INTERACTIVE STREAMING OF PANORAMAS AND VR WORLDS

R. Schäfer, P. Kauff, R. Skupin, Y. Sánchez, C. Weißig

Fraunhofer Heinrich Hertz Institute, Germany

ABSTRACT

Virtual Reality (VR) has recently become a hot topic in the media industry. VR systems enable the user to navigate in real or virtual worlds. The dimensions of these worlds may range from 180 degree cylinders up to complete 360 degree spheres. Such systems require cameras, which are able to capture these worlds and transmission systems, which enable users to view these worlds on their VR devices. Fraunhofer HHI has developed an interactive streaming system, which provides highest quality and does not limit the image resolution neither on the production nor on the end device side. It consists of one or several omnidirectional cameras, which can be placed into a scene like sport stadiums, theatres, fairs etc. In HHI's OmniCam360 there are 10 micro HD cameras, delivering a 360 degree panorama of 10.000 x 1.920 pixels. As the data rate of such video is by far too high to be transmitted to an end device, three solutions to stream the content to end devices have been developed. A user can then interactively navigate in the panorama.

INTRODUCTION

In recent years, patterns of media consumption have been changing rapidly. Video material is now viewed on screens ranging in size from an IMAX cinema, through to large domestic projection and flat-panel displays, down to tablet PCs and mobile phones. At the same time the resolution of displays is constantly increasing, 4k displays are already state of the art and can be bought at affordable prizes. Even first 8k devices appear on the market and this trend will surely continue.

Another significant change in media consumption habits is the level of interactivity that consumers are increasingly expecting. With web-based media it is commonplace to scroll to parts of a web page that are of particular interest, or to use Google Earth to examine a particular part of the world in detail. In a 'first person shooter' computer game, the player can look around in all directions, and expects the soundscape to rotate to match his viewpoint. This level of control has not been possible with traditional video-based media, where the program director has generally determined the view of the scene with which the user is presented.

These general trends are the cradle of the recent hype in Virtual Reality (VR), which has been triggered by the acquisition of Oculus Rift by Facebook and by new VR devices like Gear VR from Samsung and others from Microsoft, Sony, Razer etc. Industry analyst firm

CCS Insight has just published a report – Augmented and Virtual Reality Device Forecast, 2015-2019 – stating the amount of AR and VR devices sold will rise from 2.5 million this year, to 24 million in 2018 (1). It forecasts that the market will be worth more than $4 billion.

VR systems enable the user to navigate in real or virtual worlds. The dimensions of these worlds may range from 180 degree cylinders up to complete 360 degree spheres.

On the production side a number of new 180 - 360 degree cameras have already been launched or have been announced. They either combine a number of cameras to scan panoramas or the complete 360 degree surrounding or they use single cameras with wide angle lenses (e.g. fisheye) or curved mirrors (e.g. parabolic front mirror). Single cameras are easy to handle, however their resolution is limited which results in rather poor image quality. If more than one camera is used, a number of technical issues occur: The cameras have to be synchronized, parallax errors may occur, different sensitivities of the cameras have to be compensated and the images of the single cameras have to be stitched together. In addition, the resulting video format may become very large, which results in problems for viewing, storage and transmission.

Fraunhofer HHI has developed a complete interactive streaming system, which overcomes all the above mentioned problems. It consists of one or several omnidirectional cameras (OmniCam360), which can be put into a scene like sport stadiums, theatres, fairs etc. An OnmiCam360 consists of a number for Micro HD cameras which cover a desired viewing angle. In HHI's OmniCam360 there are 10 micro HD cameras, delivering a 360 degree panorama of about 10.000 x 1.920 pixels.

As each camera delivers an HD frame in portrait format, the frames of all cameras have to be combined. This processing is done by a so called Real Time Stitching Engine (RTSE), which can also render a number of HD frames from arbitrary parts of the panorama with arbitrary zoom factors, covering either the complete panorama or parts of it.

