WHITE PAPER

Enabling High Quality Volumetric VOD Streaming Over Broadband and 5G

An investigation of Volumetric VOD streaming compression techniques

---

Authors: Anna Qingfeng Li, William Cheung, Robert Kawiak, Dusty Robbins – Intel
Contributors: Michael Chen & Pete Quesada – Comcast
Tonaci Tran, Jason Juang & Caoyang Jiang – HypeVR

Introduction

As announced at CES 2019, Comcast is working with Intel to help build a platform to power the ultra-connected experiences of the future, from the core network all the way into the home and beyond. In April 2019, at NAB (National Association of Broadcasters), cinematographers discussed how they are innovating new ways to tell stories with tools such as volumetric capture. Sportscasters are also getting into the game with recent examples such as the Sky Scope golf feature for the 2019 Open Championship in July. These efforts toward scalable streaming and reproduction arrive at an exciting time.

For volumetric content, the industry as a whole has been focused on improving the quality of volumetric capture from commodity depth cameras such as Realsense D415. For this, they need a scalable commercial solution to stream that content over the network. To accelerate streaming, the industry requires guidance on which technologies can be applied in VoD streaming use cases, based on requirements and trade-offs of these technologies in the areas of latency, bitrate, and quality. Initial VoD use cases supported include work-based training, sports training, singer songwriter and small stage broadcasts.

Contents

Volumetric content captured in raw pointcloud format, even after generating the 3D pointcloud, would require 1Gbps of bandwidth to stream satisfactorily. This means a home or office would require a 1Gbps connection to stream this type of content, severely limiting the type of devices and users that can access this content. Remote rendering approaches exist that can deliver this type of content using lower bandwidth, but those solutions would require a motion to photon latency of 100ms per frame or more, whereas the tolerance on most mobile devices for motion to photon latency today is only16-30ms today.

To address these problems, we will illustrate a Volumetric VOD framework that will allow users to instantaneously stream high quality volumetric content with minimal latency, and under 100Mbps. This makes the volumetric content streamable over 5G speeds, making it possible to reach a wider range of customers over broadband and mobile networks. We will also illustrate a tiling technique reduces bandwidth up to 50%, as well as two compression schemes that can further reduce bandwidth requirements by 5-10X[1,2,3].

---

1. Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products.
2. Configurations: [Testing Configuration: Server: Core i7-7700 3.6 GHz, Client: Apple iPad* Pro 3rd generation + Test: Intel Real Sense Captured Volumetric Video + Robert Kawiak + 09/03/2019].
3. For more information go to www.intel.com/benchmarks

Volumetric VOD Streaming Challenges

We encountered two main challenges while working to provide a solution for Volumetric VOD streaming. The first challenge encountered was meeting the motion-to-photon latency of users' good immersive media QoS. The second challenge was in delivering the volumetric content within the bandwidth constraints of the 5G and broadband networks in order to make the content as accessible as possible.

The way we meet the latency challenge is by streaming the volumetric video content directly to the user so they are able to access the full 3D objects. This allows each user's motion to produce a new viewpoint instantly on the device. The latency is essentially the render time of the device which is ~16ms plus buffering, as most devices can render ~60FPS today. The alternative is to render the scene from the user's perspective on an Edge server or the Cloud and then send the rendered frame as a video frame. However, that would have latency results in the ~80-180ms range or higher, and not acceptable for AR immersive usages driven on mobile devices.

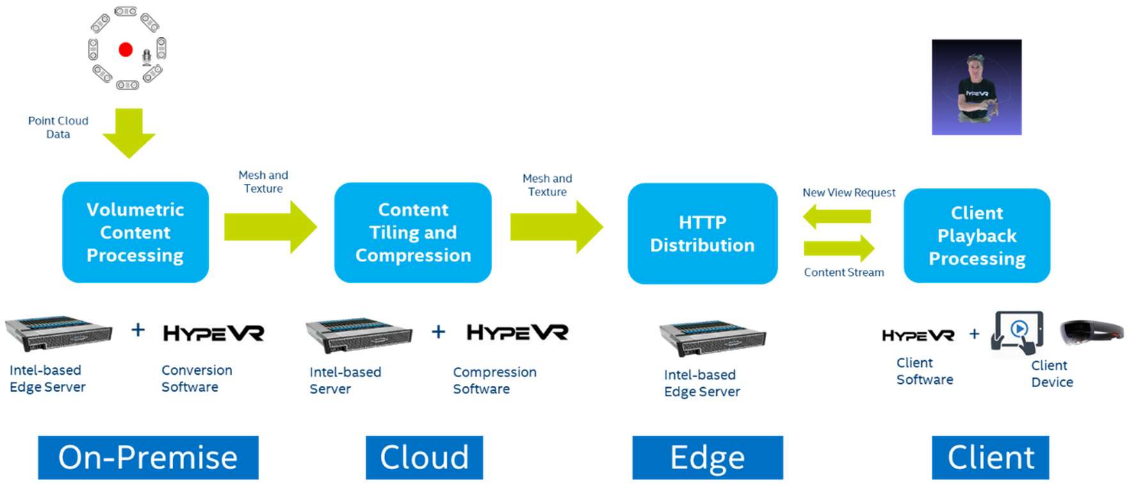
To deliver this content within the network constraints of 5G and conventional Wi-Fi for at-home use, we are targeting a bandwidth of ~100Mpbs per user. To meet this second challenge, we need to reduce the bandwidth requirements of the volumetric content from ~1Gbps down to ~100Mpbs, which is a bandwidth reduction of 10X[1,2,3].

