

# Adaptive 360 VR Video Streaming: Divide and Conquer!

Mohammad Hosseini<sup>\*†</sup>, Viswanathan Swaminathan<sup>†</sup>

<sup>\*</sup>University of Illinois at Urbana-Champaign (UIUC)    <sup>†</sup>Adobe Research, San Jose, USA  
Email: shossen2@illinois.edu, vishy@adobe.com

**Abstract**—While traditional multimedia applications such as games and videos are still popular, there has been a significant interest in the recent years towards new 3D media such as 3D immersion and Virtual Reality (VR) applications, especially 360 VR videos. 360 VR video is an immersive spherical video where the user can look around during playback. Unfortunately, 360 VR videos are extremely bandwidth intensive, and therefore are difficult to stream at acceptable quality levels.

In this paper, we propose an adaptive bandwidth-efficient 360 VR video streaming system using a divide and conquer approach. We propose a dynamic view-aware adaptation technique to tackle the huge bandwidth demands of 360 VR video streaming. We spatially divide the videos into multiple tiles while encoding and packaging, use MPEG-DASH SRD to describe the spatial relationship of tiles in the 360-degree space, and prioritize the tiles in the Field of View (FoV). In order to describe such tiled representations, we extend MPEG-DASH SRD to the 3D space of 360 VR videos. We spatially partition the underlying 3D mesh, and construct an efficient 3D geometry mesh called *hexaface sphere* to optimally represent a tiled 360 VR video in the 3D space. Our initial evaluation results report up to 72% bandwidth savings on 360 VR video streaming with minor negative quality impacts compared to the baseline scenario when no adaptations is applied.

## I. INTRODUCTION

Advances in computing hardware and networking technologies with support of high bandwidth have enabled the use of new 3D media such as 3D immersion and 360-degree VR video applications in the recent years. 360 VR videos are immersive spherical videos, mapped into a 3D geometry as shown in Figure 1, where the user can look around during playback using a VR head-mounted display (HMD). This gives viewer a sense of depth in every direction.

Despite the promising nature of 360 VR videos, existing 360 VR video applications are still restricted to lower resolutions compared to their 2D counterparts. Unfortunately, a major challenge is how to efficiently transmit the bulky 360 VR video streams to bandwidth-constrained devices such as wireless VR HMDs given their high bitrate requirements. Especially with the 4K video resolution being widely viewed as a functional minimum resolution for current HMDs, and 8K or higher being desired, these new media are extremely bandwidth intensive and difficult to stream at acceptable quality levels. Thus there must be a balance between the requirements of streaming and the available resources on the display devices. One of the challenges to achieving this balance is that we need to meet this requirement without much negative impact on the user's viewing experience. While our work is motivated by the 360 VR video applications with 8K and 12K resolutions and the data rate issues that such rich multimedia system have, a semantic link between FoV, spatial partitioning, and stream prioritization has not been fully developed yet for the purpose

of bandwidth management and high performance 360 VR video streaming. Hence, we propose to utilize this semantic link in our dynamic adaptations.

In this paper, we propose FoV-aware streaming adaptations for efficient delivery of high-resolution 360 VR videos to bandwidth-limited HMDs. We spatially divide the videos into multiple tiles while encoding and packaging, use MPEG-DASH SRD to describe the spatial relationship of tiles in the 360-degree space, and prioritize the tiles in the viewport. We then extend our tiling process to the 360-degree VR environment to *conquer* the intense bandwidth requirements using viewport adaptation techniques. To achieve that, we spatially partition the underlying 3D mesh, and construct an efficient 3D geometry mesh called *hexaface sphere* to optimally represent a tiled 360 VR video in the 3D space. On the other hand, due to the absence of a fine-grained prioritized mechanism, most 360 VR streaming systems today handle all the portions of the spherical view as equally important, resulting in sub-optimal resource usage. Our approach is to deliver higher bitrate content to regions where the user is currently looking and is most likely to look, and delivering lower quality level to the area outside of user's immediate viewport. Our initial evaluation results using a real-platform wireless HMD and multiple 360 VR video benchmarks show that our adaptations significantly reduces the amount of bandwidth required to deliver a high quality immersive experience, and increases the overall 360 VR video quality at a given bandwidth.



