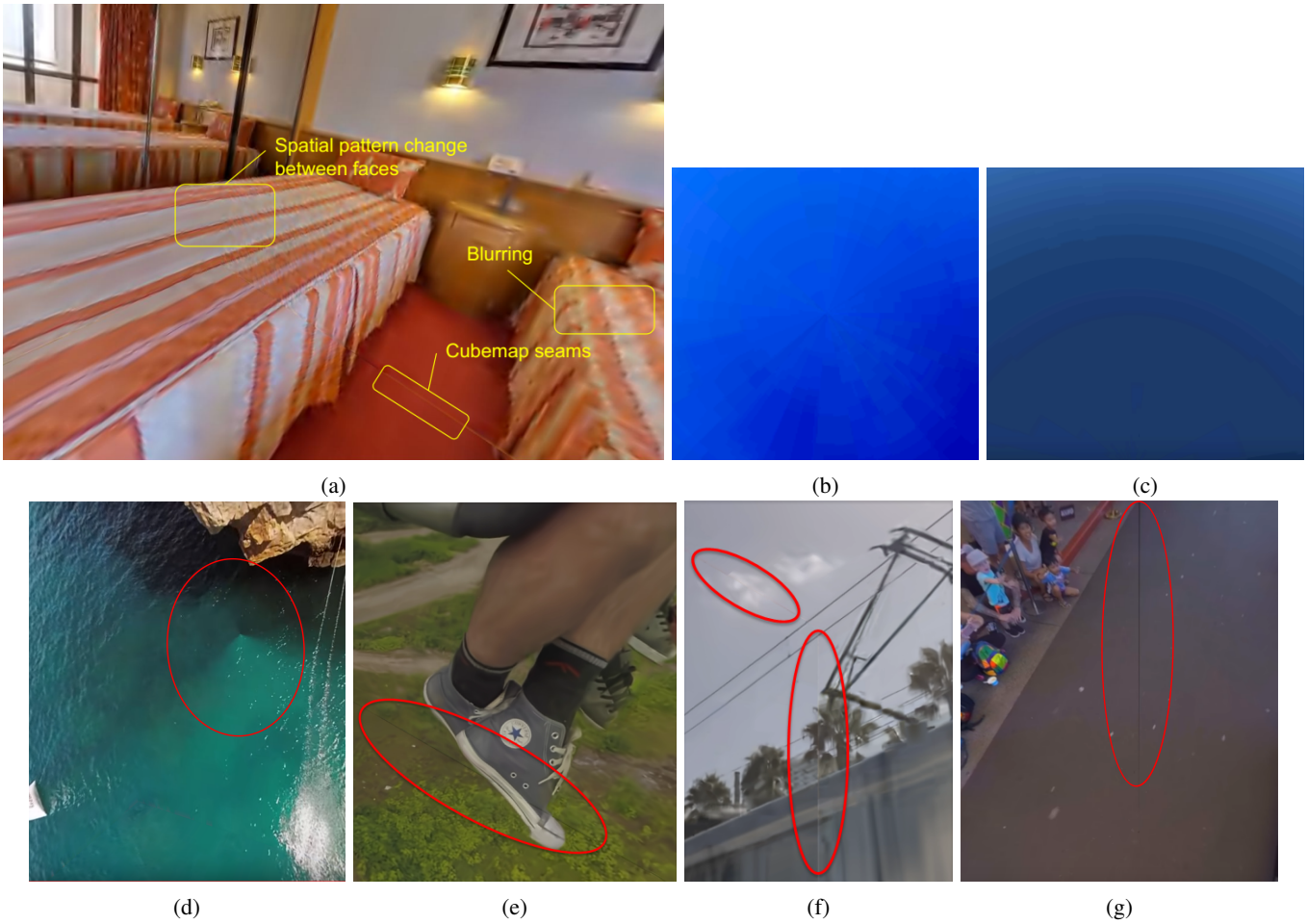


Fig. 7: Examples of compression artifacts on 360-degree content: (a) blurring, cubemap seams, and spatial pattern changes (due to jpeg2000 compression); (b) radial blocking; (c) radial banding; (d) visible pole; (e)-(f) cubemap seams artifacts; and (g) equirectangular seam artifact.

and few studies have investigated them in the context of stereoscopic 3D video [61]. The impact of these distortions on the Quality of Experience (QoE) of immersive applications when the compressed content is projected to the viewport and viewed through an HMD is still largely overlooked [62].

In the *viewport-aware* adaptive streaming schemes, the ability of the system to predict how the user navigates the content can impact the artifacts perceptible by the end user. Besides the typical DASH-based distortions, new artifacts may appear.

In the *viewport-dependent projection* approach, besides the temporal quality fluctuations in the field of view, there is also the possibility of the user experiencing *quality fluctuations* and *rebuffering events on head movement*. Also, when the viewport is composed of regions with different qualities, *spatial quality fluctuations in the viewport* may become annoying.

The *tile-based* approach is also subject to *spatial qualities fluctuations* (when the viewport is composed by tiles of different qualities). In addition, the *tile borders* may become visible and, in extreme cases, a portion of the viewport could be missing (*incomplete viewport*) if the client has not downloaded the corresponding tile.

From the above visual artifacts, *spatial quality fluctuations* in both viewport-based adaptive streaming can also impact the two views differently, leading to stereoscopic 3D artifacts.