A Volumetric VOD Delivery Framework

To meet these challenges, we built a volumetric streaming framework that delivers high quality volumetric captured video content directly to mobile devices in an AR environment over 5G. Our framework will allow users to view the volumetric content from all angles and walk around the captured subjects. The volumetric streaming framework is capture-agnostic but is designed to stream content that is converted into textured 3D content. We chose this format for ease of adoption on the client side since most render engines, such as Unity and Unreal, can easily accept this format. In addition, the 3D mesh format is effectively a dissemination from the raw camera capture and reduces the bandwidth requirements for streaming this content, depending on the quality and detail of the mesh.

Once the content is created into a series of object, material, and texture files which represent the volumetric video content from frame to frame, we apply our own 3D tiling techniques to the volumetric objects in the scene. This allows us to break up the volumetric content so that only meshes and textures that should be visible to the user will be loaded to their devices for streaming. After we have processed the tiling of the content, we apply a compression technique on the server and send the tiled, compressed content to the edge server/CDN for content distribution.

The edge server and CDN then distributes the tiled compressed content to each user's device based on the user's viewpoint or position. Because volumetric content is true 3D content, the user can walk all around the volumetric objects in Virtual or Augmented Reality space and only see a section of the objects at any given point. This delivery framework focuses on distributing the content that is visible to the user both to reduce bandwidth and to reduce the render requirements on the client side. The graphic below demonstrates the delivery framework.



---

1. Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products.
2. Configurations: [Testing Configuration: Server: Core i7-7700 3.6 GHz, Client: Apple iPad* Pro 3rd generation + Test: Intel Real Sense Captured Volumetric Video + Robert Kawiak + 09/03/2019].
3. For more information go to www.intel.com/benchmarks

Volumetric Compression Methodologies

To deliver this framework in 5G network conditions, we must apply significant compression to the volumetric content. We investigated various 3D compression techniques that can reduce the bandwidth requirements of the volumetric content down to ~100Mpbs[1,2,3]. In addition to the target compression ratio, we must limit our compression techniques to ones that have real time decompression capabilities for the content size we are targeting. In our investigations, we came across two such 3D compression techniques that can meet the requirements outlined above. The first is from a company named HypeVR, which uses a temporally-consistent compression methodology. The second is Draco[1, 2, 3], an open source project from Google, which focuses on spatial compression. In the below section, we will include descriptions of both compression techniques as well as performance data highlighting compression ratios and encode / decode latencies. We are also exploring applying the temporal compression methodology towards texture compression and standard media compression such as JPEG on the texture files.

HypeVR

The Mesha codec [6], developed by HypeVR [6], is a high quality, proprietary, lossy volumetric video codec designed for storing, streaming and playing back volumetric content that has texture resolution up to 4K and mesh quality up to 60K vertices. HypeVR's Mesha codec consists of two independent compression modules; texture compression and mesh compression. The texture compression module is carried out via standard video codecs such H.264, H.265 and VP9, while the mesh compression module is carried through HypeVR's proprietary mesh sequence codec [6]. This mesh codec first quantizes continuous geometric attribute values into discrete values. Next, spatial (within a frame) and temporal (between frames) redundancies, in the quantized attributes of the mesh sequence, are analyzed, extracted, entropy-coded and embedded into a proprietary bit stream format. For a typical temporally coherent volumetric video (consisting of groups of 8 frames on average), the mesh compression ratio is about ~14x. As the length of the temporally consistent frame group increases, the mesh compression ratio increases proportionally[6].

