# High Quality 360° Video Rendering and Streaming

Louay Bassbouss[1], Stephan Steglich[1], Martin Lasak[1]

[1]Fraunhofer FOKUS, Berlin, Germany

E-mail: [1]{firstname.lastname}@fokus.fraunhofer.de

*Abstract:* **This paper introduces a new solution that facilitates the deployment and consumption of immersive and interactive media by ensuring optimal network delivery and media playback. We will address in this paper two main challenges a) the efficient streaming of high quality 360° video content using existing content delivery networks (CDNs) and without the need for additional bandwidth comparing to traditional video streaming and b) the playback of 360° content even on devices with limited processing resources and programmatic capabilities.**

**Keywords:** 360° Video, VR, Immersive Media, Streaming, HbbTV, MSE, Cloud Rendering

## 1 INTRODUCTION

Immersive video has been around for some time, dating back to the “A Tour of the West” short movie from 1955 [1] and the ‘Circarama’ format (also known as “Circle-Vision 360°” [2]). It re-emerged a couple of times, though mostly as a showcase exhibit at trade fairs and fairground rides and often more as a spin-off of interactive VR experiences than as real-world video presentation in its own right. Only recently did the market situation change. Affordable cameras with sufficient resolution became available to allow professionals and interested amateurs to create 360° movies. Stitching software became good enough to stitch the videos and hide the seams with reasonable quality. Networks became fast enough to allow end-users to stream 360° video content to their devices. TVs, smartphones and tablets are sufficiently powerful and have the necessary sensors and quality to handle the content and react on view changes without noticeable delay. After 60 years, immersive video has reached the mass market.

Most efforts in these area, however, have been aimed at the technical challenges creating and viewing of 360° video. As 360° video is starting to reach a wider audience, the need arises to pay attention to the use of such content in a realistic commercial environment. For this, two issues need to be addressed. The efficient distribution of 360° content and the added-value that it can bring content providers.

On the distribution side, almost all current solutions stream the full 360° content to the end-user device, where only about 8% is actually presented to the viewer, while the other 92% are disregarded, causing a huge bandwidth requirement. (While the actual amount depends on the video mapping (usually by “Equirectangular Projection” [3]), the video codec, viewing angle and the area looked at by the user, on average the viewer sees about 1/12 of the available sphere). This currently means that a 4k UHD (3840x2160) [11] 360° source video provides approximately HD (1280x720) [12] output video resolution for a field of view of 60° vertically and 106,67° horizontally (106,67°=60°x16/9 to keep 16/9 aspect ratio. The aspect ratio is relevant to consider on flat screens but not necessary on head-mounted displays). Figure 1 shows an example. Inversely, to allow the end user to experience 360° content in 4k UHD field of view resolution, the source 360° video must have a resolution of 11520x6480 (between 8k and 16k).

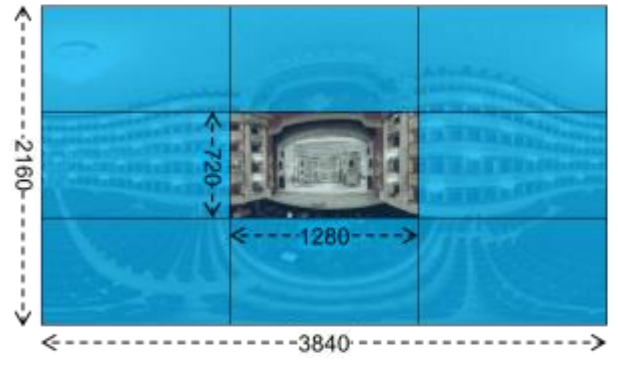


**Figure 1: 4k UHD source video results in HD output field of view.**

## 2 360° VIDEO RENDERING AND STREAMING

The key idea of the solution in this paper is to render 360° videos in the cloud and stream individual field of views to the client instead of streaming the whole video. Figure 2 shows the three possible rendering and streaming options:

1) **360° Client Rendering**: the whole source video will be delivered to the client. A 360° renderer processes the video on the target device and displays the requested field of view.
2) **360° Cloud Rendering**: the source 360° video will be live rendered on the server and only the requested field of view will be streamed to the client. The client is a simple video player. The server needs to start a session for each client.

**Corresponding author:** Name, Affiliation, Address, Phone, E-mail

3) **360° Cloud Pre-Rendering**: from client point of view, it is the same as option 2). The difference is on the server. In a pre-processing step, the server prepares and stores the video content is a way that no video processing is needed anymore.