Fig. 8: Example of *spatial quality fluctuation* artifacts due to tiling (highlighting the tile borders). (Adapted from [52])

VI. Artifacts from Displays

Even with a perfectly captured, transmitted, and received mono or stereo omnidirectional image, artifacts still can appear due to technical limitations of the current displays. Among the common 360-degree visualization techniques, the HMD mode is the most challenging one. Indeed, all the artifacts of traditional displays, such as *aliasing*, *blurring*, *motion blur*, etc., may also affect HMDs. In addition, new distortions that are specific to HMDs can appear because, compared with traditional displays, the HMD is very close to the user's eyes, it has a wider field of view, and, more importantly, it physically moves with the user's head. Such an interaction between a user's and display movement is unique to HMDs, and it can cause new artifacts that have not been considered in traditional displays, and that can even break the sense of presence or, worse, they can make the user physically uncomfortable.

Designed for supporting an immersive visual experience, most HMDs are composed of a display device attached to the head (and providing stereoscopic vision) and an optical and a head-tracking system. The purpose of the optical system (see Fig. 9) is twofold. First, the close distance between the user's eyes and the HMDs requires an optical system to support comfortable content viewing. Second, it serves to optically magnify the content presented on the screen, supporting a Field of View (FoV) closer to the natural human viewing. The head tracking system allows the system to update the content presented to the user based on his head position.

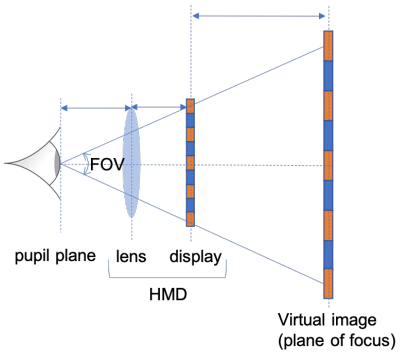


Fig. 9: Simplified schematics of HMD

Different optical design systems have been used with the goal of supporting a larger FoV and comfortable viewing on HMDs. These systems have varying trade-offs in weight, field of view, light transmittance, and image quality [63], but they all suffer from *optical distortions*. For instance, some lenses can cause *chromatic aberrations* at the edge of the FoV. Currently, higher-quality HMD displays, such as Oculus and HTC, have changed the design to incorporate fresnel lens [64] features. Although these new lenses improve on the rectification of the *chromatic aberrations*, they bring another problem sometimes referred to as *"god-rays"* or *"flare"*, which is characterized by the appearance of a halo at the FoV's edges. This is mainly due to the light that is falsely redirected through the fresnel steps.

On both of the aforementioned lenses types, the magnification characteristics of HMDs is done by applying a significant *pincushion* distortion through the lenses. Such a distortion must be rectified by applying a distortion in the other direction, usually a barrel-distortion shader toward the end of the rendering process. The required amount of distortion is display-specific, and if it is not done properly it may also result in a *barrel* or *pincushion* distortion perceived by the end user. In both *chromatic aberrations* and *geometrical distortion*, shaders can be used to try to mitigate the visible effects. [65]

Then, when watching 360-degree video on most of the current HMDs, it is possible to see a fixed lattice pattern (such as the one shown on Fig. 10) named the *screen-door effect*. Such a pattern mainly occurs because having the screen very close to viewers eyes as in an HMD, it is actually possible to see the spacing between the pixels. The screen-door effect is certainly not a new phenomenon but has been mostly solved for the viewing distance of current digital TVs and projectors. For current HMD displays, this is still an issue, and it may be solved in the coming years with higher resolutions.

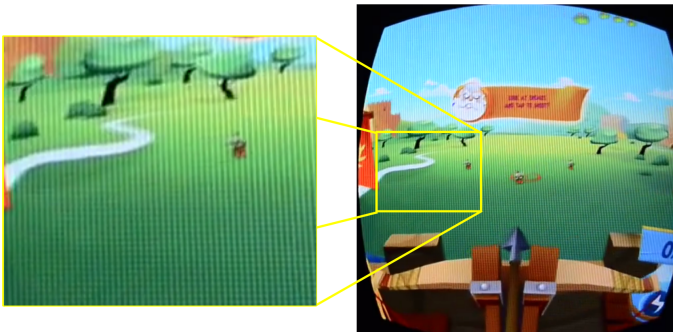


Fig. 10: Example of the screen-door effect. [2]

Motion-to-photon delay is another artifact that is specific to HMDs. It is defined as the time perceived by the end-user between his movement and the full response on the display screen [66]. Despite being an annoying artifact, motion-to-photon delays may also induce motion sickness. Ideally, to achieve a full sense of presence, no motion-to-photon delay should be perceived.

While motion-to-photon is a well-studied phenomenon in VR, and current high-quality displays have been improving in this area, another phenomena, named *smearing*, related to the pixel's persistence has become more visible. (Fig. 11 shows an example of how smear is perceived.) Smearing is caused by an intrinsic interaction between pixel persistence on a moving display and the Vestibulo-Ocular Reflex (VOR). When focusing on one object and rotating our head, the eyes counter-rotate this movement due to the VOR to keep the image of the object focused [67].

Finally, most of the common stereoscopic displays-related distortions are still present in HMDs. For instance, since the current commercially available displays do not provide eye-tracking technologies accommodation-convergency rivalry is still a problem for HMDs, and can create several problems

[2]Frame extracted from https://youtu.be/V7uTnOYLhZA

[3]Image adapted from: http://blogs.valvesoftware.com/abrash/why-virtual-isnt-real-to-your-brain-judder/

(a) (b)

Fig. 11: Example of smear from persistency. (a) A rendered scene as seen without head movement. (b) The same scene as perceived by a moving head (smeared by 2deg). [3]

such as eyestrain, blurred image, and misperception of distance, size, depth, or speed of objects [63].