The proprietary bit stream supports frame segmentation, random key frame (intra-frame) seeking, and timestamping. It is also designed with MPEG compatibility in mind. For this reason, both compressed textures and mesh sequences can be packaged into an MPEG container format (MP4) for ease of volumetric video streaming [6]. For content distribution, this means a simple HTTP-based service (which is offered by a wide range of cloud storage services) is sufficient to serve the volumetric VoD to the public.

The Mesha codec is under active development and the next release aims to improve mesh compression by another 30% to 50%. To realize this goal, new inter-frame prediction models for vertex attributes will be developed to better capture inter-frame redundancies. Additionally, more accurate and self-adaptive probability models will be designed to improve symbol coding, and rate-distortion models will be added for generating different quality bit streams [6].

DRACO

Draco is a vector quantization compression methodology based on the Edgebreaker [2, 3] compression scheme, which utilizes a methodology known as triangle traversal encoding[4]. This methodology creates a triangle spanning tree composed of stripes. The algorithm iteratively processes the mesh it is working to encode in a breath first traversal [4]. It uses the vector location information itself, and local and global predictors to “predict” the trajectory of the triangle/vector connectivity, e.g. how the triangles are connected to each other [3, 5]. Because the encoder and decoders use the same formula when the vector data passes through the compression and decompression path, they should ideally have the same pattern of prediction traversal when encoding or decoding [3].

As the encoder compresses, it makes predictions about the way in which the next triangle in the mesh is related to the previous triangles and updates its prediction patterns based on the results of its next pass. The decoder is essentially a reconstruction of the triangle mesh, based on the same predictive algorithm and created on encode, about the way in which the triangles are connected. It compresses the triangle meshes by visiting them in a depth-first order, generates a string of descriptors (one per triangle), about the way in which they are connected to each other. This allows the entire mesh to be “rebuilt” by attaching new triangles to previously reconstructed ones. These descriptors can be represented by 2 bits per triangle and be optimized to be 1.73 bits per triangle [3].

Draco’s Edgebreaker compression scheme is lossless, because it can fully reconstruct the original triangle mesh that was originally assigned to the algorithm. However, Draco does apply an additional quantization compression to the resulting data from Edgebreaker, which has different levels of quantization, and is lossy. The levels to which you can apply the quantization compression used in Draco will be further discussed in the compression results below [1].

Volumetric Video Compression Results

---

1. Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products.
2. Configurations: [Testing Configuration: Server: Core i7-7700 3.6 GHz, Client: Apple iPad® Pro 3rd generation + Test: Intel Real Sense Captured Volumetric Video + Robert Kawiak + 09/03/2019].
3. For more information go to www.intel.com/benchmarks

The sample test case used in this study is content that was shot and constructed from Real Sense camera [7]. The raw point cloud content was captured by HypeVR and post-processed into 3D mesh data represented by OBJ or object files and textures represented by simple JPEG files. The specific sample we used was captured by 8 real sense cameras and has roughly 65K mesh vertices and 2K textures. The content has 900 frames, captured at 30 FPS. This content will be provided along with our volumetric VOD pipeline as sample test content to verify the volumetric streaming pipeline.
In this section, we will illustrate the different compression techniques we investigated, working in conjunction with our tiling techniques and the compression results we obtained from this data. We will illustrate how these two 3D compression techniques, and the compression results, enable us to deliver high-quality volumetric content over 5G and typical in-home broadband. We performed our measurements on a Core i7-7700 3.6 GHz server and an Apple iPad* Pro 3rd generation client device, with the Apple A12X bionic processors. Since the SW works purely on the CPU, there is some run-to-run variation, but both the compression ratios and the decode times we shared below did not show drastic difference from run to run[1,2,3].

The main limitation to enabling our VOD usage, with respect to compression, is the client FPS performance or the client decode performance and server-to-client bandwidth. Therefore, we are choosing to highlight the decode performance and compression ratios of both compression methodologies. One caveat is that our decode time, shown in the tables below, is not simply the decode performance of the compression techniques, but rather a sum of the actual decode time and the time required to convert the data to a format usable by OpenGL, Metal, Vulkan, etc. This time does not include file operations. We are categorizing it in this way because it has a direct impact on the playback frame rate on the client device. Here, we can see that HypeVR has a distinct advantage in decode performance, which fits well with our VOD usage and is able to keep up with the volumetric video framerate we required[1,2,3].. There are optimizations we plan to work on in the future to improve decode performance, such as multi-threading the decoder, and optimizing both the decode and format conversion for Draco. This is currently under investigation.