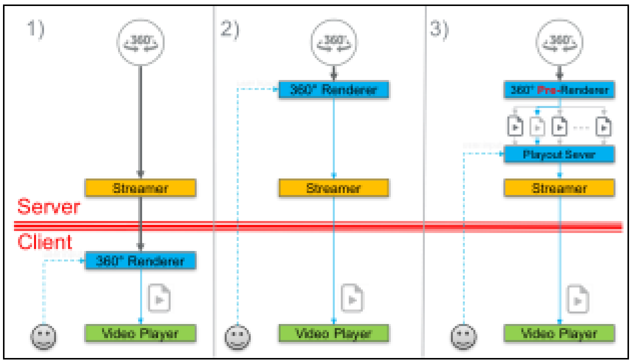


**Figure 2: 360° Rendering and Streaming Options**

Each of these options will be discussed separately in the following subsections.

## 2.1 360° Client Rendering

360° Client Rendering is the most widely used option in existing 360° video players like Facebook [5] and YouTube [6] players. When using this option, the whole 360° video will be delivered to the client and the 360° video player needs to process each received frame and calculates the requested field of view. Currently, 360° video playout is not distinguished from traditional video playout. It uses the same content distribution servers and network protocols as other video formats. It is a known problem that more than 90% of 360° video content is not in the visual field of an individual user - meaning that most of the video content streamed over the network to the user is discarded. This problem is specific to 360° video, as most VR content is rendered only for the areas the users can see. Furthermore, the device platform must offer APIs to process each frame of the 360° video stream and render the field of view. The required processing resources for a smooth video playback depends from different factors like video resolution, encoding and 360 projections. The advantages and disadvantages of this option are summarized below:

\+ Existing Content Delivery Networks (CDNs) [8] can be used without any modification. This means that video processing is not required on the server.

\+ Low motion-to-photon latency (Motion-to-Photon latency is the time needed for a user movement to be fully reflected on a display screen [7]).

– Significant additional bandwidth is required.

– Significant additional processing resources on the client are required.

– Programmatic interfaces for 360° video processing and rendering are required.

– Additional hardware requirements in terms of GPU [10] in order to decode and process original video content. Only new devices currently available on the market support 4k UHD [11], which means only a 720x400 output video resolution.

This option is suitable for devices where the user interaction requires a low motion-to-photon latency. This is the case for example on head-mounted displays where motion sensors are used to capture the view port coordinates. "*Low motion-to-photon (< 20ms) latency is necessary to convince the mind of the user that he is in another place (Presence)" [7]. "A high motion-to-photon latency makes a poor virtual reality experience and causes motion sickness and nausea*" [7]. As depicted in Figure 2 option 1), the captured view port coordinates are processed locally and will be not sent over the network to a server. This means the photon-to-motion latency depends only from the processing capabilities of the target device. In other words, the 360° video player needs to capture sensor inputs, process a frame of the source video and calculates the projection and display the output frame of the corresponding field of view in less than 20ms.

## 2.2 360° Cloud Rendering

To reduce the bandwidth requirements, Fraunhofer FOKUS offers a solution [9] that renders 360° videos in the cloud and streams only individual field of views (FOVs) to clients like TVs, PCs and mobile devices. Another advantage of this solution is that there are no additional requirements in terms of processing and programmatic capabilities on the client since the 360° video processing and view calculation happens in the cloud. The client is a simple video player and needs only to play a live stream that contains an individual field of view at a specific time.

