

Creating and sharing immersive 360° visual experiences online

Masterthesis Master of Arts Zeitabhängige Medien
Sound – Vision – Games

Laura Fergusson de la Torre



Hochschule für Angewandte
Wissenschaften Hamburg
Hamburg University of Applied Sciences

Department Medientechnik

First evaluator: Prof. Dr. Robert Mores

Second evaluator: Prof. Wolfgang Willaschek

Hamburg, Thursday, 15 August 2019

1 CONTENTS

1	Contents	1
2	introduction.....	3
2.1	abstract	3
2.2	motivation and overall purpose.....	3
3	Main Research.....	4
3.1	Definition and differentiation between VR, AR and MR	4
3.2	human interaction – VR seen through a headset.....	6
3.2.1	user interaction with the virtual environment.....	6
3.3	framing content into immersion	12
3.4	Technological Status quo and Background	14
3.4.1	Software and technical parameters.....	14
3.4.2	distortion and line segment detection	19
3.5	Sharing and showing 360° online	24
3.5.1	sharing immersive content online independently	24
3.5.2	Hardware.....	28
4	Author's project: the 360 explant video for the DESY Hamburg	30
4.1	Project story.....	30
4.2	Production workflows.....	31
4.2.1	MODO 3D equirectangular Rendering	31
4.2.2	360° video rendering and upload to YouTube	33
4.2.3	static immersive visualizations	36
4.3	Reference use cases and scenarios for immersive media	37
4.3.1	journalistic VR.....	37
4.3.2	entertaining purposes	39
4.3.3	scientific purposes.....	41
5	Outlook: future (pervasive) uses of immersive media content	43
5.1	massive commercialization of hardware for immersive experiences.....	43

5.2	user generated VR content and online communities	44
5.2.1	Google street view	44
5.2.2	VR social media: a parallel reality?	46
6	Bibliography.....	48
6.1	image index.....	48
6.2	Literature and websites	49

2 INTRODUCTION

2.1 ABSTRACT

this paper is a practical introduction into immersive photography, 3d rendering and postproduction of images and videos suited for the web. The scope includes topics such as human interaction, technological status quo of visual VR, AR and mixed media use cases for educational, scientific and entertaining purposes. The main research is followed by practical examples based on the documentation of one large virtual reality project developed by the author over a period of six months. To conclude an outlook on the massification of VR in the web and VR social media is discussed along with information about the latest and upcoming VR experiences for the masses.

2.2 MOTIVATION AND OVERALL PURPOSE

With VR the creator can immerse viewers in imaginary or a mix of real worlds with imaginary elements, like the microscopic world or walking on the surface of other planets. VR is a medium that can take viewers to places, show ideas and dreams somebody else experienced or imagined. Creators can use VR to tell stories beyond words and images, by letting others experience, feel and navigate a place as if it were possible to open a window for others to peek into strangers' dreams. The science of dreams and dreaming itself has always amazed me. This is the main reason why I dedicated most of my time while studying to VR projects. This paper is a compilation of the practical experiences I have gathered along my studies in regard to the development of a good storyline along with the technical setting needed for immersive stories to come to life.

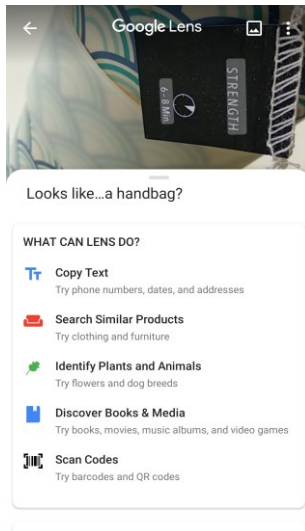
3 MAIN RESEARCH

3.1 DEFINITION AND DIFFERENTIATION BETWEEN VR, AR AND MR

Immersive media can be clustered into 3 main categories: virtual reality, mixed reality and augmented reality. For the latter two categories it is crucial to use for playback special devices for ensuring a fully immersive experience.