In the tables below, Draco is shown to have several compression levels, whereas HypeVR is shown to have several GOP levels. For Draco, the compression levels are associated with the quantization compression applied to the data files after Draco has run its triangle traversal compression. Level 0 means no quantization, and level 1, 7 and 10 represent increasing levels of compression [1]. Although quantization is lossy, we were not able to observe significant artifacts created by this quantization, but we do plan to follow up our current study with PSNR impacts of these compression formats[1,2,3]. HypeVR's GOP levels or group of pictures, similarly increases the compression ratio of their Mesha compression technique by increasing the number of frames they group together, and is lossy[1,2,3]. We plan to also extend the PSNR study to HypeVR's compression as well.

In the Draco case, the levels of quantization significantly impact the decode performance on the client device. One thing to keep in mind is that the decode performance shown below is not the only latency that can impact user experience, since there's still a render pass needed on the client device. But, for the type of content we are displaying with 2K textures and ~65K vertices, the render performance on the device (Apple iPad* Pro 3rd Generation) is ~100-150FPS, therefore the overall client device experience has its bottleneck in the decode performance[1,2,3]. In the tables below, we can clearly see the HypeVR has a significant advantage in decode performance[1,2,3].

	Level 0	**Level 1**	**Level 7**	**Level 10**
Draco Decode [ms]	6.54	9.47	20.26	26.00
Draco Decode [FPS]	152.91	105.60	49.36	38.46

	gop=1	**gop=4**	**gop=8**
HypeVR Decode [ms]	3.82	2.68	2.52
HypeVR Decode [FPS]	261.60	372.47	397.19

In compression ratio, Draco has an advantage when the number of triangles are high. If you look at the tables below and look at the "no tiling" section in both the Draco and HypeVR tables, Draco has a much higher compression ratio at with no tiling, or ~65K vertices. However, when you take tiling into account, where each of our tiles may have a significantly lower number of vertices than 65K, you can see the compression ratios start to narrow. Currently, we are testing a 16, 32 and 65 tile solution. In all three cases, each tile will have significantly fewer triangles per tile than ~65k. And, as you can see from the tables below, Draco will have a lower compression ratios at those reduced mesh sizes [1,2,3]. For HypeVR on the other hand, the compression ratios are fairly consistent regardless of mesh size, and this again is consistent with a temporal compression methodology [1,2,3]. And although the compression ratio may be lower than Draco, the decode performance makes it a compelling compression technique for our volumetric VOD usage [1,2,3].

1. Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products.
2. Configurations: [Testing Configuration: Server: Core i7-7700 3.6 GHz, Client: Apple iPad* Pro 3rd generation + Test: Intel Real Sense Captured Volumetric Video + Robert Kawiak + 09/03/2019].
3. For more information go to www.intel.com/benchmarks

		Draco - compression level 1		Draco - compression level 7		Draco - compression level 10	
Number tiles	**uncompressed stream size [kB]**	**compressed stream size [kB]**	**compression ratio**	**compressed stream size [kB]**	**compression ratio**	**compressed stream size [kB]**	**compression ratio**
No tiles	3950971	200968	20	136375	29	134309	29
16	3905315	247295	16	172843	23	174200	23
32	3941841	259119	15	183341	22	185696	21
64	3925993	271351	15	194265	20	197377	20

		HypeVR compression - gop=1		HypeVR compression -gop=4		HypeVR compression - gop=8	
Number tiles	**uncompressed stream size [kB]**	**compressed stream size [kB]**	**compression ratio**	**compressed stream size [kB]**	**compression ratio**	**compressed stream size [kB]**	**compression ratio**
No tiles	3950971	697034	6	340656	12	281527	14
16	3905315	752458	5	371123	11	307843	13
32	3941841	767263	5	379281	10	314887	13
64	3925993	722929	5	358275	11	297723	13

Conclusion

By making the above enhancements to Edge servers, we are able to create a scalable volumetric VOD delivery framework that can support VOD streaming of volumetric video while meeting the motion-to-photon latency for user interactivity, all within the bandwidth constraints of broadband and upcoming 5G services. We are able to show that by applying compression techniques such as HypeVR's Mesha and Google's Draco, along with our tiling technique, we are able to take high-quality volumetric content and make it streamable in a AR environment with minimal motion to photon latency. These results suggest significant efficiency gains for practical volumetric content streaming over broadband and 5G networks, just as cinematographers and sportscasters are exploring these new technologies. The future seems bright here indeed!