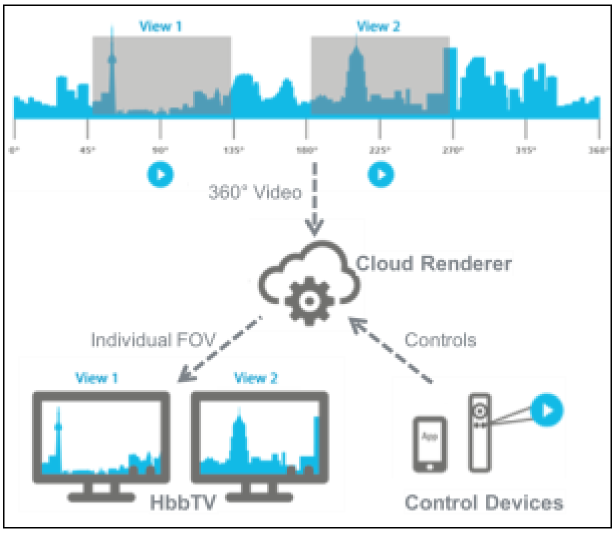


**Figure 3: 360° Cloud Live Rendering**

Hybrid TV Terminals like Hybrid broadcast broadband TVs (HbbTV) [4] fall into the category of devices that are suitable to use this solution as depicted in Figure 3. Devices as HbbTV Terminals are not capable, in terms of programmatic features, of performing the necessary image transformations for rendering 360° video content. Our primary goal at Fraunhofer FOKUS is to bring innovative video experience on traditional TV screens. Our activity is driven by the huge interest of content providers especially from broadcasters to bring 360° video experience to TV sets. HbbTV is one of the options broadcasters can use to offer this kind of features together with the broadcast service. Viewers can experience video content with freely selectable views on their primary video viewing device using the TV remote control or a companion screen application. All video controls are sent to the Cloud Renderer either using a companion screen application or via the HbbTV application in case of using the TV remote control. The advantages and disadvantages of this option are summarized below:

- \+ No additional bandwidth is required comparing to traditional video streaming.
- \+ No processing resources are required on the client.
- \+ Any video player can be used. There is no need for additional APIs to process the video.
- \+ No special hardware requirements comparing to traditional video playback.
- ± Since all user commands are sent to the cloud renderer, the motion-to-photon latency depends from network latency and other factors like buffered amount of video data.
- – Existing CDNs cannot be used since the server renders the field of view for each client in a separate session and the output stream is delivered to the client over a persistent connection.
- – The 360° video rendering requires a lot of processing resources which have a direct impact on scalability.

## 2.3 360° Cloud Pre-Rendering

The 360° Cloud Pre-Rendering option solves the scalability issue of the 360° Cloud Rendering described in previous subsection. Instead of live-rendering of the 360° source video for each client in a separate session, the source video will be pre-rendered in a pre-processing step. The 360° Cloud Render from previous subsection can be used for pre-rendering. The output FOVs for different combinations of viewport coordinates will be made available in a video storage as depicted in Figure 4. Therefore, the 360° Cloud Playout Server needs only to prepare the video stream for the requested viewport just by reading the video data from corresponding FOV files available in the video storage and collecting the segments into a live stream which will be delivered to the client. The 360° video player is the same as in option 2).

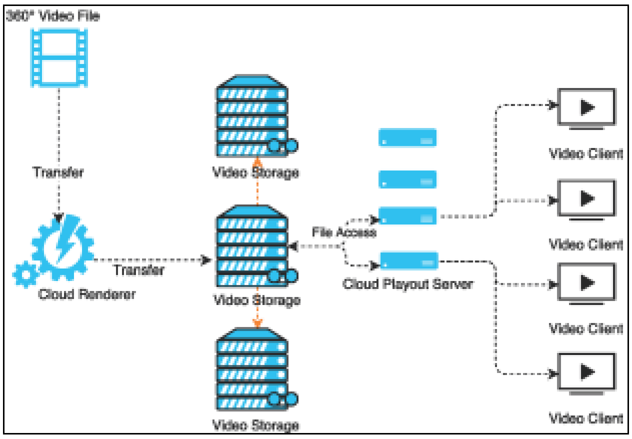


**Figure 4: 360° Pre-Rendering**

The advantages and disadvantages of this option are summarized below:

- \+ No additional bandwidth is required comparing to traditional video streaming.
- \+ No processing resources are required on the client.
- \+ Any video player can be used. There is no need for additional APIs to process the video.
- \+ No special hardware requirements comparing to traditional video playback.
- \+ No video processing on the server is needed during streaming.
- ± Motion-to-photon latency depends from different factors like network latency and segment length.
- – Existing CDNs cannot be used because each client has its own persistent streaming connection to the server.
- – Additional cloud storage for the pre-rendered video files is needed. The amount of additional storage depends from the overlapping factor during pre-rendering.

## 2.4 360° Video Playback on Web Clients

The Fraunhofer FOKUS video solution is not only applicable for HbbTV terminals, but also for any device that runs a web browser like desktop, mobile and streaming devices (e.g. Chromecast, Android TV, Amazon Fire TV, etc.). By using existing W3C APIs like Media Source Extension (MSE) [13], we have implemented the entire logic of the 360° player in the browser which is currently not possible in HbbTV due the missing MSE API. In a proof-of-concept implementation we were able to play 360° video in any browser that supports MSE. The source 360° video is pre-rendered using the “360° Cloud Pre-Renderer” described in section 2.3 and the output video files are made available on a cloud storage like Amazon Simple Storage Service (S3) [14]. The prepared pre-rendered video files contain segments for different combinations of field of views and video qualities. Existing content delivery networks (CDNs) can be used to host and stream the pre-rendered videos without any additional requirement comparing to traditional video streaming. Figure 5 shows an example how the MSE 360° player switches to the next field of view. The player can first switch to the new field of view after the playback of current segment of the current field of view is completed. A segment consists of a group of pictures (GOPs) [15] starting with an image frame (I-Frame) and followed by a number of difference frames (P-Frame, B-Frame or D-Frame). Video decoders can decode a difference frame if only all previous frames in the segment are decoded. This is why a switch is possible only at the end of a segment.