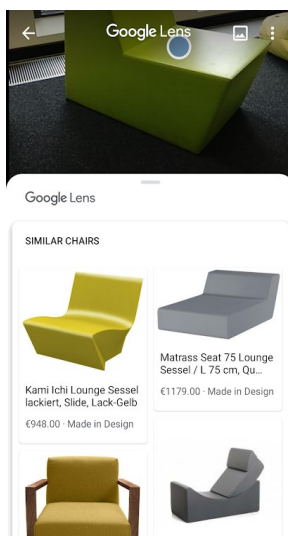
The most accessible and simple way to display Virtual Reality (VR) is by using WebVR, which consists on showing the virtual world on a flat screen either full-screen or in a browser frame. This method is compatible with smartphones, tablets and PCs which makes it easy to access by most end users, but the perceived quality of experience might vary strongly depending on which device the user uses to watch the experience. To ensure that the end user will most likely have a Fully-Immersive VR experience it is advisable to use methods like Multi-Sensory Cinema or CAVE.

Augmented Reality (AR) is an overlay of computer-generated content placed over the real-world objects where information elements such as menus or signs blend in real time over the environment. AR can have zero occlusion, low intrusiveness and high interaction. AR is the most user friendly way to create wireless experiences for end-users. An already patented AR solution is Google Translate “Word Lens” which uses the users smartphone camera to translate and identify signs, menus and any kind of text in real-time from one language to another.



*Image 1 - google lens mobile interface identifies text on an object
Source: self-made*

Mixed Reality (MR) is a step further than AR: here the overlay of synthetic content such as solid objects or people in a room blend in with objects in the real world in real time. Mixed Reality experiences have occlusion: the computer-generated objects can be visibly covered by objects in the physical environment. By using MR, for example an user could overlay a digital a cloth he/she is interested to buy over a table at home to generate a preview of how the cloth would look like in real life.



*Image 2 - object identification in google lens mobile app
Source: Self-made*

To conclude, Virtual reality does not mix or interact with the real environment and can be fully computer generated while Mixed Reality and Augmented Reality rely on the real environment. All MR is AR, but not all AR is MR. AR simply overlays information over the real world while MR interacts with it (Irvine et al., n.d.).

3.2 HUMAN INTERACTION – VR SEEN THROUGH A HEADSET

3.2.1 USER INTERACTION WITH THE VIRTUAL ENVIRONMENT

How valuable is it for the viewer to be able to interact with characters and objects inside a virtual world? How does interaction affect the structure and perceived quality of the story? How are virtual experiences better exhibited? In this chapter the content of the storyline in relation to the end-user's interaction with the digital environment will be discussed.

3.2.1.1 IMITATING REALITY: A VIRTUAL STORYTELLING EXPERIENCE IN A MUSEUM

The first example of an academic research project took place in the District Six Museum, in Cape Town where a group of researchers developed a storytelling experience for visitors to preserve personal experience narratives from two voluntary storytellers named Joe and Noor in an interactive virtual environment (VE).

The reason why the Museum decided to “preserve” stories from these two particular volunteers is because of the historical value they add to the community: this district was the first part of the city where freed slaves, merchants, artists and immigrants lived together peacefully, a representation of the opposite of the oppressive government. The storytellers that grew up in the 6th district in the 1960s are now aging and will most likely sooner or later not be able to come to the museum to tell their stories themselves. The museum curators wanted to preserve the stories of the district in an interactive format that imitated real life storytelling because tales of personal experience are a great format to teach visitors about historical events, particularly when told by those who experienced the events themselves.

The researchers aimed to imitate real life storytelling in the VE: the content of the most popular stories was preserved while still providing interaction possibilities to the listeners. Classical digital storytelling is usually used to presenting real-life stories that listeners are not able to influence or interrupt but only to comment, leading to story experiences with little or no interactivity. Researchers designed a VE that allowed highly interactive story experiences with little authorial control.