After stitching, the panorama can be delivered to fixed or mobile devices. As the data rate of a 10.000 x 1.920 video is by far too high to be transmitted to an end device, there mainly two ways to solve this problem: 1) The complete panorama is sent to the edge of the network and the end device sends the coordinates and the zoom factor of the selected sector directly to the edge server, where this sector is rendered to a HD video. 2) The panorama is split into a number of tiles of different resolutions and the client request the tiles of the selected sector from the server. Both solutions will be explained in this paper.

PANORAMA CAMERA SYSTEM

As known from computer vision and projective geometry, the optimal multi-camera arrangement for capturing panoramic videos requires that the focal points of all camera views coincide in a common point (e.g. central point of a cylinder or sphere; see Figure 1, left) (1). In case of capturing static 2D panoramas, this condition is usually achieved by rotating a single camera at a tripod with a revolving camera head around the entry pupil of the lens. For video, however, this approach

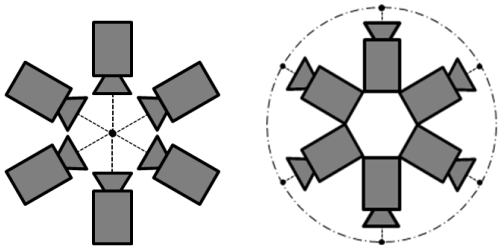

Figure 1 – Possible camera arrangements ; left: optimum from theory; right: star-like approach

is impractical due to the need of using multiple cameras simultaneously on one hand and the physical dimensions of each camera on other hand (apart from the fact that the cameras would shoot each other in the most cases). Hence, many commercial solutions capture video panoramas with a star-like approach (see Figure 1, right) (3,4). In this case, the focal points of all cameras are located on a common circle, where the optical axes are located perpendicular to the arc at the radial axes of the circle. This approach works reasonably well as long as only far distant objects or objects close to one particular plane (e.g. playing field in sports arena) appear in the scene. However, the existence of a non-zero parallax angle does not allow seamless stitching in case of close objects in the overlap area (e.g. a close object moves from camera view to the next in front of a long-distant background).

A more sophisticated solution can be achieved by using special mirror-rigs. In this case all cameras look, for instance, bottom-up into a pyramid shaped mirror with 45-degree tilt angles. If in this case all cameras and mirrors are aligned and calibrated correctly, it is possible to superimpose the virtual images of all focal points in one common central point behind the mirrors. Since the first applications in the 60s, many further system approaches have been proposed and since then a considerable progress has been made due to the advent of digital cameras and digital processing capabilities (2, 3, 4).

Figure 2 – OmniCam360

One of the most recent high-quality systems, the OmniCam360, has been developed by HHI. In its current version it is equipped with 10 micro-HD cameras used in portrait format. Hence, after stitching the 360 degree panorama ends up with a total resolution of 10.000 x 1.920 pixels. Due to the usage of the portrait format, the vertical field-of-view is about 60 degree, a feature that is extremely useful for many applications in the field of immersive media. Compared to other mirror-based solutions, its form factor is relatively small (50x50 cm) and it is quite light weight (25 kg). Although the current version is equipped with 10 cameras and 360 degree, the general concept behind it is open and scalable. If desired, the same concept can be used for other types of cameras (e.g. 4k micro cameras or digital cinema cameras) or for any panoramic format between 120 and 360 degrees.

A special property of OmniCam360 is its very accurate calibration capability. The illustration in Figure 3 depicts a horizontal section through the mirror pyramid at the plane where the optical axes intersect the mirror surfaces and, with it, how the virtual images of the focal points are located behind the mirrors. Note that the cameras look bottom-up and that the mirrors deflect the optical axes horizontally in radial direction. In a first step the rig is calibrated such that all virtual images of the focal points coincide in the center point C of the mirror pyramid (see Figure 3, left). This initial state refers to the optimal camera arrangement from Figure 1 left. Although this initial and optimal state allows a parallax-free stitching for scenes with a depth range from zero to infinity, it is not really suitable under real working conditions. If all cameras have a common focal point in the center of the mirror pyramid, there would be no overlap between the different tiles due to a hard cut at the mirror edges. Hence, there is no possibility to blend pixels between adjacent image