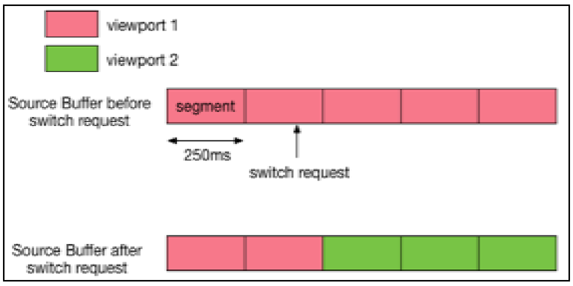


**Figure 5: MSE 360° Video Player**

With all the advantages of this solution, the limitation regarding motion-to-photon-latency which depends from different factors like network latency, buffering strategies and segment duration still exists. The 20ms maximum latency in case of head-mounted displays is difficult to achieve.

## 3 CONCLUSION AND OUTLOOK

In this paper we presented a new solution to render and stream 360° videos without the need for additional bandwidth, processing resources and programming interfaces comparing to traditional video streaming. We also introduced a 360° video player for web clients that uses the W3C MSE API which allows us to implement the entire player logic on the client and takes advantage of content delivery networks. A proof-of-concept implementation for all options and components mentioned in this paper are available. As next step, we will bring this technology to most relevant devices and platforms and improve the user experience by reducing the motion-to-photon latency using intelligent caching algorithms that takes many parameters into consideration like FOV, available bandwidth and screen resolution of target devices. We will also start soon a pilot together with broadcasters addressing HbbTV.

## References

[1] “A Tour of the West” (1955, USA), Director Peter Ellenshaw, Production Co. Walt Disney Pictures, The Internet Movie Database, Retrieved 10 November 2016, from http://www.imdb.com/title/tt0048742/.

[2] “Circle-Vision 360°”, Wikipedia: The Free Encyclopedia. Wikimedia Foundation, Inc., Retrieved 10 November 2016, from https://en.wikipedia.org/wiki/Circle-Vision_360%C2%B0.

[3] “Equirectangular projection”, Wikipedia: The Free Encyclopedia. Wikimedia Foundation, Inc., Retrieved 10 November 2016, from https://en.wikipedia.org/wiki/Equirectangular_projection.

[4] “Hybrid Broadcast Broadband TV” (HbbTV), Retrieved 10 November 2016, from https://www.hbbtv.org.

[5] “Facebook 360 Video”, Retrieved 10 November 2016, from https://facebook360.fb.com/.

[6] “Google VR”, Retrieved 10 November 2016, from https://vr.google.com/jump/

[7] “What is Motion-To-Photon Latency?”, Retrieved 10 November 2016, from http://www.chioka.in/what-is-motion-to-photon-latency/.

[8] “Content delivery network”, Wikipedia: The Free Encyclopedia. Wikimedia Foundation, Inc., Retrieved 10 November 2016, from https://en.wikipedia.org/wiki/Content_delivery_network.

[9] “Fraunhofer FOKUS Cloud-based 360° Video Playout for HbbTV”, Retrieved 10 November 2016, from https://www.fokus.fraunhofer.de/go/360.

[10] “Graphics processing unit”, Wikipedia: The Free Encyclopedia. Wikimedia Foundation, Inc., Retrieved 10 November 2016, from https://en.wikipedia.org/wiki/Graphics_processing_unit.

[11] “4K resolution – Ultra HD”, Wikipedia: The Free Encyclopedia. Wikimedia Foundation, Inc., Retrieved 10 November 2016, from https://en.wikipedia.org/wiki/4K_resolution#Ultra_HD.

[12] “High-definition video”, Wikipedia: The Free Encyclopedia. Wikimedia Foundation, Inc., Retrieved 10 November 2016, from https://en.wikipedia.org/wiki/High-definition_video.

[13] “W3C Media Source Extensions”, Retrieved 10 November 2016, from https://www.w3.org/TR/media-source/.

[14] “Amazon Simple Storage Service (S3) - Cloud Storage”, Retrieved 10 November 2016, from https://aws.amazon.com/s3/.

[15] “Group of pictures”, Wikipedia: The Free Encyclopedia. Wikimedia Foundation, Inc., Retrieved 10 November 2016, from https://en.wikipedia.org/wiki/Group_of_pictures.