VII. Measuring 360-degree content quality

The visual quality of classical images and videos is generally measured by a global quality index that (ideally) integrates all the possible sources of distortion into a single or a few values. However, as aforementioned, the sources of distortions in 360-degree videos are numerous and quite different, and their combination into a global index is far from trivial.

Table II summarizes the different types of distortions commonly found in 360-degree video. In the table, we broadly categorize the artifacts into four categories: spatial, temporal, stereoscopic, and navigation. *Spatial artifacts* are those related to still image compression and can appear in both images and videos. *Temporal artifacts* are those related to the temporal evolution of images and appear only on video. *Stereoscopic artifacts* are those related to binocular vision. *Navigation artifacts* are those that only appear while the user navigates through the scene.

The ultimate way to assess the 360-degree visual quality is through *subjective* tests, which can shed light on the way the different distortions interact together. Such tests, however, are time consuming and expensive. Thus, *objective* metrics have been proposed for omnidirectional video in the past few years. However, it is quite challenging to capture all the effects that impact the QoE of 360-degree videos, and much work remains to be done in this area, in particular, with regards to perceptually optimized metrics. The rest of this section presents some of the current approaches for both objective and subjective quality assessment of 360-degree content and discusses some of the open research challenges.

A. Objective metrics

Currently, the main approaches for objectively assessing the quality of 360-degree content are: (1) the use of well-known objective metrics for 2D content computed on the planar domain; (2) the use of well-known objective metrics for 2D content computed on the viewports; (3) objective metrics specifically developed for 360-degree content.

The use of standard 2D image and video metrics, e.g., PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity), directly in the planar domain is straightforward, but they do not properly model the perceived quality of the 360-degree content. For instance: (1) they give the same importance to the different parts of the spherical signal, which besides being sampled very different from classical images, also have different viewing probabilities (and then different importance); (2) even for traditional images, most of these metrics are known for not performing very good at representing the subjective quality. The use of more perceptually-optimized metrics [68] on the planar domain, however, is not straightforward, and new metrics specifically developed for the 360-degree content must be developed.

The use of well-known 2D objective metrics computed on the viewports is an interesting approach, in which, N viewports, for different viewing directions, are generated for both the original and the distorted content, and the 2D metric is computed individually for each of these viewports. Then, the overall 360-degree quality metric can be computed by aggregating the individual viewports. If the used 2D metric properly models the human perceptibility, in theory, it could be a good approximation of the overall 360-degree. Viewport-based metrics using PSNR/SSIM have been discussed [36]; but probably more perceptually-optimized metrics [68] could better model approximate the 360-degree quality. In any case, how to choose which viewports should be computed (which, in theory, could be arbitrarily large) and how they should be pooled on a unique score is far from trivial.

Spherical-PSNR (S-PSNR) [69], Craster Parabolic Projection PSNR (CPP-PSNR) [70], Weighted-to-Spherical-PSNR (WS-PSNR) [71], and S-SSIM [72] are some of the first attempts to objectively measure the quality of 360-degree images. In S-PSNR, sampling points uniformly distributed on a spherical surface are re-projected to the original and distorted images respectively to find the corresponding pixels, followed by the PSNR calculation. In CPP-PSNR, the PNSR is computed between samples in the Craster parabolic projection (CPP) domain [70], in which pixel distribution is close to that in the spherical domain. The pixels of the original and distorted content are first projected to the spherical domain and then mapped to a CPP domain, where the PNSR is computed. In WS-PSNR, the PSNR computation at each sample position is performed directly in the planar domain, but its value is weighted by the area on the sphere covered by the given sample position. Different weight patterns may be used for different projections. S-SSIM is a similar approach, but using SSIM instead of PSNR [72].

All the above objective metrics, however, fail in *properly considering the perceptual artifacts* in a 360-degree processing chain, as discussed herein. For instance, the *visible seams* artifacts due to compression —which is usually easy to perceive— may be hidden in current full-frame objective metrics, because the samples along the seams are only a small percentage of the samples in the frame or viewport [73]. Thus, objective metrics that detect individual artifact reliably and efficiently, and that can build on the perceptual features of these artifacts are still necessary. This paper contribution is one step in this direction.

Finally, another critical issue today towards the development of perceptually-optimized objective metrics is the *lack of a common quality 360-degree dataset* (for both monoscopic and stereoscopic content) to be used for various dimensions including processing (fusing, stitching, editing), encoding, delivery,

TABLE II: Summary of the visual distortions in 360-degree content.

