Abstract—Omnidirectional (or 360-degree) images and videos are emergent signals in many areas such as robotics and virtual/augmented reality. In particular, for virtual reality, they allow an immersive experience in which the user is provided with a 360-degree field of view and can navigate throughout a scene, e.g., through the use of Head Mounted Displays. Since it represents the full 360-degree field of view from one point of the scene, omnidirectional content is naturally represented as spherical visual signals. 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 on 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 omnidirectional 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. However, the omnidirectional imaging pipeline has some particularities that induce specific distortions if not handled properly. First, the geometry of the content capture system is spherical, rather than planar. To reuse existing file formats and algorithms designed for perspective signals, the spherical signal is warped into a planar representation. The resulting planar signal, however, is not a classical natural visual signal. 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.

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 available on the market provide up to full High Definition (HD) display resolution. Since these devices usually provide a 110-degree of field-of-view, thus 4K resolution is being 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. Thus, 360-degree planar signals with resolution of 12K or higher will have to be efficiently stored and transmitted soon. Consequently, new coding and transmission techniques, able to cope with increasingly high data rates and to satisfy user’s expectations regarding visual quality, 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.

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 and stereoscopic 3D signals. 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. 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. 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. 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 that we consider in this paper, from acquisition to consumption by the end user via an HMD. Each step is briefly described hereafter.

A. Acquisition

Different optical systems have been proposed in the past to capture wide field of view signals. 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. 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.

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. For video stitching, additional video synchronization (if the individual sensors are not finely synchronized) and video stabilization (for moving cameras) may be necessary.

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). 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 some of the stitching problems, which are discussed later). 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.
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 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 space needed to store and transmit it. Most of the omnidirectional video systems re-use the same encoding tools as classical video solutions, such as H.264, H.265, VP9, or AV1. The main challenge with omnidirectional coding resides in mapping the content into rectangular frames that are typical 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 cube MAP 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 stereoscopic omnidirectional 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, multiview coding, etc. [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 stereoscopic 3D 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 communications. In addition, 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) 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.

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 viewpoint.

In the tile-based approach [41], [44], [49], [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 to be visualized, 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: a portion of the sphere surface 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 more immersive features and challenging of HMD-based approaches (i.e., it can introduce new distortions still not fully understood), with regards to consumption, this paper focus mainly on the HMD-based 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

Today, the most common omnidirectional recording systems are composed of multiple cameras, which altogether capture omnidirectional views of real-world scenes. 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 stereo 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 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 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 some visually annoying artifacts, in the multicamera rig. The same principle in the horizontal direction leads to the depth field 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).

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 capturing 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 3D 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 (i.e., neighboring regions on the spherical domain may end up not being neighbors on the planar representation, or vice-versa) on the planar representation. Different projections may imply different geometrical distortions and discontinuities regions. Fig. 5 shows examples of geometrical distortions and discontinuities resulted from the ERP and CMP map projections. Even though these distortions are not directly viewed by the end-user, the interaction between the geometrical distortions and 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 in the signal visualized by the end-user (see Fig. 6).

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. 6); 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
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.

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

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 these 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 omnidirectional geometry. Thus, they will 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 may be perceived as a blocking radial pattern by the end-user (see Fig. 7(b)). 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 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...
Fig. 6: Examples of (a) aliasing and blurring, (b) visible poles; and (c) visible seams due to projection and resampling.

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 on the properties of the visible seams during consecutive frames, it is also possible to end up with flickering seams in the temporal domain.

On 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 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”.

Moreover, the use of compressors unaware of the geometry of omnidirectional videos also result 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 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 is 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

| 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. |
| Staircase 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. |
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 \cite{58} is also another potential source for cross distortions. The cardboard effect may also be introduced by compression.

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 adaptive 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, rebufferring 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 \cite{59}, \cite{60} and few studies have investigated them in the context of stereoscopic 3D video \cite{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 \cite{62}.

In the viewport-aware adaptive streaming schemes, the ability of the system to predict how the user navigates the content can also 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 rebufferring 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, depending on the adopted tiling scheme, 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 magnify the content presented on the screen, supporting a comfortable content viewing. Second, it serves to optically eyes and the HMDs requires an optical system to support head-tracking system. The purpose of the optical system (see Fig. 9) is twofold. First, the close distance between the user’s head (and providing stereoscopic vision) and an optical and a head-tracking system. The 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.

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, 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 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.

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 a 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 resolution displays.