After listening to the real-life stories, researchers identified a schema that most interesting personal experience narratives followed:

1. abstract (story's beginning)
2. orientation (provide context)
3. complicating action (the main event – usually something unexpected)

4. resolution (the outcome)
5. evaluation (comment on why the story is interesting, this can happen near the end or throughout the story)
6. coda (signals the end)

good storytellers establish an emotional connection with audiences that is difficult to imitate digitally. The researchers' approach was to first observe real-life oral storytelling ethnographically to identify how experienced storytellers' bond with audiences. **Structuring the narrative before starting to design was crucial in this case because this experience aimed to preserve the stories by imitating real life narrative.**



Image 3 - Start of the Virtual experience where the user is part of a virtual audience
Source: (Ladeira et al., 2011)

The VE was displayed using PCS and consisted of an imitation of the real-life storytelling experience: a room full of people listening to real life stories of Joe and Noor that from time to time were present live to tell stories in the museum. The viewers had two options of interactivity: users could type questions to the storyteller and/or the storyteller will ask users a question they can answer by typing.

Researchers used open source software to create the VE and ensured that any affordable computer was able to playback the experience without failing. The software used to code the game was Microsoft's XNA Game Studio and the 3d models and room were created in Blender 3D. To animate the characters, researchers inspired themselves on Joe and Noor's gestures and movements during storytelling.

The storytelling interaction framework is simple and allows spontaneous interaction: the user is part of a virtual audience facing the two storytellers (Joe and Noor) and can move and look

around using standard keyboard and mouse controls. Interactions were possible at any point during a story by pressing the space bar to 'raise the hand' to ask a question; the message "You may press SPACE to put up your hand" was always displayed on the screen. For every interaction possibility keyword matching was used to recognize the user's input.

The storytellers were programmed with a reserve of possible responses that were fetched based on simple keyword matching of the user input to find an appropriate response. If no matching response was found, the storyteller responds by saying "I don't know". Occasionally the storyteller would encourage the user to raise the hand by saying "If you have any questions, raise your hand", and waiting for the user to press Space after.

The second possible interaction scenario is an exchange of information between the user and the storyteller: the storyteller asks a question to the user and a typing box is displayed for the user to type an answer. Each question has a 'final correct answer'; if the user enters this answer, the storyteller gives positive feedback and the interaction ends. If the user enters an unrecognized or incomplete answer, the storyteller will continue to encourage the user to give another answer by saying: "No, try again". To ensure that this interaction does not become an endless loop, the user has three tries to answer the question before a virtual audience member answer.

To test the effectiveness of the exchange structures, researcher conducted a user evaluation with 150 university students. Participants saw the storytelling VE while being observed by a researcher in a closed room making ad-hoc observations (thinking out loud). Most participants showed an initial positive response by laughing, speaking, exclaiming and nodding. Most participants reported frustration at the inability of the storytellers to answer their questions. The prototype was limited by the number of recordings the researchers received from Joe and Noor in real life.

User questions could also be made more efficient by using more sophisticated artificial intelligence techniques than simply matching keywords. The student trial was nevertheless necessary to dig deeper into linguistics and discourse analysis to identify a structure in the recorded stories to fit this into well-defined interaction patterns. (Ladeira et al., 2011)

3.2.1.2 SOCIAL IMMERSIVE MEDIA

A different approach called social immersive media (SIM) to create and exhibit virtual experiences was developed and documented over a 10-year period by Snibbe and Raffle. This method uses interactive camera/projector systems and has been implemented mainly in

museums for public science exhibitions. The storylines of the experiences rely on following 3 concepts: philosophical goals, design principles and interaction techniques that aim to generate emotional responses and social engagement through touch and movement. **What is different about SIM is that does not rely on graphic user interfaces (GUI) but instead on the “language of cinema” aiming to make users actors within the virtual experience.**

There is a growing interest for Interface designers to implement immersive strategies in collaboration with interactive graphics systems to engage users beyond the GUI world. Virtual reality systems aim to artificially generate all the sensory input channels for a single person by placing them in a completely synthetic world; this synthetic interactive world is what can be defined as social immersive media:

Instead of having a classical GUI interaction between the user and the system takes place in a shared social space using a person’s entire body as the “input device” without using any electronics or controllers attached to the body. SIM experiences engage people's whole bodies to “orchestrate” an exhibit aiming to trigger stronger emotional responses and user engagement. Users’ bodies interact with projected interactive graphics and video through a camera or sensor set-up. Even though Standard video is a known method for emotionally engaging visitors, museums increasingly include exhibits that physically engage visitors.