	Capturing	Encoding	Transmission	Display
Spatial	Optical distortions (individual cameras): • blurring by defocus • barrel distortions • pincushion distortions • mustache distortions • noise • chromatic aberrations Stitching artifacts: • discontinuities (e.g., misaligned/broken edges: • missing objects parts • exposure artifacts • black circle / blurred circle Blending artifacts: • Visible color- and luminance- mismatches of regions withing an ODI. • exposure artifacts • visible seams due to color- and luminance- mismatches • ghosting / duplicated objects Warping artifacts: • geometrical distortions / visible deformation of objects	Projection: • geometrical distortions • aliasing – circular pattern aliasing • blurring • ringing • radial pattern close to the poles (due to oversampling on ERP) • visible seams (due to sampling) Compression: • blocking – mosaicing effect – staircase effect – false edges – warped blocking artifacts ∗ radial blocking pattern (erp) ∗ perspective projected blocks (cmp) ∗ identifiable underlying 3D geometry • blurring • ringing • basis pattern effect • color bleeding • visible seams, due to – color bleeding – crossing faces block transform – loop filter • spatial pattern changes between faces (cmp)	Channel distortions: • data loss • data distortion Viewport-aware adaptive streaming: • spatial quality fluctuation • tiling artifacts	Rendering: • aliasing • blurring • ringing Display limitations: • Optical distortions – pincushion distortions – chromatic aberrations – god-rays or flare – • screen-door effect
Temporal	• motion blur • channel mismatch • motion discontinuities – appearing / disappearing objects – dynamic geometrical distortions – dynamic ghosting – wobbling artifacts	Compression: • flickering – mosquito noise – fine-granularity flickering – coarse-granularity flickering • jerkiness • floating – texture floating – edge neighborhood floating • flickering seams • spatial pattern changes when crossing the faces	• delay • video freezing • quality fluctuations •	
Stereoscopy	• depth plane curvature • keystone distortion • cardboard effect	Projection: • ghosting (caused by disocclusion) Compression: • cross-distortions • cardboard effect	Channel distortions: • data loss • data distortion (binocular) Viewport-aware adaptive streaming: • cross distortions (due to spatial quality fluctuations)	Display limitations: • crosstalk as inter-perspective aliasing and ghosting • viewing dependent binocular aliasing • accommodation / convergence rilvary • lattice artifacts
Navigation / Head movement			Viewport-aware adaptive streaming: • video freezing • spatial quality fluctuation • quality fluctuation Tile-based viewport-aware adaptive streaming: • spatial quality fluctuations • tile borders • incomplete viewport	Display limitations: • motion-to-photon delay • smear from persistence

and rendering/consumption. It is not clear, for instance, how the few available datasets [74], [75], [20], [76] cover the visual distortions introduced by state-of-the-art 360-degree pipelines, and thus how they can be effectively used as benchmark for perceptual-based quality metrics and the design of optimized processing algorithms. The contribution of this paper can also help in analyzing the current datasets and on the development of new, more perceptually relevant, ones.

B. Subjective studies

The lack of standard quality 360-degree datasets is also due to the lack of standardized methodologies for the subjective quality assessment of 360-degree content, which is still in active debate in the research community. For instance, through the Immersive Media Group (IMG) [4], the Video Quality Expert Group (VQEG) is actively pursuing the development and standardization of methodologies for the subjective assessment of 360-degree visual content. Currently, however, the research community still did not reach a consensus on the best practices for subjective assessment of 360-degree content.

Some recent efforts have been made in adapting subjective methodologies from classical image/video quality assessment to 360-degree content. Initial tests have been performed on viewing the rendered viewports on traditional displays [77], [78], while others have been performed using HMDs [69], [79]. On the one hand, visualizing the viewports on standard displays lacks the important immersive features that can only be assessed when the user is wearing an HMD. On the other hand, the adaptation of traditional subjective methods for the immersive viewing through HMDs is far from trivial because it needs to take into account at least that: there are important differences in displays (e.g., increased FoV and magnification of the content); the user is immersed in the content; and that the content can both induce the sense of presence and sickness.

As discussed in Section VI, the different displays specifications, e.g., resolution, and supported FoV, may have a direct impact on the visual quality perceived on subjective studies. Indeed, as previously mentioned, different HMD lenses may change how the spatial display resolution is perceived and introduce different artifacts. Thus, it is critical that during the subjective experiments, researchers specify both the displays and the rendering/adaptation of the content to the specific display. Moreover, there is still a lack of cross-device studies that allow researchers to better understand the impact of the display features on the overall quality assessment.

The fact the user is immersed in the content, and free to navigate with 3DoF in the content completely changes the QoE perspective when compared to classical subjective studies. First, by only looking at a fraction of the captured scene at a given time, the user may not perceive an artifact if he is not looking to the "right place", and some quality issues may go unnoticed. Since different people might look at different parts of the content, *visual attention* and *salient regions* are more important on the subjective quality assessment of 360-degree content. Some studies have been considering such importance and datasets have been proposed to develop these ideas. Currently, such data have been used mainly for improvements on streaming, but they can (and should) also be used to improve quality metrics. Second, the ideal viewing sequence duration is not necessarily the same being standardized for traditional methods, since the user may need some time to adapt and understand where he is in the content.

Finally, all the subjective tests performed to the date, including the ones using HMDs [69], [79], focus on the overall signal quality, and do not provide insights on the impact of specific artifacts and, e.g., how they might cause the user to lose the sense of presence (immersion-breaking artifacts). A better understanding of the impacts of the perceptibility of individual artifacts and its impacts on user's QoE will only be possible by performing *psychophysical visual studies* [80] specifically designed for these artifacts, which are still to appear in the scope of 360-degree content consumed through HMDs. We expect that new studies for the other artifacts presented in this paper will start to appear soon in the literature.

C. Beyond visual quality

It is also important to highlight that visual quality alone is not enough for measuring QoE in VR. VR is much broader than just the visual experience, and for a complete VR quality framework, besides measuring the visual quality, it is also necessary to quantify other parameters that have not been discussed in this study. For example, VR Audio, HMD ergonomics (e.g., weight, weight balance, pressure, fit and finish, temperature, and overall hygiene) [81], [82], user discomfort, and usability are all important factors in defining a global VR quality of experience.

In this paper, we have been mainly concerned visual distortions on current monoscopic and stereoscopic 360-degree images and videos, which allows for a 3DoF experience. New approaches based on multiple 360-degree views, point clouds, and volumetric videos with potential to support both 3DoF+ and 6DoF are also expected to appear in the future, and they bring their own issues for visual quality assessment.