tiles. In order to obtain at least some overlap, the focal points of the cameras are then moved slightly out of the center in radial direction (Figure 3, right). By this off-center shift it becomes possible to regulate a scene-adaptive trade-off between sufficient overlap for blending and parallax-free stitching. In practice, the system is usually operated with a radial-shift of about 5 mm, resulting in a blending area of about 10 pixels and a parallax-free stitching of scenes in a range from about 1 meter to infinity. This adjustment is achieved by mechanical means in the construction phase. As long as cameras, optics and off-center adjustments are kept unchanged, the OmniCam360 is only calibrated once and no further re-calibration is needed at the set.

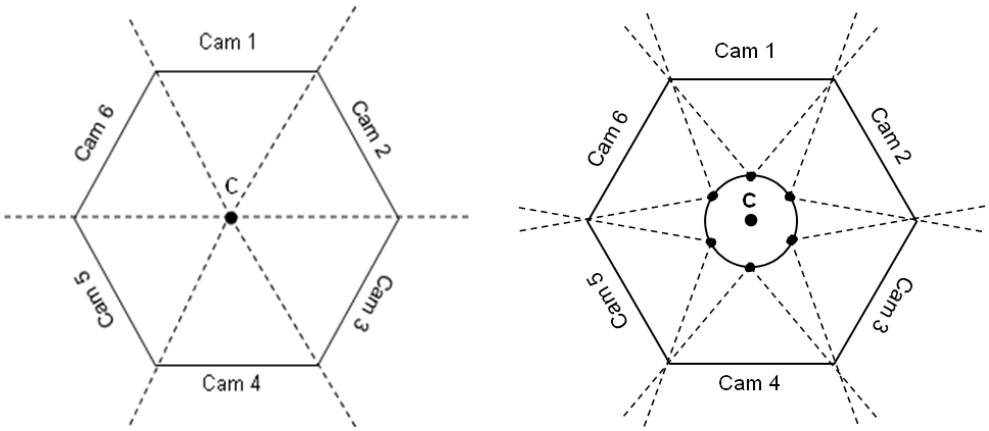

Figure 3 – Optimal mirror-based arrangement (left), radial off-centered arrangement (right)

Figure 4 shows an example of the whole OmniCam360 processing for a sports production with a particularly high depth range of the captured scene at the outer left and right blending areas. All the necessary processing steps are performed in real time by the Real Time Stitching Engine (RTSE) developed by HHI.