Fig. 1. Visual structure of a 360 VR video

## II. BACKGROUND AND RELATED WORK

In this section we briefly present different concepts and categories of state-of-the-art related to our proposed approach.

### A. Dynamic Adaptive Streaming

One of the main approaches for bandwidth saving on bandwidth-intensive multimedia applications is adaptive streaming. Dynamic Adaptive Streaming over HTTP (DASH) specifically, also known as MPEG-DASH [1], [2] is an ISO standard that enables adaptive bitrate streaming whereby a client chooses the video segment with the appropriate quality (bit rate, resolution, etc.) based on the constrained resources such as bandwidth available at the client.

As a part of ISO/IEC 23009-1:2015, a new amendment called Spatial Relationship Description (SRD) has been added to MPEG-DASH standard, which allows to define spatial

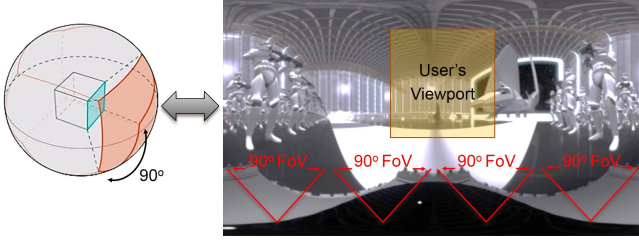


Fig. 2. An example visual view of how a 90-degree spherical viewport is mapped on a raw 360 video (*Star Wars*)

relationships between spatially partitioned multimedia content [3]. The SRD feature explores how to combine several spatially-related videos in a tiled combination, while at the same time provides backward compatibility with regular definition of adaptation sets. It provides a unique approach to tiled streaming in the perspective of the facilitated ultra high-resolution video display, specifically in the context of immersive environments such as those seen in 360 VR videos. There have been some work exploring the features of MPEG-DASH SRD. Le Feuvre *et al.* in their work [4] explored spatial access of 2D video contents within the context of MPEG-DASH, and discussed how the tiling feature of SRD extension can enable that. In another work [5], D'Acunto *et al.* explored the use of MPEG-DASH SRD to partition a video into sub-parts to provide a zooming feature inside 2D video contents. In this work, we extend the semantics of MPEG-DASH SRD towards the context of 360 VR videos, and use that to partition the bulky 360 VR videos into spatially related tiles in the 3D space for the purpose of view-aware adaptation. Our aim is to explore prioritization of the tiles by assigning highest resolution only to those tiles that are relevant to the user's FoV and lower resolution to tiles outside of the user's FoV.

### B. Prioritized Multimedia Streaming

Generally, different parts of multimedia can have different importance given various settings such as view, region, or the context. Hosseini *et al.* [6], [7] adopted prioritization techniques towards efficiently transmitting, rendering, and displaying bulky 3D contents to resource-limited devices given the importance of various 3D objects in the context. Similarly, in the context of 3D tele-immersive systems, the authors in [8] studied stream prioritization in regards to bandwidth savings. Their approach assigns higher quality to parts within users' viewport given the features of the human visual system.

In this paper, we build upon the concepts from these works to implement an adaptive prioritized view-aware streaming approach to reduce the bandwidth requirements of 360 VR video streaming.

## III. METHODOLOGY

Similar to the context of 3D graphics, the visual experience of 360 VR videos is also based upon texturing. Textures are used to provide surface wrapping for a 3D object mesh, with 3D textures being a logical extension of the traditional 2D textures. 360 VR videos are created by mapping a raw 360 video as a 3D texture onto a 3D geometry mesh, often a sphere, with the user at the center of that geometry as shown in Figure 1. In this 360-degree environment however, a user is viewing only a small portion of the whole raw 360-degree video at any

given time. Therefore, a user's FoV is also only a small portion equivalent to a specific confined region on the 3D object mesh which is *spatially* related to the corresponding portion of the raw content. For example, the Samsung Gear VR HMD offers a 96-degree FoV, meaning it can only almost cover a quarter of a whole 360-degree-wide content horizontally. Figure 2 illustrates this concept. The left subfigure shows an example 90-degree viewport as projected on a spherical 3D geometry, while the right subfigure shows how the mapping of the viewport corresponds to that of a given frame on a raw 360-degree video.