VIII. Conclusion

By reviewing and characterizing the common artifacts in state-of-the-art end-to-end 360-degree video workflows, this paper contribution is an important step towards the design of more effective algorithms, applications, and in the development of perceptual-based quality metrics for 360-degree content (which is still an open research problem). Being aware of the artifacts, understanding their sources, and impact on the human visual system can also provide new insights on how to measure, avoid, and compensate for them. Indeed, overall, the consideration of the human visual perception in 360-degree images and video processing is an important issue to take into account towards optimized VR services.

[4]https://www.its.bldrdoc.gov/vqeg/projects/immersive-media-group.aspx

References

[1] B. Micusik, "Two view geometry of omnidirectional cameras," Ph.D. dissertation, Czech Technical University in Prague, 2004.

[2] Z. Chen, Y. Li, and Y. Zhang, "Recent advances in omnidirectional video coding for virtual reality: Projection and evaluation," *Signal Processing*, vol. 146, pp. 66 – 78, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0165168418300057

[3] F. D. Simone, P. Frossard, P. Wilkins, N. Birkbeck, and A. C. Kokaram, "Geometry-driven quantization for omnidirectional image coding," in *Proc. of the Picture Coding Symposium*, 2016.

[4] B. Petry and J. Huber, "Towards effective interaction with omnidirectional videos using immersive virtual reality headsets," in *Proceedings of the 6th Augmented Human International Conference*. ACM, 2015, pp. 217–218.

[5] M. Hosseini, "View-aware tile-based adaptations in 360 virtual reality video streaming," in *2017 IEEE Virtual Reality (VR)*, March 2017, pp. 423–424.

[6] F. De Simone, P. Frossard, C. Brown, N. Birkbeck, and B. Adsumilli, "Omnidirectional video communications: new challenges for the quality assessment community," November 2017.

[7] M. Yuen and H. Wu, "A survey of hybrid MC/DPCM/DCT video coding distortions," *Signal Processing*, vol. 70, no. 3, pp. 247–278, Nov. 1998. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0165168498001285

[8] M. Yuen, "Coding artifacts and visual distortions," in *Digital video image quality and perceptual coding*, H. Wu and K. Rao, Eds. CRC Press, 2005, pp. 123–158.

[9] A. Unterweger, "Compression artifacts in modern video coding and state-of-the-art means of compensation," in *Multimedia Networking and Coding*, R. A. Farrugia and C. J. Debono, Eds. IGI Global, 2013, pp. 28–49. [Online]. Available: http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-4666-2660-7

[10] K. Zeng, T. Zhao, A. Rehman, and Z. Wang, "Characterizing perceptual artifacts in compressed video streams," B. E. Rogowitz, T. N. Pappas, and H. de Ridder, Eds., Feb. 2014, p. 90140Q. [Online]. Available: http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2043128

[11] L. Meesters, W. IJsselsteijn, and P. Seuntiens, "A Survey of Perceptual Evaluations and Requirements of Three-Dimensional TV," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 3, pp. 381–391, Mar. 2004. [Online]. Available: http://ieeexplore.ieee.org/document/1273547/

[12] A. Boev, D. Hollosi, and A. Gotchev, "Classification of stereoscopic artefacts," *Mobile3DTV Project report, available online at http://mobile3dtv. eu/results*, 2008. [Online]. Available: http://sp.cs.tut.fi/mobile3dtv/results/tech/D5.1_Mobile3DTV_v1.0.pdf

[13] A. Boev, A. Gotchev, and K. Egiazarian, "Stereoscopic Artifacts on Portable Auto-stereoscopic Displays: What matters?" p. 6, 2009.

[14] P. Hanhart, F. De Simone, M. Rerabek, and T. Ebrahimi, "3d Video Quality Assessment," in *Emerging Technologies for 3D Video*, F. Dufaux, B. Pesquet-Popescu, and M. Cagnazzo, Eds. Chichester, UK: John Wiley & Sons, Ltd, Apr. 2013, pp. 377–391. [Online]. Available: http://doi.wiley.com/10.1002/9781118583593.ch19

[15] S. Knorr, S. Croci, and A. Smolic, "A Modular Scheme for Artifact Detection in Stereoscopic Omni-Directional Images," in *Proceedings of the Irish Machine Vision and Image Processing Conference*, 2017.

[16] T. Ebrahimi, A.-F. Perrin, C. Bist, and R. Cozot, "Measuring quality of omnidirectional high dynamic range content," A. G. Tescher, Ed. SPIE, Sep. 2017, p. 38.

[17] R. Schatz, A. Sackl, C. Timmerer, and B. Gardlo, "Towards subjective quality of experience assessment for omnidirectional video streaming." IEEE, May 2017, pp. 1–6. [Online]. Available: http://ieeexplore.ieee.org/document/7965657/

[18] H. Duan, G. Zhai, X. Min, Y. Zhu, Y. Fang, and X. Yang, "Perceptual Quality Assessment of Omnidirectional Images," p. 5, 2018.

[19] H.-t. Lim, H. G. Kim, and Y. M. Ro, "VR IQA NET: Deep Virtual Reality Image Quality Assessment using Adversarial Learning," *ArXiv e-prints*, Apr. 2018.

[20] M. Xu, C. Li, Z. Wang, and Z. Chen, "Visual Quality Assessment of Panoramic Video," *arXiv preprint arXiv:1709.06342*, 2017.