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. Despite being an annoying artifact, motion-to-photon delays may also induce motion sickness. Ideally, to

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

Fig. 9: Simplified schematics of HMD

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 HMD displays. In addition, new distortions that are specific to HMDs can appear due to the fact that, compared with traditional displays, the HMD is very close to the users 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.

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, 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 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.

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 a 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 resolution displays.

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. Despite being an annoying artifact, motion-to-photon delays may also induce motion sickness. Ideally, to

Fig. 9: Simplified schematics of HMD

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 HMD displays. In addition, new distortions that are specific to HMDs can appear due to the fact that, compared with traditional displays, the HMD is very close to the users 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.

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, 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 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.

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 a 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 resolution displays.

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. Despite being an annoying artifact, motion-to-photon delays may also induce motion sickness. Ideally, to

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

Fig. 9: Simplified schematics of HMD
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 such as eyestrain, blurred image, and misperception of distance, size, depth, or speed of objects [63].

![Fig. 10: Example of the screen-door effect.](image)

![Fig. 11: Example of smear from persistency.](image)

A rendered scene as seen without head movement. (b) The same scene as perceived by a moving head (smearred by 2deg). [7]

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 impacts the QoE of 360-degree videos, and much more 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

The current global objective quality metrics for 360-degree content —such as standard PSNR and SSIM, viewport-based PSNR/SSIM [36], S-PSNR [68], WS-PSNR [69], and S-SSIM [70]— are first attempts to measure the quality of 360-degree content. The use of standard image metrics such as PSNR and SSIM directly in the planar domain is straightforward, but unfortunately: (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 being very good at representing the subjective quality. The viewport-based PSNR/SSIM metrics apply PSNR and SSIM on the generated viewports, which is closer to what the users really see. However, this brings the issue on how to generate representative viewports, whose number, in theory, can be arbitrary large.

All the above objective metrics, however, fail in *properly considering the perceptual artifacts in a 360-degree processing chain*, as discussed in this paper. 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 [71]. Thus, objective metrics that detect each artifact reliably and efficiently, and that can build on the perceptual features of these artifacts are still necessary. This paper contribution is a first step towards this direction.

Finally, another important 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,
TABLE II: Summary of the visual distortions in 360-degree content.

| Spatial          | Optical distortions (individual cameras): | Projection: | Channel distortions: | Rendering: |
|------------------|-------------------------------------------|-------------|----------------------|------------|
|                  | blurring by defocus                       | geometrical distortions | data loss            | aliasing   |
|                  | barrel distortions                         | aliasing    | data distortion      | blurring   |
|                  | pin-cushion distortions                    | = circular pattern aliasing | spatial quality fluctuation | ringing    |
|                  | mustache distortions                       | blurring    | tiling artifacts      |            |
|                  | noise                                       | ringing     |                      |            |
|                  | chromatic aberrations                      | radial pattern close to the poles (due to oversampling on ERP) |                      |            |
| Stitching artifacts: | discontinuities (e.g., mis-aligned/broken edges:) | visible seams (due to sampling) |                      |            |
|                  | missing objects parts                      |                          |                      |            |
|                  | exposure artifacts                         |                          |                      |            |
|                  | black circle / blurred circle              |                          |                      |            |
| Blending artifacts: | Visible color- and luminance- mismatches of regions within an ODI. |                          |                      |            |
|                  | exposure artifacts                         |                          |                      |            |
|                  | visible seams due to color- and luminance- mismatches |                          |                      |            |
|                  | ghosting / duplicated objects             |                          |                      |            |
| Warping artifacts: | geometrical distortions / visible deformation of objects |                          |                      |            |
| Optical distortions (individual cameras): | blurring by defocus | geometrical distortions | data loss | aliasing |
|                  | barbell distortions                         | aliasing    | data distortion      | blurring   |
|                  | pin-cushion distortions                    | = circular pattern aliasing | spatial quality fluctuation | ringing    |
|                  | mustache distortions                       | blurring    | tiling artifacts      |            |
|                  | noise                                       | ringing     |                      |            |
|                  | chromatic aberrations                      | radial pattern close to the poles (due to oversampling on ERP) | spatial quality fluctuation | tiling artifacts |
| Stitching artifacts: | discontinuities (e.g., mis-aligned/broken edges:) | visible seams (due to sampling) | data loss | aliasing |
|                  | missing objects parts                      |                          |                      |            |
|                  | exposure artifacts                         |                          |                      |            |
|                  | black circle / blurred circle              |                          |                      |            |
| Blending artifacts: | Visible color- and luminance- mismatches of regions within an ODI. |                          |                      |            |
|                  | exposure artifacts                         |                          |                      |            |
|                  | visible seams due to color- and luminance- mismatches |                          |                      |            |
|                  | ghosting / duplicated objects             |                          |                      |            |
| Warping artifacts: | geometrical distortions / visible deformation of objects |                          |                      |            |
| Optical distortions (individual cameras): | blurring by defocus | geometrical distortions | data loss | aliasing |
|                  | barbell distortions                         | aliasing    | data distortion      | blurring   |
|                  | pin-cushion distortions                    | = circular pattern aliasing | spatial quality fluctuation | ringing    |
|                  | mustache distortions                       | blurring    | tiling artifacts      |            |
|                  | noise                                       | ringing     |                      |            |
|                  | chromatic aberrations                      | radial pattern close to the poles (due to oversampling on ERP) | spatial quality fluctuation | tiling artifacts |

