On Streaming Services for Omnidirectional Video and its Subjective Assessment

Igor D.D. Curcio

Introduction

Virtual Reality applications and services that use omnidirectional video are recently making the highlights of news releases related to the most advanced consumer electronics technologies. In particular, streaming of 360-degree video content is one of the most compelling applications in this area. It is technically very challenging, among other reasons, because of the mismatch between the required transmission bandwidth for video and the network bit rates available today for consumers. The Moving Picture Expert Group (MPEG) is currently working on the first standard called Omnidirectional Media Format (OMAF) [1] to be used as common industry platform for encoding, storing, and the delivery of 360-degree video. This letter introduces some of the challenges related to subjective assessment of a streaming system for 360-degree video, and introduces a new metric that could be utilized in the assessment process.

Streaming services for 360-degree video lack of a proper standardized methodology and procedures for subjective assessment. Currently, there are several open issues, which require further research and standardization.

360-Degree Streaming Systems

A streaming system can be depicted in simplified form as in Figure 1. Here, the rendering device is a *Head Mounted Display (HMD)*. Between the server and the HMD there is typically a

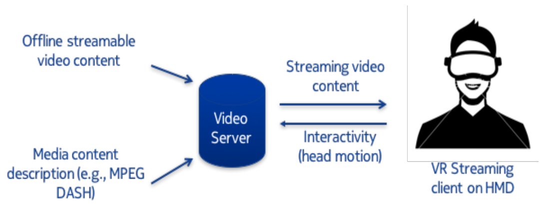


Figure 1. Example reference system for streaming omnidirectional video.

network connection which is wired or wireless (e.g., cellular or Wi-Fi) and that is affected by variable latencies.

When dealing with 360-degree video consumed on an HMD, there are, among others, few key parameters that are critical in such a system:

- The HMD *Field Of View (FoV)*.
- The size of the *foreground viewport*, which is the portion of omnidirectional video visible to the viewer in the horizontal and vertical directions. We call the other portions of the video which are not visible at a given time instant as *background viewport* (including part of the top and bottom portions of the 360-degree video).
- *Motion-to-Photon (MTP) Delay*, which is the elapsed time between the head motion to an orientation outside of the foreground viewport, and the subsequent system reaction to render a refreshed high quality viewport on the HMD. This is a factor that heavily impacts system interactivity.

The role of the above parameters in a streaming system for omnidirectional video will be clearer in the following.

Streaming techniques

Given the delivery of omnidirectional video is quite bandwidth hungry, compared to traditional 2D video, one of the main challenges for a successful streaming service is, more than ever, the deliver of the best possible visual quality using the smallest bandwidth.

360-degree video could be transmitted at a uniform quality, without differentiating between foreground and background viewports. This is the case when each rendered bit counts, and no compromises on video quality are permitted. This streaming technique is also referred to as *Viewport Independent Delivery (VID)*. From a bandwidth perspective, VID is quite demanding since the highest video quality is required all the time within the 360-degree space.

A great amount of transmission bandwidth could be saved by making use of the *Viewport Dependent Delivery (VDD)* technique. In this case, the foreground viewport is streamed at higher visual quality, whereas the background viewport(s) is (are) streamed at lower visual quality. This is a reasonable bandwidth saving mechanism, because for a given HMD orientation, the background viewport is not visible (and therefore not rendered).

As the HMD moves outside of the current foreground viewport, the visual quality may degrade for a certain time, until the system provides to stream and render a new foreground viewport corresponding to the new HMD orientation (*viewport switch*). Such time is essentially the MTP delay. The shorter this delay is, the better the user experience of such a system is. The larger (or unpredictable) this delay is, the more unusable becomes this system in terms of interactivity. Also, fatigue and motion sickness may be often experienced by a viewer.

Subjective Quality Assessment

The assumption that subjective assessment procedures for 2D/3D video used for several decades do apply also in the case of omnidirectional video watched with HMDs is too simplistic. In the latter case, the watching conditions and the immersion levels are different and, therefore, further research is needed. Here are some of the main challenges that currently lack of standardized methodology procedures.

Duration of the test video sequences: because of the wider FoV, compared to traditional video viewed on a 2D flat panel display, the length of video clips shall be sufficient such that all 360-degree content is watched and assessed. To perform such task, more time is needed for a test subject in order to explore the content in all directions. It appears logical that different durations for test video sequences may be envisaged in case of 180-degree content, as opposed to 360-degree content. A test case duration may be extended by allowing a looping function for a given video test sequence, and scored by a test subject when the judgment has been comfortably formed by a test subject. However, the duration of each test sequence (or test case) cannot be too long in order 1) not to produce fatigue symptoms on the test subjects, 2) to avoid that test sessions become too long and unmanageable, and 3) to avoid that test subjects forgive the impairments located in an early location of the test sequence (temporal forgiveness).

Full video assessment: for some use cases, such as 180-degree video, partial video subjective assessment may be sufficient. However, in the general streaming case of VDD of omnidirectional video, the subjective evaluation should not be limited to viewing a particular viewport orientation corresponding to a small portion of the whole video. This may happen, for example, if a viewer concentrates to watching only some details of the whole video in a quasi-still orientation, neglecting all other parts. For instance, there should be a way for the test subjects to form an opinion score based on the overall 360-degree video quality assessment, possibly without incurring in side effects, such as motion sickness or nausea.