[21] Y. Zhang, Y. Wang, F. Liu, Z. Liu, Y. Li, D. Yang, and Z. Chen, "Subjective Panoramic Video Quality Assessment Database for Coding Applications," *IEEE Transactions on Broadcasting*, pp. 1–13, 2018. [Online]. Available: https://ieeexplore.ieee.org/document/8350375/

[22] W. Zou, F. Yang, and S. Wan, "Perceptual video quality metric for compression artefacts: from two-dimensional to omnidirectional," *IET Image Processing*, vol. 12, no. 3, pp. 374–381, Mar. 2018. [Online]. Available: http://digital-library.theiet.org/content/journals/10.1049/iet-ipr.2017.0826

[23] Y. Yagi, "Omnidirectional sensing and its applications," *IEICE Transactions on Information and Systems*, vol. 82, no. 3, pp. 568–579, 1999.

[24] L. E. Gurrieri and E. Dubois, "Acquisition of omnidirectional stereoscopic images and videos of dynamic scenes: a review," *Journal of Electronic Imaging*, vol. 22, no. 3, p. 030902, Jul. 2013. [Online]. Available: http://electronicimaging.spiedigitallibrary.org/article.aspx?doi=10.1117/1.JEI.22.3.030902

[25] R. Szeliski *et al.*, "Image alignment and stitching: A tutorial," *Foundations and Trends® in Computer Graphics and Vision*, vol. 2, no. 1, pp. 1–104, 2007.

[26] W. Xu, "Panoramic Video Stitching," Ph.D. Thesis, University of Colorado, Boulder, 2012.

[27] W. Jiang and J. Gu, "Video stitching with spatial-temporal content-preserving warping." IEEE, Jun. 2015, pp. 42–48. [Online]. Available: http://ieeexplore.ieee.org/document/7301374/

[28] F. Pearson, II, *Map Projections: Theory and Applications*. CRC press, 1990.

[29] S. Peleg, M. Ben-Ezra, and Y. Pritch, "Omnistereo: panoramic stereo imaging," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 3, pp. 279–290, Mar. 2001. [Online]. Available: http://ieeexplore.ieee.org/document/910880/

[30] C. Schroers, J.-C. Bazin, and A. Sorkine-Hornung, "An omnistereoscopic video pipeline for capture and display of real-world vr," *ACM Trans. Graph.*, vol. 37, no. 3, pp. 37:1–37:13, Aug. 2018. [Online]. Available: http://doi.acm.org/10.1145/3225150

[31] R. Anderson, D. Gallup, J. T. Barron, J. Kontkanen, N. Snavely, C. Hernández, S. Agarwal, and S. M. Seitz, "Jump: virtual reality video," *ACM Transactions on Graphics*, vol. 35, no. 6, pp. 1–13, Nov. 2016. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2980179.2980257

[32] J. Tan, G. Cheung, and R. Ma, "360-Degree Virtual-Reality Cameras for the Masses," *IEEE MultiMedia*, vol. 25, no. 1, pp. 87–94, Mar. 2018. [Online]. Available: doi.ieeecomputersociety.org/10.1109/MMUL.2018.011921238

[33] H. Ishiguro, M. Yamamoto, and S. Tsuji, "Omni-directional stereo," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 257–262, Feb 1992.

[34] C. Wu, Z. Tan, Z. Wang, and S. Yang, "A dataset for exploring user behaviors in vr spherical video streaming," in *Proceedings of the 8th ACM on Multimedia Systems Conference*, ser. MMSys'17. New York, NY, USA: ACM, 2017, pp. 193–198. [Online]. Available: http://doi.acm.org/10.1145/3083187.3083210

[35] X. Corbillon, F. De Simone, and G. Simon, "360-Degree Video Head Movement Dataset." ACM Press, 2017, pp. 199–204. [Online]. Available: http://dl.acm.org/citation.cfm?doid=3083187.3083215

[36] Z. Chen, Y. Li, and Y. Zhang, "Recent advances in omnidirectional video coding for virtual reality: Projection and evaluation," *Signal Processing*, vol. 146, pp. 66–78, May 2018. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0165168418300057

[37] "Next-generation video encoding techniques for 360 video and vr," 2016. [Online]. Available: https://code.fb.com/virtual-reality/next-generation-video-encoding-techniques-for-360-video-and-vr/

[38] Khronos Group, *The OpenGL Graphics System: A Specification*, Version 4.4 (Compatibility Profile) ed., J. Leech, Ed., 2014. [Online]. Available: https://www.opengl.org/registry/doc/glspec44.compatibility.pdf

[39] M. Cagnazzo, B. Pesquet-Popescu, and F. Dufaux, "3d Video Representation and Formats," in *Emerging Technologies for 3D Video*, F. Dufaux, B. Pesquet-Popescu, and M. Cagnazzo, Eds. Chichester, UK: John Wiley & Sons, Ltd, Apr. 2013, pp. 102–120. [Online]. Available: http://doi.wiley.com/10.1002/9781118583593.ch6

[40] T. Stockhammer, "Dynamic adaptive streaming over http –: Standards and design principles," in *Proceedings of the Second Annual ACM Conference on Multimedia Systems*, ser. MMSys '11. New York, NY, USA: ACM, 2011, pp. 133–144. [Online]. Available: http://doi.acm.org/10.1145/1943552.1943572

[41] M. Hosseini and V. Swaminathan, "Adaptive 360 VR video streaming: Divide and conquer!" *CoRR*, vol. abs/1609.08729, 2016. [Online]. Available: http://arxiv.org/abs/1609.08729

[42] T. El-Ganainy and M. Hefeeda, "Streaming virtual reality content," *CoRR*, vol. abs/1612.08350, 2016. [Online]. Available: http://arxiv.org/abs/1612.08350

[43] K. K. Sreedhar, A. Aminlou, M. M. Hannuksela, and M. Gabbouj, "Viewport-adaptive encoding and streaming of 360-degree video for virtual reality applications," in *2016 IEEE International Symposium on Multimedia (ISM)*, Dec 2016, pp. 583–586.