One well-known 5th century Confucius quote summarizes the benefit of interactive learning: *“I hear, and I forget. I see and I remember. I do and I understand.”*

To succeed at SIM Snibbe and Raffle identified following design principles for ensuring engaging immersive experiences:

1. *“Visceral. The media is experienced physically and emotionally, through whole-body interactions, before it is experienced symbolically or rationally.*
2. *Responsive. The media responds immediately, clearly and predictably to users’ actions.*
3. *Continuously variable. The media, like our natural environment, is continuously changing with infinite variability, usually through simulated response to the user’s image, silhouette, sound, location, gestures or other uniquely identifiable features.*
4. *Socially scalable. Interactions are designed to share with others. Furthermore, interaction, representation, and users’ engagement and satisfaction should become richer as more people interact.*

5. *Socially familiar. The media should augment and reinforce existing collocated social behaviours.*
6. *Socially balanced. Interaction equally emphasizes a user's awareness of herself, other users, and the media itself. (Snibbe and Raffle, 2009, p. 1449)*

The SIM approach is inspired by the book philosophy of phenomenology (Merleau-Ponty, 2013) basing its design principles on the idea that “reality” only exists as an interplay between consciousness, the perceived world, and our bodies. To put it simple: each human has a “gut” or primitive mode of perception that precedes the analytic, symbolic and linguistic.

For making media that engages with the visceral part of a person's mind (visceral design philosophy) it is necessary to abandon language and all meaning related to it and replace it with non-symbolic and enactive representations of knowledge. This philosophy has two important positive side-effects that come from focusing on visceral experience and avoiding language: users of all ages can engage and understanding across cultures without modification or translation is possible. Younger viewers (4-10 years old) tend to absorb educational and cultural concepts only as a secondary to their main experience of physical play.



Image 4 - Interactive Exhibitions in Museums: Three drops (left) and Boundary Functions (Right)
 Source: Snibbe and Raffle, 2009

Snibbe and Raffle focused were as well inspired by a well-known time-based medium for emotional engagement and storytelling: Cinema. Principles of cinema and animation proved to be useful in designing social immersive media experiences with one exception: the cut, one of the most powerful elements of cinema causes viewers to disengage from an immersive experience.

In the interactive scientific Exhibition “Three Drops” (Image 4) users take a digital shower under a projected showerhead. When viewers stand in front of the projection the water splashes, puddles, and flows around their shadow. After a while viewers dive into the

microscopic world to see and understand physical properties of water. When magnified, water behaves different: a drop of water becomes a giant ball that viewers can catch and throw, and the final stage visitors get to see how individual water molecules look like.

Visitors can engage with this exhibit using full body movements: showering, bouncing with simulated water droplets or by standing still to see how water molecules hover around their bodies. Unlike cinema, since users arrive continuously, the "Narrative" of each interaction must be designed to remain understandable independently of when an user decides to enter into the experience.

A convenient but misguided approach to social immersive media is to imitate GUI in social space, where bodies replace cursors and the space is a screen holds a series of buttons. The problem with this approach is that human social and physical interaction is naturally not digital, consisting of singular specific events, but rather continuous, comprising continuously varying movements and sounds.

In unmediated full- body interactive experiences users expect that objects respond continuously and directly to the changing full-body gestures, rather than using the body to act as a pointer that activates buttons and widgets. For example, replaying user movements creates a lively feedback between users' recording and observing which promotes attention.