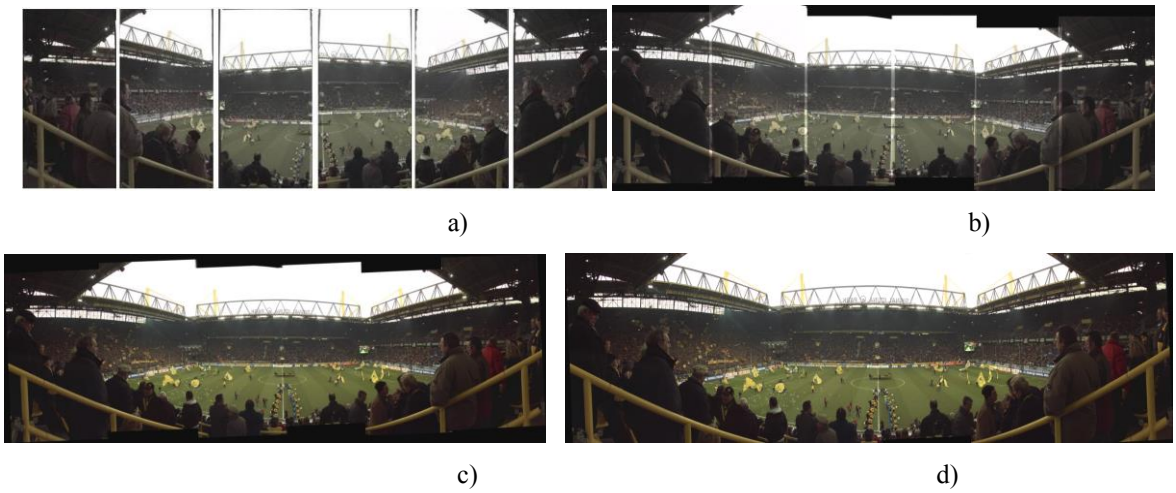

Figure 4 – Processing steps of panorama generation with OMNICAM: a) original camera views; b) geometrical correction and warping; c) photometrical correction, colour matching and blending; d) final cut-out of panoramic view by cropping.

Production examples

During the last two years, a large number of content has been produced with the OmniCam360 (see Figure 5). Different genres have been covered such as sport events, classical and pop concerts as well as documentary films. Some of the highlights are:

Sports:

- FIFA World Cup Qualification (Cologne)
- FIFA World Cup final (Rio de Janeiro)
- ESPN X-Games (Munich)

Pop music:

- Countig Crows, UK Tour (London)
- Bon Jovi, 'Because We Can' Australian Tour (Brisbane)
- Herbert Grönemeyer, 'Dauernd Jetzt' Deutschland Tour (Berlin)

Classical music:

- Berliner Philharmoniker, Anniversary concert “50 years of the Berlin Philharmonie”
- Berliner Philharmoniker, Concert on the occasion of the 25th anniversary of the fall of the Berlin Wall
- Rundfunkchor Berlin, Human Requiem, Neue Nationalgallerie Berlin
- Rundfunkchor Berlin, LOVER, Kraftwerk Berlin

Figure 5 – Classic, pop and sports are typical fields of applications

PANORAMA STREAMING SYSTEMS

As already mentioned in the introduction, panorama streaming systems allow the transmission of huge video worlds. As normal transmission pipes are much too narrow, to stream the complete panorama to the end device, only the Region of Interest (RoI) is transmitted (Figure 6). Depending on the end device, the RoI may be a fixed window (e.g. an HD frame), which scans the panorama. Such an operation mode is typically used in VR devices like Oculus Rift or Samsung VR Gear. If the

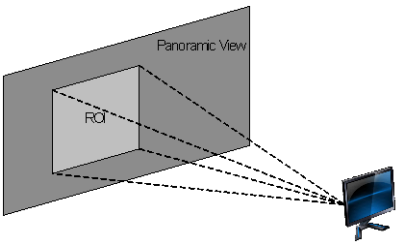

Figure 6 – Region of Interest (RoI) in a panorama

end device is a smart phone or a tablet, it is desirable to have a zoom function, which allows to zoom out, such that the RoI covers the complete panorama or to zoom in, such that the RoI covers only a small portion of the panorama. In both cases, the RoI is displayed in a fixed format, e.g. as an HD video. Now the problem is, to deliver the RoI to the end device and to allow a smooth navigation, without transmitting the entire panorama in its original resolution. In principal there are two solutions to this problem. The first consists in rendering the RoI at the transmitting side and to transmit the RoI as HD video to the end device. The second solution consists in subdividing the panorama into tiles, which may have different resolutions and may overlap. All the tiles are then streamed in parallel as HD videos to the CDN, e.g. using HTTP streaming, and the end device picks up those tiles, which it wants to display.

Server-side rendering

The end device sends the coordinates and the zoom factor of the RoI to the renderer, the selected window is rendered as an HD video and the HD video is streamed to the end device (Fiure 7). This is a quite simple system, however it is a quite expensive solution, as each end device needs its own renderer at the transmitting side. Such a rendering server may be placed in the edge of a mobile network, because in a given cell only a limited number of users may use such a service. However, a network operator would need to install rendering servers in his complete network, which will be quite costly. Such a system has been developed by Fraunhofer HHI and demonstrated at the Mobile World Congress 2013 for the first time. It is also used in the OmniCam system itself to feed a tablet as mobile control monitor, because a visual inspection of all parts of the panorama is easily possible with this streaming method.