One of the major challenges in streaming 360-degree VR videos is the high bandwidth demands. To decrease the bandwidth requirements, our approach is to assign higher quality to parts within a user's viewport, and use lower quality to parts which are not within the immediate viewport of the user. This approach also makes it possible to stream tiles inside the viewport at highest resolution, at or near the native resolution of the HMD, virtually enabling a total resolution of 8K and higher in the 360-degree VR environment. To achieve that, our approach consists of two parts. First, the raw 360-degree video is spatially partitioned into multiple tiles. Using the features of MPEG-DASH SRD, a reference space is defined for each tile, corresponding to the rectangular region encompassing the entire raw 360-degree video. Second, we partition the underlying 3D geometry into multiple segments, each representing a subset of the original 3D mesh with a unique identifier. Finally, a mapping mechanism is defined for spatial positioning of the tiles on the 3D space, so that each tile be textured on its corresponding 3D mesh segment.

### A. 3D Geometry Construction: Hexaface Sphere

Most of the cameras available nowadays output 360 panoramic videos using a equirectangular format. While these videos are mapped into a flat projection for storage, they are inherently spherical. To achieve a spherical view as a common practice, a 3D sphere mesh is created surrounding the virtual camera inside the main virtual scene. Every individual photo-sphere, an image of a equirectangular 360 video frame, is wrapped on the internal

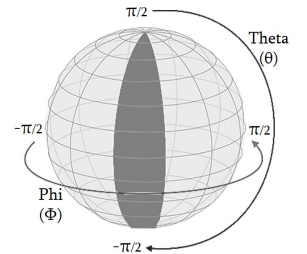


Fig. 3. A 3D sphere mesh. The highlighted area represents a slice.

surface of the sphere. The stereoscopic depth of 360 VR video requires two photo-spheres be stored and shown side-by-side representing the small disparity of left eye and right eye. The sphere is setup in such a way that it contains vertex locations and texture coordinates to achieve equirectangular mapping. The front faces of each of the rendered sphere polygons is culled to enable their internal surface showing as opposed to external showing. As a part of our geometry construction, we programmatically create a 3D sphere mesh in code to be able to control and further modify the geometry, and set the normal vectors of all of the mesh segments to point inside, towards the center of the sphere, achieving internal showing.

In order to create a 3D sphere mesh, we create an array of vertices by procedurally generating triangles for rendering. We approximate the sphere and provide a quality parameter

**Algorithm 1:** Our process to generate a 3D sphere mesh

---

$M$ : the number of stacks  
 $N$ : the number of slices  
 $\forall m : 0 \leq m \leq M$  and  $\forall n : 0 \leq n \leq N - 1$ , calculate and store a spatial point  $P(x, y, z)$  such that:  
 $P_x \leftarrow \sin(\pi \times m/M) \cdot \cos(2\pi \times n/N)$   
 $P_y \leftarrow \sin(\pi \times m/M) \cdot \sin(2\pi \times n/N)$   
 $P_z \leftarrow \cos(\pi \times m/M)$   
 Draw the line segments between the each vertex.

---

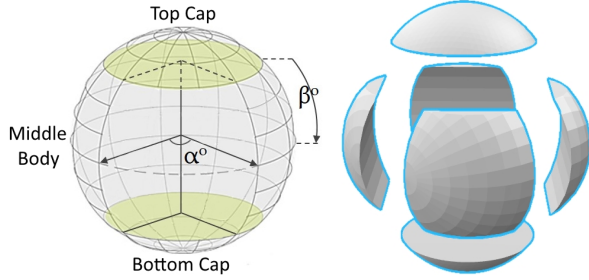


Fig. 4. Visual overview of a generated hexaface sphere.