Snibbes approach to design is balanced mixture of human body movement and layout of interface objects. More specifically researchers relied on Thomas and Johnson's seven principles of animation (Thomas and Johnston, n.d.) to evoke emotionally expressive animated responses:

1. *Easing in and easing out. Natural actions ease in slowly and come to a halt gradually. Ignoring this principle produces robotic movement and abrupt change.*
2. *Overlapping action. Natural movements overlap in time. For example, a person does not walk toward a door, halt, then reach for the doorknob. Rather, he simultaneously extends his arm while walking towards the door.*
3. *Follow through. Natural actions do not abruptly end, but rather follow through, in the same way a baseball swing does not abruptly end after hitting the ball.*
4. *Staging. Since users can only pay attention to one thing at once, the designer must direct users' attention to different portions of an experience at different times. This is accomplished by choosing when and where on-screen elements change. In general, there should be one dominant action upon which a user's attention is focused.*

5. *Squash and stretch. Real objects deform as they move and interact with the world, either by squashing and stretching as fleshy bodies do, or by motion blur, when captured by a camera.*
6. *Exaggeration. Since animated representations are usually simplified compared to the natural world, in order to get the same or greater emotional impact, the designer must exaggerate features, motions, and reactions.*
7. *Timing. Human beings are very sensitive to the timing of movements. The same animated poses can be interpreted very differently depending on the time taken to interpolate between them. For example, a head turning in 30ms could mean a character heard a gunshot, while a head turning over 500ms indicates he is feeling pensive (Snibbe and Raffle, 2009, p. 1493).*

Additional to the animation principles described above researchers identified important spatial and social guidelines that assure a successful exhibit: the exhibit should be placed where visitors will naturally walk into it, there should be enough space to ensure that interaction with the exhibit doesn't end up in an accident between visitors and all exhibits should be accessible to the disabled. For security and social reasons it is just as important for visitors to see each other as it is for them to see the exhibit encouraging social interaction and group learning (Snibbe and Raffle, 2009).

3.3 FRAMING CONTENT INTO IMMERSION

A broad spectrum of topics can be in theory fitted into an immersive experience, nevertheless there are situations and stories that are exhibited at its best in an immersive way. A good example of getting the best out of immersive images is EasyTour, a digital heritage image archive reframing concept: Mazzoleni et al. had the idea of turning an archive of historical images into an augmented intelligent panoramic tour to maximize interest into this topic.

The spread of a high bandwidth network, the ability to purchase high - quality digital cameras at a convenient price and the availability of open - source programs for manipulating and composing digital images, make 360 panoramic imagery a technology used in a growing number of websites. Panoramic images are wide screen images particularly suited for landscapes, where a lot of sceneries can be put together.

Even though this technology is highly appealing, panoramic imaging is still not widely adopted as a massive technology for sharing knowledge online. Two main aspects can be attributed to such limitations: the first aspect is that most panoramas available online are not

connected with the rest of information of the website and the second is limited user-environment interaction.

Mazzoleni et al. believe that panoramic imaging technology should be used to create environments that invite users to interact with all the heritage information represented in the panoramas by hiding the complexity of the underlying DBs and archives.

With help of EasyTour experience designers will be able to enhance the information presented in the panorama by directing end-user's attention to elements that are most relevant. What happens during a real visit is recreated digitally: the expert guide does not wait until the visitor asks for additional information about an object but points out elements that are most likely relevant to a specific target of visitors.

The user can create panoramas and edit hot spaces with EasyTour. The scene map's hot area corresponds to a panoramic view and the hot areas in each panorama correspond to "tour into the picture" sub-sights. The user can virtually tour the scenes by modelling the sight navigation into the global model and local model.

Panoramas could become a new user-generated content-based interaction medium. The idea is to follow Flickr and Bubbleshare's approach: those websites that offer the user the ability to store and share their own photos, ultimately enriched with other multimedia content. Bubbleshare, for example, enables its users to publish photo albums with text, audio comments, and notes of other visitors.