---

1. Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products.
2. Configurations: [Testing Configuration: Server: Core i7-7700 3.6 GHz, Client: Apple iPad* Pro 3rd generation + Test: Intel Real Sense Captured Volumetric Video + Robert Kawiak + 09/03/2019].
3. For more information go to www.intel.com/benchmarks

Glossary

AR/VR	Augmented Reality/Virtual Reality
CDN	Content Delivery Network
CES	Consumer Electronics Show, which occurs each January
Cloud	Cloud server
CPU	Central Processing Unit
Edge	Edge server
Entropy	Entropy encoding, a lossless compression that represent frequently used bits with few bits
FPS	Frames Per Second
GHz	Giga Hertz
GOP	Group of Pictures
GPU	Graphics Processing Unit
H.264 & H.265	Video coding standards developed by ITU together with ISO/IEC
HTTP	HyperText Transfer Protocol
Intra-frame	A frame of data that is encoded without referencing other frames
Inter-frame	A frame of data that is encoded by referencing one or more frames
IR	Infrared
JPEG	Joint Photographic Experts Group
Mesh	A set of triangles that are connected by the common edges
MPEG	Motion Pictures Experts Group
MP4	A digital multimedia container format used to store compressed video and audio data
NAB	National Association of Broadcasters, which has their annual conference in mid-April
On Prem	On Premises
OBJ	Object File, a typical 3D image format that contains 3D coordinates, texture maps, polygonal faces, etc.
PSNR	Peak signal-to-noise ratio
QoS	Quality of Service
Quantization	Mapping a large data set to a smaller one
Texture	Surface texture or color information applied on a 3D generated model
Triangle	In 3D Graphics, a construct made of 3 vertices
VCD	Visual Cloud Division
VOD	Video On Demand
Vector	A quantity that has magnitude and direction, represented by a directed line
Vertex	A point where two more or curves meet
Volumetric Video	Video technique that enables capture and playback in 3D space
VP9	Video coding standard developed by Google

References

1. Google Inc. and other Authors, Draco 3D Data Compression, https://google.github.io/draco/, Accessed 09 Sept 2019
2. Rossignac, J.: Edgebreaker: connectivity compression for triangle meshes. IEEE Trans. Vis. Comput. Graph. 5, 47–61 (1999)
3. Rossignac, J., Safonova, A., Szymczak, A., 3D Compression Made Simple: Edgebreaker on a Corner-Table.
4. Maglo, A., Lavoue, G., Dupont, F., Hudelot, C.: 3D mesh compression: survey, comparisons and emerging trends, ACM Computing Surveys, Vol. 9, No. 4, Article 29, September 2013
5. Devillers, O., Gandoin, P.M., Geometric Compression for Interactive Transmission. Research Report – 3910, INRIA. 2000.
6. HypeVR Inc., Mesha Codec https://www.hypevr.com/
7. Intel Inc., Real Sense Camera https://www.intelrealsense.com/, Accessed 20 July 2019

1. Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products.
2. Configurations: [Testing Configuration: Server: Core i7-7700 3.6 GHz, Client: Apple iPad* Pro 3rd generation + Test: Intel Real Sense Captured Volumetric Video + Robert Kawiak + 09/03/2019].
3. For more information go to www.intel.com/benchmarks

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors.

Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products.

For more information go to www.intel.com/benchmarks.

Performance results are based on testing as of 09/03/2019 and may not reflect all publicly available security updates. See configuration disclosure for details. No product or component can be absolutely secure.

Configurations: [Testing Configuration: Server: Core i7-7700 3.6 GHz, Client: Apple iPad* Pro 3rd generation + Test: Intel Real Sense Captured Volumetric Video + Robert Kawiak + 09/03/2019].

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Performance varies depending on system configuration. Check with your system manufacturer or retailer or learn more at [intel.com].

No license (express or implied, by estoppel or otherwise) to any intellectual property rights is granted by this document.

Intel disclaims all express and implied warranties, including without limitation, the implied warranties of merchantability, fitness for a particular purpose, and non-infringement, as well as any warranty arising from course of performance, course of dealing, or usage in trade.

This document contains information on products, services and/or processes in development. All information provided here is subject to change without notice. Contact your Intel representative to obtain the latest forecast, schedule, specifications and roadmaps.

The products and services described may contain defects or errors known as errata which may cause deviations from published specifications. Current characterized errata are available on request. No product or component can be absolutely secure.

Copies of documents which have an order number and are referenced in this document may be obtained by calling 1-800-548-4725 or by visiting www.intel.com/design/literature.htm.

---

1. Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products.
2. Configurations: [Testing Configuration: Server: Core i7-7700 3.6 GHz, Client: Apple iPad* Pro 3rd generation + Test: Intel Real Sense Captured Volumetric Video + Robert Kawiak + 09/03/2019].
3. For more information go to www.intel.com/benchmarks