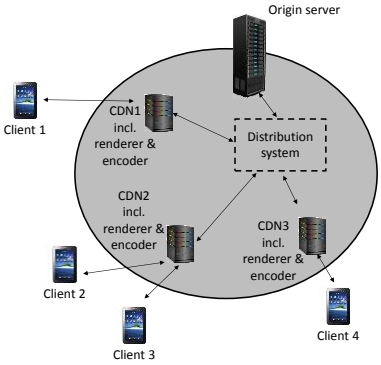

Figure 7 – Distribution network with server-side rendering

Tiled Streaming

Another approach consists in dividing the complete panorama into tiles of e.g. HD resolution and to stream all the tiles in parallel to the network using HTTP streaming. The end device then selects and decodes only those tiles, which are to be displayed (Fig. 8, top). The problem of such a solution consists in a smooth navigation and in zooming, because in most cases the desired RoI does not exactly correspond to a tile, but it will cover an area coverings parts of up to four tiles (Fig. 8, middle). In the case of zooming, the viewing window may even cover a large amount of tiles or even the complete

Figure 8 – Tiling of panorama (top); RoI covering 4 tiles (middle); zoomed-in RoI covering multiple tiles (bottom)

panorama (Fig. 8, bottom). However, most end devices can only decode one or two streams simultaneously. The solution to this problem is the usage of overlapping tiles, as shown in Fig. 9. In addition, different resolution layers will be transmitted, all rendered in the same format as the basic tiles (Figure 8, top).

However, this may result in a quite large number (>100) of tiles, and an equal number of renderers and encoders on the transmitting site and an equal number of streams distributed to the CDN. But even with overlapping tiles and different resolution levels smooth navigation with continuous zoom is not possible, therefore some compromises have to be made and the overall optimisation of the system is quite complex. Such a system has been implemented by HHI and will be presented at the IBC show at the Fraunhofer booth in hall 8.

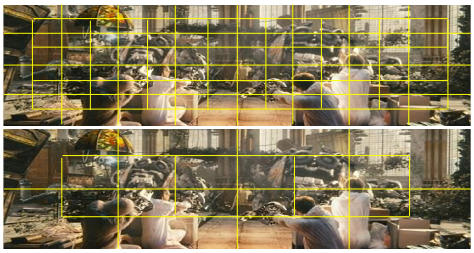

Figure 9 – Overlapping tiles (top) and overlapping tiles of different resolution (bottom)

Compressed Domain Tile Stitching

Building upon the Tiled Streaming approach, a third approach is Compressed Domain Tile Stitching, which allows using a single decoder instance in the end device. As most target end devices are not capable of decoding a multitude of tiles in parallel, the former approach usually relies on encoding a massive amount of overlapping tiles. A given user can then find a suitable panorama tile for the desired RoI.

Through using a set of smaller tiles to coverthe RoI, overlapping tiles are no longer required and hence, server-side encoding complexity is reduced. This technique relies on H.265/HEVC, which introduces tiles as a coding tool for parallelization. An HEVC picture can be divided into a grid of independent tiles. Vice versa, a picture can be stitched together from individually encoded videos given that they fulfil a small set of encoding constraints. The underlying lightweight stitching technique operates on coded bitstreams of neighbouring tiles and fuses them into a single picture within a common bitstream as illustrated in Figure 10 using horizontal tiling. Without entropy decoding and apart from minor high-level syntax adjustments, this technique allows to copy the largest part of the original bitstream, i.e. actual entropy coded slice data, into the new common bitstream without modification.

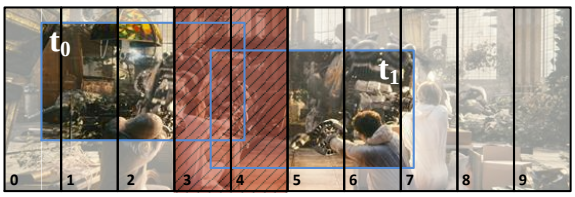

Figure 10 – Stitched Panorama tiles.