To ensure a more human immersive experience, familiar design concepts such as hotspots and overlays of information as text must be adapted to the end user's language. Semantic Hotspots are an extension of traditional panoramic image hotspots. They can be connected to one or more regions of a panorama like traditional hotspots. While traditional hotspots are associated with actions (to be executed when the user clicks on the region), semantic hotspots are associated with concepts within the panorama based on the cultural setting that describes the information. **Instead of a pointer hovering over a region, a conversation or words in a conversation are the new triggers for (inter)actions.**

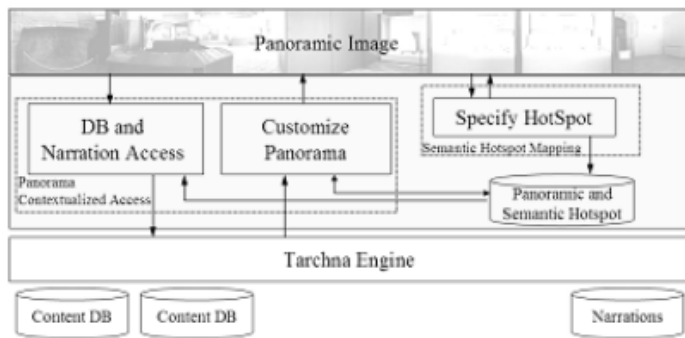


Image 5 - System Architecture of EasyTour
 Source: (Mazzoleni et al., 2006)

To draw an example: the experience designer aims to specify the panorama of an ancient tomb. The “Tarchna” Engine automatically presents a preview of the additional information of all the tombs available in the data sources by fetching similar concepts on the semantic model. In this way an insulated story can be interconnected with similar stories enabling a whole new architecture of information and path exploration possibilities for end-users.

When the user selects a narration created by an experience designer, the system will extract and highlight the selected theme-related semantic hotspots in the panorama. If a narration is associated with hotspots belonging to multiple panoramas, the user can preview them and find out which objects are associated with the narration she has chosen. EasyTour encourages viewers to actively participate in rewriting history while improving the showcasing of heritage related content (Mazzoleni et al., 2006).

3.4 TECHNOLOGICAL STATUS QUO AND BACKGROUND

3.4.1 SOFTWARE AND TECHNICAL PARAMETERS

3.4.1.1 EQUIRECTANGULAR PROJECTIONS AND DRAWINGS

Virtual reality panoramas in social media are becoming increasingly popular. Facebook, Google, and Flickr provide users with relatively easy ways to upload and share their own 360-degree photographic panoramas as VR experiences. Any image in equirectangular perspective with correct metadata injected can be uploaded and the rendering software on each platform will project it automatically, by giving any creator the possibility to share without coding abilities a simple VR static scene.

Anybody with enough money can buy a 360-degree compatible device or camera to take and publish their own equirectangular images. Other more time and talent demanding possibilities are creating and rendering a drawing with 3d modelling software or to draw by

hand. Several artists have used and shared equirectangular drawings: an interesting collection of these drawings on Flickr’s artistic panorama group can be seen on albums by David Anderson (Anderson, 2011) and on-location drawings by Gerard Michél (Michel, 2008). There is a growing interest among illustrators and urban sketchers to stand out by creating and sharing both curvilinear and anamorphic perspectives on their creations online.

VR panoramas are not simple to create merely by hand, the artist must have knowledge in descriptive geometry. A central perspective can be defined as a conical (anamorphosis) projection onto a surface, depicted after in a flattening map (the surface) to a plane. The anamorphic method includes cylindrical perspective, (hemi-)sphere perspective, and any overall spherical perspective possibilities.