Fair within-subject and between-subjects assessment: there might be the chance that the parts (and/or time instances) of a 360-degree test sequence viewed and assessed by a subject may differ from the parts viewed by the same subject for another test condition (or by another test subject when assessing the same test sequence). For ensuring a fair evaluation procedure, the test methodology should enable comparable test results for the same subject (or for several subjects) also in the case of not perfectly identical watching patterns for different test cases. In other words, it should be easy to verify whether, for different test cases, the same or different subjects view "the same thing at the same time". When this condition is not met, there should also be a way to measure how far apart are different viewing patterns for different test conditions.

Similarity Ring Metric (SRM)

A new metric for the last of the challenges introduced in the previous section is presented here. Such metric measures essentially the degree of similarity of a set of watching patterns [2]. For simplicity, the remainder of this discussion will focus on Yaw which, by convention, measures the horizontal FoV.

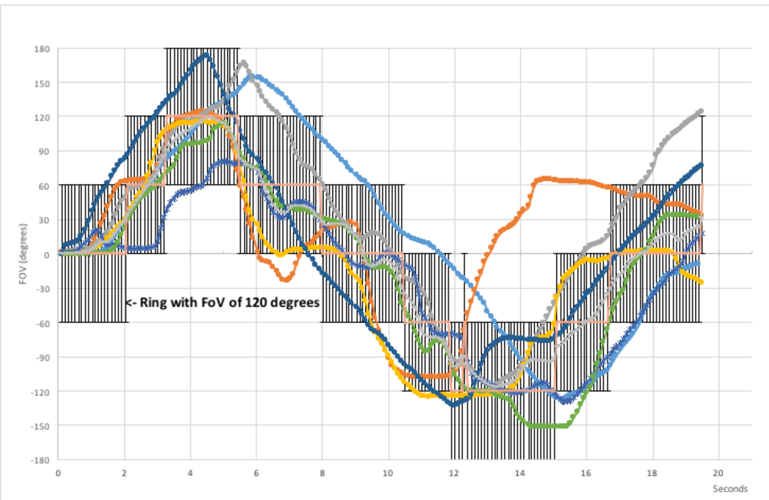


Figure 2. Example Ring with size of 120 degrees. The ring moves through the curves at discrete steps according to the foreground viewport orientation. Here the test subjects were instructed to follow a specific motion pattern without any speed constraints.

A typical plot of Time vs. Yaw could look like in Figure 2. Each curve represents the watching pattern of one (or more) subjects when viewing the same omnidirectional test sequence. It can be clearly seen that, for each time instant, the curves follow the same rough direction, but they are far apart by a certain distance. In practice, it is rare that all curves overlap (i.e., exploit a perfect watching similarity), since each test case carries some elements of variability even within the same test subject. These elements are direction and speed of motion while watching a video with an HMD.

However, it is possible to verify if the aggregate set of curves falls within a certain range. We could ideally think of this range as a "ring" (see Figure 2). The goal is then to check if the ring can travel through all curves from the beginning to the end of a test clip. If this occurs, it means that all clips (i.e., the curves) have been watched with high similarity.

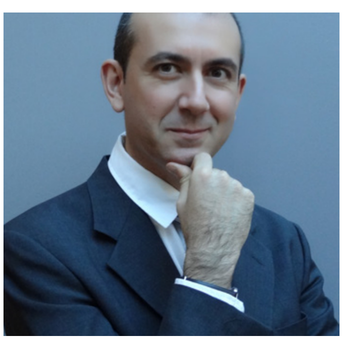
More specifically, if the curves are related to several test cases of the same sequence watched by a test subject, high similarity means that the subject has been watching and assessing the same content at the same time. Differently, if the curves are related to different test subjects that evaluate a particular clip, high similarity means that the test subjects have been watching and assessing the same content at the same time.

The ring size is a critical parameter here, and is determined by the HMD FoV or the content FoV. In Figure 2, the ring size is the size of the foreground viewport, which is 120 degrees.

As it is difficult to achieve a SRM of 100%, it is convenient to define a *Similarity Threshold* ST, e.g., 80%. In this way, a rejection criterion could be established: for example, the results of a particular subjective test set could be rejected if SRM < ST.

Future Developments

The support for multi-dimensional SRM with pitch and roll is one of the development areas. Furthermore, it is worth mentioning that the SRM could also be tailored to tiled streaming supported, for example, by the HEVC video codec. Further research advances are envisaged also in the area of foveated streaming as soon as adequate hardware for gaze tracking will be available in mainstream HMDs.



Igor D.D. Curcio *is Principal Scientist at Digital Media Laboratory of Nokia Technologies, Tampere, Finland.*

References

[1] "Text of ISO/IEC FDIS 23090-2 Omnidirectional Media Format", ISO/IEC JTC1/SC29/WG11 Doc. N17235, Macau, China, Oct. 2017.

[2] "Similarity Ring Metric (SRM) for Subjective Evaluation of 360-Degree Video", 3GPP TSG SA WG4, Tdoc AHVIC-090, Nokia, Mar. 2017.