As the stitching operation is comparatively lightweight, a server should easily be able to serve a multitude of users with individual bitstreams. Likewise, such processing could also be carried out at the client side after transport, which allows to use established adaptive transport schemes such as MPEG-DASH. The resulting bitstreams are HEVC standard compliant and produce the desired RoI when fed to a standard HEVC decoder.

A challenge for such an approach is to deal with changes of the RoI within the panorama, e.g. through user interaction as depicted in Figure 10 using blue rectangles to indicate the RoI at time instants t_0 and t_1. In the classic Tiled Streaming approach described before,

panorama tiles that are newly encompassed in the RoI have to provide intra-coded random access at high bitrate cost, while the tiles that remain within the RoI can use efficient temporal prediction. Tile 3 and 4 in Figure 10 belong to the latter category. Thus, most of the times, only a subset of the RoI tiles contribute to bitrate peaks at such RoI change events. Using Compressed Domain Tile Stitching, all stitched tiles change spatial position within the RoI during such an event. Therefore, random access usually would have to be provided for all tiles leading to a significant peak bitrate increase compared to classic Tiled Panorama Streaming without stitching. In order to resolve this disadvantage, Fraunhofer HHI has developed a lightweight processing that operates in the Compresssed Domain (5). Through inserting artificial reference pictures into the bitstream at RoI change events with low complexity and for almost zero bitrate cost, tiles that remain in the RoI, e.g. Tile 3 and 4 in Figure 10, can use temporal prediction as in the classic Tiled Panorama Streaming approach. Hence, amongst other advantages, the peak bitrate behaviour using Compressed Domain Tile Stitching with a single decoder instance on the end device is comparable to classic Tiled Panorama Streaming that requires multiple decoders.

CONCLUSIONS

Virtual reality is currently one of the hottest topics in the media sector. Besides complete CGI panoramas, which can easily be produced with appropriate graphics platforms and rendering software. However, there is a great interest in natural content picked-up by cameras, but most of the camera systems available today suffer from poor image quality caused by low resolution or parrallax errors. HHI has developed a mirror based camera system, which overcomes these problems. It consists of 10 professional micro HD cameras which deliver a parallax free panorama of 10.000 x 1.920 pixel. Stitching is done completely automatically in real time by HHI's RTSE. Several productions including the final of the FIFA world Championship in Brasil have been made with this system during the last two years. In addition, HHI has developed two types of panorama streaming systems, which allow to transmit these panoramas at reasonable bit rates to smart phones, tablets or VR devices and which enable the user to navigate in the panoramas. A third system based on HEVC and Compressed Domain Tile Stitching is currently under development.

REFERENCES

1. P. Sturm, S. Ramalingam, J.-P. Tardif, S. Gasparini and J. Barreto, 2010. Camera Models and Fundamental Concepts Used in Geometric Computer Vision. Foundations and Trends in Computer Graphics and Vision, vol. 6, no 1–2, pp. 1-183, 2010.
2. U. Iwerks, 1963. Panoramic Motion Picture Camera Arrangement. Canadian Patent Publication, no. CA 673633, 1963.
3. K. A. Tan, H. Hua and N. Ahuja, 2004. Multiview Panoramic Cameras Using Mirror Pyramids. Trans. on Pattern Analysis and Machine Intelligence, Vol. 26, no7, 2004.
4. C. Weissig, O. Schreer, P. Eisert and P. Kauff, 2021.The Ultimate Immersive Experience: Panoramic 3D Video Acquisition. Proc. MM 2012, Klagenfurt, Austria, January 2012.
5. Y. Sanchez, R. Skupin and T. Schierl, 2015. Compressed Domain Video Processing for Tile based Panoramic Streaming using HEVC. Proceedings of IEEE International Conference on Image Processing (ICIP), Quebec, Canada, September, 2015