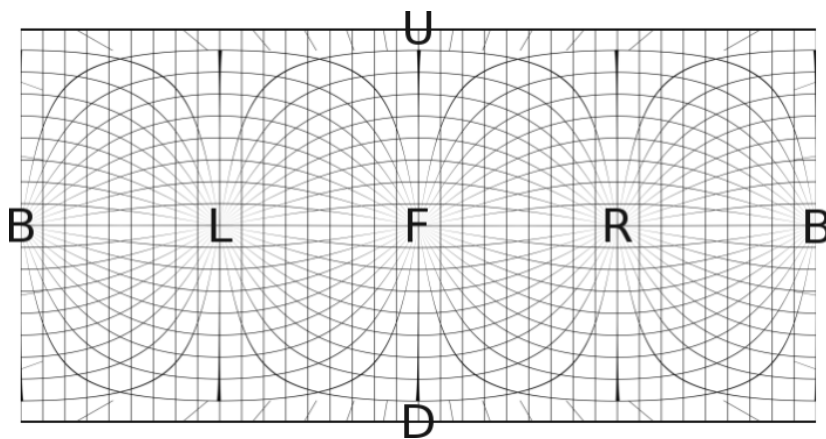


Image 6 - Equirectangular grid with Up, Back, Left, Forwards, Right, and Down
Source (“Drawing a Spherical Panorama,” 2016)

To draw an equirectangular image correctly it is necessary to “solve” and understand the perspective: a delimitation of all lines and of their vanishing points as shown in image 6 is to be plotted before starting to draw by hand. Equirectangular perspective can be simply plotted point-by- point by a computer, but a way to make it solvable with simple tools requires talent (Araujo, 2017).

An alternative approach to process equirectangular content effectively is described in the paper “Translating the Viewing Position in Single Equirectangular Panoramic Images” (Sandnes and Huang, 2016). This investigation proposes a post-production method for flat panoramic images resulting in an illusion for the end user of moving freely in three dimensions. This method works for panoramas in closed space with flat ground and flat vertical walls.

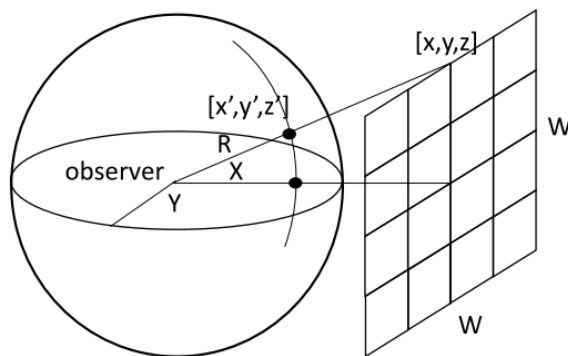
Full simple equirectangular panoramic images are static and have double width relative to height - representing 360 degrees horizontally and 180 degrees vertically. The end user can control the viewing direction and field of view by turning the body and the head without

walking. Only when the user is not walking, a convincing impression of being present in the actual location is experienced.

Researchers were therefore looking a method that allows a viewer to move around static panoramas. This method works best with panoramas containing squared and straight objects, which often contain flat faces. Example scenarios include moving around inside a room defined by one panoramic image or moving from one point of view (image a) to another (image b) with a smooth transition between the two “close” panoramic images.

The method uses a so-called Hough transform to attempt identifying flat vertical faces, or walls, in the panorama. Using the face information, the panoramic image is decomposed into the respective faces so that when the viewing position is moved, a new panoramic image will be reconstructed based on the predefined polygons (flat faces).

The process consists of computing the optical flow field between adjacent cubic panoramas, and then a morphing algorithm is used to achieve the motion between the two panoramas. In this way achieving pleasant visual transitions over realistic representations is possible.



*Image 7 viewer image projection onto a perpendicular plane
Source: Sandnes and Huang 2016*

Sandnes and Huang's strategy is based on interpreting an original point of view and translating that equirectangular panorama as seen from a changed position. The conversion happens in three steps: firstly, the walls (planes) are recognized; secondly these planes are extracted from the panorama and thirdly the point of view is interpreted so that a new panorama is created based on the new point of view.