to account for the trade-off between more smoothness and more triangles to render (representing highest quality) and rendering speed (lower spherical smoothness and chunkier shape). Figure 3 illustrates a visual view of the structure of a 3D sphere mesh. Our quality parameter is a combination of two major parameters which together control the smoothness of the sphere: a) number of slices, which represent the number of vertical cuts. Let's assume each slice collides with a sphere's perimeter at a vertical degree  $\theta$  which  $-\pi/2 \leq \theta \leq \pi/2$ ; and b) number of stacks, which is the number of horizontal cuts determining the number of rows of rectangles. Let's assume each stack collides with a sphere's perimeter at a horizontal degree  $\phi$  which  $-\pi/2 \leq \phi \leq \pi/2$ . Algorithm 1 presents pseudo-code of our process to create 3D spherical mesh.

Next, in order to generate 3D mesh segments, we partition the 3D sphere into multiple different 3D meshes, in a two-step process. In the first step, we split the sphere into 3 major parts:

- The top cap, which includes meshes from the top point of the sphere (i.e. top pole where  $\theta = +\pi/2^\circ$ ) to the top stack of the middle body ( $\theta = +\beta^\circ$ ), totaling  $(\pi/2 - \beta)^\circ$ ,
- The middle body which is a ring of triangles stretching from the bottom stack of the top cap ( $\theta = +\beta^\circ$ ) to the top stack of the bottom cap ( $\theta = -\beta^\circ$ ), totaling  $2\beta^\circ$ ,
- The bottom cap, which includes triangles ranging from the bottom stack of the middle body ( $\theta = -\beta^\circ$ ) to the bottom point of the sphere (i.e. bottom pole where  $\theta = -\pi/2^\circ$ ), for  $\pi/2 - \beta$  degrees.

360-degree videos formatted in equirectangular projection can contain redundant information at the top and bottom ends, but not the middle body. So in the second step, we further split the middle body into multiple 3D meshes, each covering  $\alpha^\circ$  of the entire 360-degree wide screen. The number of second-step partitions, and therefore  $\alpha^\circ$ , can be a function of the horizontal FoV of the VR HMD platform. We split the middle body into four segments with  $\alpha^\circ = \pi/2$ , so that each segment has a 90-degree cover, and similarly set  $2\beta = \pi/2$  following the default settings for the vertical FoV. Overall, with this procedure, our

projection will result into a combination of *six* 3D meshes of a 3D sphere. We call our customized 3D geometry a *hexaface sphere* 3D mesh. Figure 4 illustrates a visual overview of our process to generate a hexaface sphere. It should be noted that our assignment of  $\alpha = \pi/2$  and  $\beta = \pi/4$  is not a hard requirement, and is derived from the default settings of our VR HMD device as well as the result of our performance and quality trade-offs. Our approach is general, and depending on the performance of underlying hardware can be extended towards higher number of tiles and 3D meshes. Similarly, our tiling process is general, and therefore can also be employed in other geometrical projections such as cube maps.

### B. Viewport Tracking

VR device orientation is usually defined using three types of rotation over the  $X$ ,  $Y$ , and  $Z$  axes. The values are inherently represented in the *Tait-Bryan* angular system, which uses a special form of Euler angles that require 3 rotations around each of the 3 axes. To avoid possible orientation tracking problems such as *gimbal lock*, the angular Euler system is transformed into a different rotational system known as a *quaternion*, which is converted into a unit rotation matrix. To enable view awareness, we use the following three steps to create valid confines of unit quaternions specifically set for each of the hexaface sphere 3D mesh segments:

- convert Euler angles to a unit quaternion representation for VR device orientation tracking,
- calculate an array corresponding to a normalized direction vector from our quaternion,
- combine the values together to compute the confines of segment-specific quaternion representations inside the hexaface sphere.

With the confines of each 3D mesh segment being defined, we then identify which segments and the corresponding video tiles intersect with a user's viewport and implement our viewport tracking at every frame. With viewport tracking, we then implement view-aware adaptation, and dynamically deliver higher bitrate content to the tiles within the user's FoV, and assign lower quality level to the area outside the user's immediate viewport.