**Temporal**

| Temporal          | motion blur                                   | flickering: |Delay |                |
|-------------------|-----------------------------------------------|-------------|------|--------------|
|                  | channel mismatch                               | = mosquito noise | video freezing |                |
|                  | motion discontinuities                         | = fine-granularity flickering | quality fluctuations |            |
|                  | appearing / disappearing objects               | = coarse-granularity flickering |                |            |
|                  | dynamic geometrical distortions               | jerkiness   |                |            |
|                  | dynamic ghosting                               | floating    |                |            |
|                  | wobbling artifacts                             | = texture floating |                |            |
|                  |                                               | = edge neighborhood floating |                |            |
|                  |                                               | flickering seams |                |            |
|                  |                                               | spatial pattern changes when crossing the faces |                |            |

**Stereoscopy**

| Stereoscopy       | depth plane curvature                        | ghosting (caused by disocclusion) | data loss |                |
|-------------------|----------------------------------------------|-----------------------------------|-----------|--------------|
|                  | keystone distortion                          | compression: |                 |             |
|                  | cardboard effect                              | cross-distortions                 | data distortion (binocular) |                |
|                  |                                               | cardboard effect                  | Viewport-aware adaptive streaming: |                |
|                  |                                               |                                   | cross distortions (due to spatial quality fluctuations) |                |
| Viewport-aware adaptive streaming: | video freezing | spatial quality fluctuation |                |            |
|                  |                                               | quality fluctuation           |                |            |
|                  |                                               | Tile-based viewport-aware adaptive streaming: |                |            |
|                  |                                               | spatial quality fluctuations     |                |            |
|                  |                                               | tile borders                    |                |            |
|                  |                                               | incomplete viewport             |                |            |

**Navigation / Head movement**

| Navigation / Head movement | Viewport-aware adaptive streaming: | Display limitations: |                |
|----------------------------|-----------------------------------|---------------------|--------------|
|                            | video freezing | spatial quality fluctuation |                |            |
|                            | quality fluctuation |                |                |            |
|                            | Tile-based viewport-aware adaptive streaming: |                |                |            |
|                            | spatial quality fluctuations |                |                |            |
|                            | tile borders |                |                |            |
|                            | incomplete viewport |                |                |            |

**Display limitations:**

| Display limitations: |                |
|---------------------|--------------|
|                            |                |
|                            |                |
|                            |                |
delivery, and rendering/consumption. It is not clear, for instance, how the few available datasets [22], [73], [20], [24] 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 for 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 common 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) 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 didn’t reach a consensus on the better methodologies for doing so.

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 [75], [76], while others have been performed using HMDs [68], [77]. On the one hand, visualizing the viewports on standard displays lacks the important immersive features (increased FoV, magnification of the content, sense of presence, motion sickness, etc.) 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; the user is immersed in the content; and that the content can both induce the sense of presence and motion sickness.

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

The fact the user is immersed in the content, and free to navigate with 3DoF the video 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”. Also, since the user is free for turning his head, he will probably not be able to see the entire scene, 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 improving streaming content, but they can (and should) also be used to improve quality metrics.

Second, the ideal viewing sequence duration, for instance, 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 [68], [77], focus on the overall signal quality, and do not provide insights on the impact of specific artifacts and, for instance, 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 [78] 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

Finally, it is 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) [79], [80], user discomfort, and usability are all important factors in defining a global VR quality of experience.

Moreover, 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. These

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 video encoder design is an important issue to take into account.

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