[44] C. Ozcinar, A. D. Abreu, and A. Smolic, "Viewport-aware adaptive 360-degree video streaming using tiles for virtual reality," *CoRR*, vol.

abs/1711.02386, 2017. [Online]. Available: http://arxiv.org/abs/1711.02386

[45] X. Corbillon, G. Simon, A. Devlic, and J. Chakareski, "Viewport-adaptive navigable 360-degree video delivery," in *2017 IEEE International Conference on Communications (ICC)*, May 2017, pp. 1–7.

[46] M. Graf, C. Timmerer, and C. Mueller, "Towards bandwidth efficient adaptive streaming of omnidirectional video over http: Design, implementation, and evaluation," in *Proceedings of the 8th ACM on Multimedia Systems Conference*, ser. MMSys'17. New York, NY, USA: ACM, 2017, pp. 261–271. [Online]. Available: http://doi.acm.org/10.1145/3083187.3084016

[47] L. Xie, Z. Xu, Y. Ban, X. Zhang, and Z. Guo, "360ProbDASH: Improving qoe of 360 video streaming using tile-based HTTP adaptive streaming," in *Proceedings of the 2017 ACM on Multimedia Conference, MM 2017, Mountain View, CA, USA, October 23-27, 2017*, 2017, pp. 315–323. [Online]. Available: http://doi.acm.org/10.1145/3123266.3123291

[48] K. Liu, Y. Liu, J. Liu, A. Argyriou, and X. Yang, "Joint source encoding and networking optimization for panoramic video streaming over lte-a downlink," in *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, Dec 2017, pp. 1–7.

[49] A. Ghosh, V. Aggarwal, and F. Qian, "A rate adaptation algorithm for tile-based 360-degree video streaming," *CoRR*, vol. abs/1704.08215, 2017. [Online]. Available: http://arxiv.org/abs/1704.08215

[50] C. Zhou, Z. Li, and Y. Liu, "A measurement study of oculus 360 degree video streaming," in *Proceedings of the 8th ACM on Multimedia Systems Conference*, ser. MMSys'17. New York, NY, USA: ACM, 2017, pp. 27–37. [Online]. Available: http://doi.acm.org/10.1145/3083187.3083190

[51] C. Timmerer, M. Graf, and C. Mueller, "Adaptive streaming of vr/360-degree immersive media services with high qoe," in *2018 NAB Broadcast Engineering and IT Conference (BEITC)*, n. available, Ed. Washington DC, USA: National Association of Broadcasters (NAB), apr 2017, p. 5.

[52] C. Concolato, J. L. Feuvre, F. Denoual, F. Maze, N. Ouedraogo, and J. Taquet, "Adaptive streaming of hevc tiled videos using mpeg-dash," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. PP, no. 99, pp. 1–1, 2017.

[53] C. Zhou, Z. Li, J. Osgood, and Y. Liu, "On the effectiveness of offset projections for 360-degree video streaming," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 14, no. 3s, pp. 62:1–62:24, Jun. 2018. [Online]. Available: http://doi.acm.org/10.1145/3209660

[54] C. Ozcinar, A. D. Abreu, S. Knorr, and A. Smolic, "Estimation of optimal encoding ladders for tiled 360° vr video in adaptive streaming systems," in *2017 IEEE International Symposium on Multimedia (ISM)*, Dec 2017, pp. 45–52.

[55] L. Bagnato, P. Frossard, and P. Vandergheynst, "Plenoptic spherical sampling," in *Image Processing (ICIP), 2012 19th IEEE International Conference on*. IEEE, 2012, pp. 357–360.

[56] J. Sauer, M. Wien, J. Schneider, and M. Blaser, "Geometry-Corrected Deblocking Filter for 360° Video Coding using Cube Representation," San Francisco, CA, 2018, p. 5.

[57] F. D. Simone, P. Frossard, N. Birkbeck, and B. Adsumilli, "Deformable block-based motion estimation in omnidirectional image sequences," in *2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP)*, Oct 2017, pp. 1–6.

[58] D. Naik, I. D. D. Curcio, and H. Toukomaa, "Optimized Viewport Dependent Streaming of Stereoscopic Omnidirectional Video," in *Proceedings of the 23rd Packet Video Workshop*, ser. PV '18. New York, NY, USA: ACM, 2018, pp. 37–42. [Online]. Available: http://doi.acm.org/10.1145/3210424.3210437

[59] M. N. Garcia, F. D. Simone, S. Tavakoli, N. Staelens, S. Egger, K. Brunnstrom, and A. Raake, "Quality of experience and http adaptive streaming: A review of subjective studies," in *2014 Sixth International Workshop on Quality of Multimedia Experience (QoMEX)*, Sept 2014, pp. 141–146.

[60] M. Seufert, S. Egger, M. Slanina, T. Zinner, T. Hosfeld, and P. Tran-Gia, "A survey on quality of experience of http adaptive streaming," *IEEE Communications Surveys Tutorials*, vol. 17, no. 1, pp. 469–492, Firstquarter 2015.

[61] A. Hamza and M. Hefeeda, "A dash-based free viewpoint video streaming system," in *Proceedings of Network and Operating System Support on Digital Audio and Video Workshop*. ACM, 2014, p. 55.

[62] R. Schatz, A. Sackl, C. Timmerer, and B. Gardlo, "Towards subjective quality of experience assessment for omnidirectional video streaming," in *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)*, May 2017, pp. 1–6.