## IV. EVALUATION

To evaluate our work, we used Samsung Gear VR HMD mounted with the Samsung Galaxy S7 smartphone with 4GB RAM and Android Marshmallow 6.0.1 as our target VR platform. We used Oculus Mobile SDK 1.0.3 along with Android SDK API 24 for development of a 360 VR video streaming application prototype based on MPEG-DASH SRD. Our VR platform provides a total resolution of 2560x1440 (1280x1440 per eye), with maximum frame rate of 60 FPS and a horizontal FoV of 96 degrees. As stated before, we set the vertical FoV of our 360 VR video prototype to 90 degrees. We prepared 5 different 360 equirectangular-formatted sample videos publicly available on Youtube as test sequences for the purpose of applying our adaptations. Table IV provides detailed information about our test video sequences.

To benefit from the features of MPEG-DASH SRD-based streaming and for the purpose of texturing the hexaface sphere mesh, we developed an executable script based on FFmpeg Zeranol 64-bit API that given a video input, spatially crops



TABLE I

360 Video	Resolution	Orig. Bitrate (Mbps)	FPS (Hz)
Waldo	3840x1920	20.0	30
Plane	3840x1920	13.1	30
Karate	1920x960	4.4	30
Star Wars	3840x2160	13.0	24
Racing Car	4096x2048	12.6	24

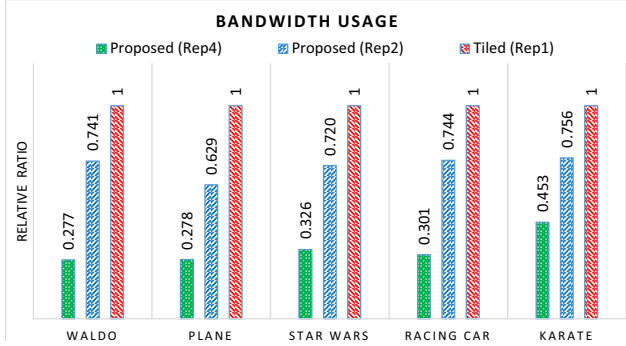


Fig. 6. A comparison of bandwidth savings of streaming different 360 VR videos, using our adaptations, and tiled-streaming with no adaptation.

the video and generates 6 different tiles as per requirements of our hexaface sphere geometry. We encoded all video segments with H.264 AVC encoder into 4 different representations using a hierarchical resolution degrading, with original resolutions decreasing polynomially at every step, as depicted in Figure 5.

We applied various sets of resolutions to different tiles to experiment how our prioritized adaptations affects the average bitrate as well as the perceived quality. Each trial of our experiment was run for a total of 30 seconds, and during each run we setup the experiment in such a way that each tile is within user's viewport for 5 seconds. We measured the bandwidth usage in terms of average bitrate, when maximum resolution is assigned for tiles within immediate FoV, and lower resolutions assigned to the other tiles. Figure 6 demonstrates results for only a small subset of our experiments on all of our benchmarks, with ratios normalized to fit within a unit. We measured the relative bandwidth usage when using our adaptations compared to the baseline case where no adaptation is applied (the 360 VR video is tiled; no viewport awareness is present, and all tiles are streamed with highest representation ( $REP_1$ )). As can be seen, the results show that our adaptations can significantly save bandwidth usage for up to 72% compared to baseline case where our adaptation approach is not employed. Figure 7 shows a sample screenshot of the experiments on *Waldo*. While the highest representation possible ( $REP_1$ - resolution of 960x1920) is delivered to the main front tile, lowest representation is delivered to the peripheral tile on the right identified by the red confines ( $REP_4$ - resolution of 240x480) Our adaptations results in minor noticeable quality impacts, sometimes not even perceptible, while maintaining the original

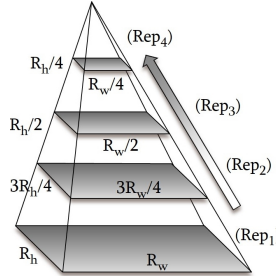
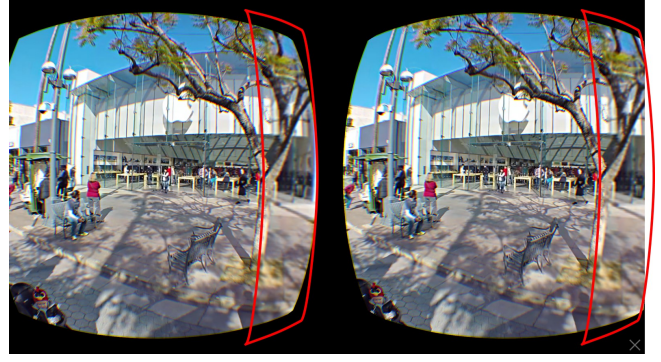


Fig. 5. Our hierarchical resolution degrading for various representations.

Fig. 7. Visual comparison of a specific frame within *Waldo* with the peripheral tiles having  $REP_4$  (resolution of 240x480)

quality for the main viewport to ensure a satisfactory user experience.

## V. CONCLUSION AND FUTURE WORK

In this paper, we proposed bandwidth-efficient FoV-aware streaming adaptations to tackle the high bandwidth demands of 360 VR videos. Our novel adaptations exploits the semantic link of MPEG-DASH SRD with a user's viewport to provide dynamic view awareness in the context of VR videos. We *divide* the bulky 360 VR videos as well as the underlying 3D geometry into spatially partitioned segments in the 3D space, and then *conquer* the huge streaming bandwidth requirements using a dynamic viewport adaptation. Our initial experimental results shows our adaptations can save up to 72% of bandwidth on 360 VR video streaming without much noticeable quality impacts.

We are currently working on an adaptive rate allocation system for tile streaming given the available bandwidth. We also plan to extend the hexaface sphere towards higher number of 3D segments and measure the bandwidth saving-performance trade-offs.

## REFERENCES

- [1] "MPEG DASH, Information technology – Dynamic adaptive streaming over HTTP (DASH) – Part 1: Media presentation description and segment formats," ISO-IEC\_23009-1, 2014-05-12.
- [2] T. Stockhammer, "Dynamic adaptive streaming over HTTP: standards and design principles," in *Proc. of 2011 ACM Multimedia Systems*, ser. MMSys '11, 2011.
- [3] O. A. Niamut, E. Thomas, L. D'Acunto, C. Concolato, F. Denoual, and S. Y. Lim, "MPEG DASH SRD: Spatial Relationship Description," in *Proceedings of the 7th International Conference on Multimedia Systems*, ser. MMSys '16. New York, NY, USA: ACM, 2016, pp. 5:1–5:8.
- [4] J. Le Feuvre and C. Concolato, "Tiled-based adaptive streaming using mpeg-dash," in *Proceedings of the 7th International Conference on Multimedia Systems*, ser. MMSys '16. New York, NY, USA: ACM, 2016, pp. 41:1–41:3.
- [5] L. D'Acunto, J. van den Berg, E. Thomas, and O. Niamut, "Using mpeg dash srd for zoomable and navigable video," in *Proceedings of the 7th International Conference on Multimedia Systems*, ser. MMSys '16. New York, NY, USA: ACM, 2016, pp. 34:1–34:4.
- [6] M. Hosseini, J. Peters, and S. Shirmohammadi, "Energy-budget-compliant adaptive 3D texture streaming in mobile games," in *Proceedings of the 4th ACM Multimedia Systems Conference*. ACM, 2013, pp. 1–11.
- [7] M. Hosseini, A. Fedorova, J. Peters, and S. Shirmohammadi, "Energy-aware adaptations in mobile 3D graphics," in *Proceedings of the 20th ACM international conference on Multimedia*. ACM, 2012, pp. 1017–1020.
- [8] M. Hosseini, G. Kurillo, S. R. Etesami, and J. Yu, "Towards coordinated bandwidth adaptations for hundred-scale 3d tele-immersive systems," *Multimedia Systems*, pp. 1–14, 2016.