[63] R. Patterson, M. D. Winterbottom, and B. J. Pierce, "Perceptual Issues in the Use of Head-Mounted Visual Displays," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 48, no. 3, pp. 555–573, Sep. 2006. [Online]. Available: http://journals.sagepub.com/doi/10.1518/001872006778606877

[64] Y. Geng, J. Gollier, B. Wheelwright, F. Peng, Y. Sulai, B. Lewis, N. Chan, W. S. T. Lam, A. Fix, D. Lanman *et al.*, "Viewing optics for immersive near-eye displays: pupil swim/size and weight/stray light," in *Digital Optics for Immersive Displays*, vol. 10676. International Society for Optics and Photonics, 2018, p. 1067606.

[65] D. Pohl, G. S. Johnson, and T. Bolkart, "Improved pre-warping for wide angle, head mounted displays," in *Proceedings of the 19th ACM symposium on Virtual reality software and technology*. ACM, 2013, pp. 259–262.

[66] J. Zhao, R. S. Allison, M. Vinnikov, and S. Jennings, "Estimating the motion-to-photon latency in head mounted displays." IEEE, 2017, pp. 313–314. [Online]. Available: http://ieeexplore.ieee.org/document/7892302/

[67] M. Regan and G. S. P. Miller, "The Problem of Persistence with Rotating Displays," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 4, pp. 1295–1301, Apr. 2017. [Online]. Available: http://ieeexplore.ieee.org/document/7829409/

[68] W. Lin and C.-C. J. Kuo, "Perceptual visual quality metrics: A survey," *Journal of Visual Communication and Image Representation*, vol. 22, no. 4, pp. 297 – 312, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1047320311000204

[69] M. Yu, H. Lakshman, and B. Girod, "A framework to evaluate omnidirectional video coding schemes," in *Proc. of the IEEE International Symposium on Mixed and Augmented Reality*, 2015.

[70] *Quality metric for spherical panoramic video*, vol. 9970, 2016. [Online]. Available: https://doi.org/10.1117/12.2235885

[71] S. Yule, A. Lu, , and Y. Lu, "WS-PSNR for 360 video objective quality evaluation," in *MPEG Joint Video Exploration Team*, 2016.

[72] S. Chen, Y. Zhang, Y. Li, Z. Chen, and Z. Wang, "Spherical Structural Similarity Index for Objective Omnidirectional Video Quality Assessment," in *2018 IEEE International Conference on Multimedia and Expo (ICME)*, Jul. 2018, pp. 1–6.

[73] P. Hanhart, Y. He, Y. Ye, J. Boyce, Z. Deng, and L. Xu, "360-Degree Video Quality Evaluation," p. 5, 2018.

[74] R. C. T. E. Anne-Flore Perrin, Cambodge Bist, "Measuring quality of omnidirectional high dynamic range content," vol. 10396, 2017, pp. 10 396 – 10 396 – 18. [Online]. Available: https://doi.org/10.1117/12.2275146

[75] A. Singla, S. Fremerey, W. Robitza, and A. Raake, "Measuring and comparing QoE and simulator sickness of omnidirectional videos in different head mounted displays," in *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)*. Erfurt, Germany: IEEE, May 2017, pp. 1–6. [Online]. Available: http://ieeexplore.ieee.org/document/7965658/

[76] C. Li, M. Xu, X. Du, and Z. Wang, "Bridge the gap between vqa and human behavior on omnidirectional video: A large-scale dataset and a deep learning model," in *Proceedings of the 26th ACM International Conference on Multimedia*, ser. MM '18. New York, NY, USA: ACM, 2018, pp. 932–940. [Online]. Available: http://doi.acm.org/10.1145/3240508.3240581

[77] J. Boyce, E. Alshina, and Z. Deng, "Subjective testing method for 360° Video projection formats using HEVC," ISO/IEC, Tech. Rep. N16892, Apr. 2017.

[78] J. H. P. Vladyslav Zakharchenko, Kwang Pyo Choi, "Quality metric for spherical panoramic video," vol. 9970, 2016, pp. 9970–9970–9. [Online]. Available: https://doi.org/10.1117/12.2235885

[79] E. Upenik, M. Řeřábek, and T. Ebrahimi, "Testbed for subjective evaluation of omnidirectional visual content," in *2016 Picture Coding Symposium (PCS)*, Dec. 2016, pp. 1–5.

[80] H. R. Wu, A. R. Reibman, W. Lin, F. Pereira, and S. S. Hemami, "Perceptual Visual Signal Compression and Transmission," *Proceedings of the IEEE*, vol. 101, no. 9, pp. 2025–2043, Sep. 2013. [Online]. Available: http://ieeexplore.ieee.org/document/6575139/

[81] Y. Yan, K. Chen, Y. Xie, Y. Song, and Y. Liu, "The Effects of Weight on Comfort of Virtual Reality Devices," in *Advances in Ergonomics in Design*, F. Rebelo and M. M. Soares, Eds. Cham: Springer International Publishing, 2019, vol. 777, pp. 239–248. [Online]. Available: http://link.springer.com/10.1007/978-3-319-94706-8_27

[82] J. Zhuang, Y. Liu, Y. Jia, and Y. Huang, "User Discomfort Evaluation Research on the Weight and Wearing Mode of Head-Wearable Device," in *Advances in Human Factors in Wearable Technologies and Game Design*, T. Z. Ahram, Ed. Cham: Springer International Publishing, 2019, vol. 795, pp. 98–110. [Online]. Available: http://link.springer.com/10.1007/978-3-319-94619-1_10