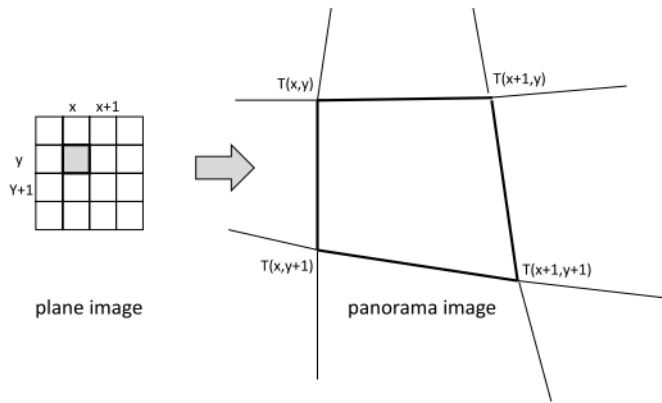


Image 8 plane to panorama mapping
 Source: Sandnes and Huang 2016

Once the plane boundaries are identified, the square image is projected back onto the respective plane within the defined boundaries to build the new face. The projection is achieved by traversing pixel points on the plane and filling these with the corresponding value of the original point projected onto the equirectangular space.

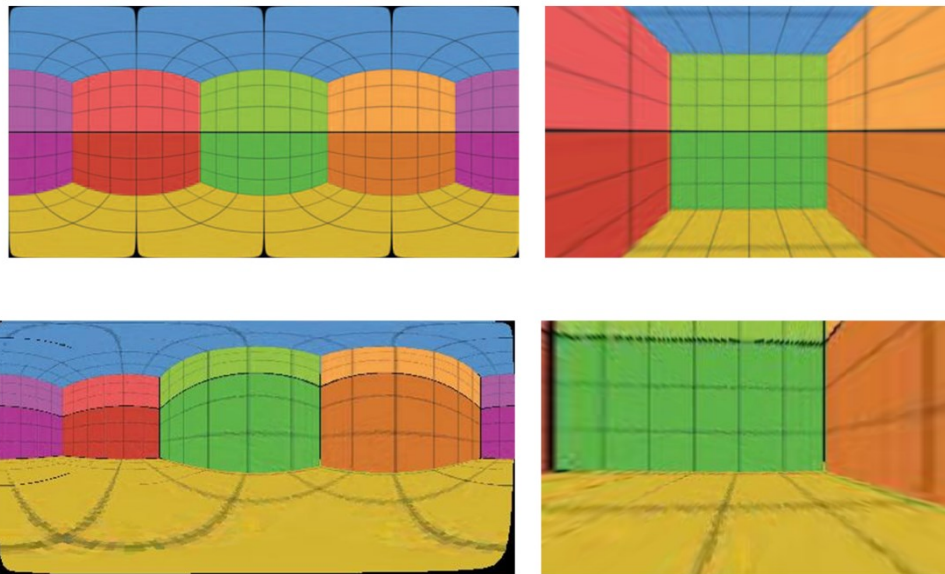


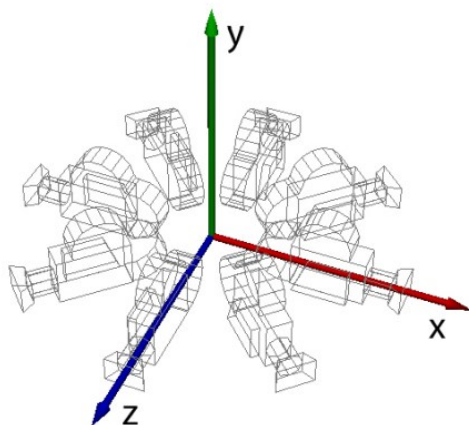
Image 9 square room example illustration before and after translation
 Source: Sandnes and Huang 2016

A square room illustration was used as an example to show that the projections after the translation transformation are nearly visually perfect. Nevertheless, this illustration has low quality and the aliasing (fogginess) effects are quite noticeable. Researchers have no solutions for anti-aliasing apart from filling in uncovered pixels using polygons. This simple illustration is a very easy case as it has regular faces that are easy to identify (Sandnes and Huang, 2016).

Going further into interactive immersive content it is important to discuss and explore possibilities with a more complex output option: panoramic immersive video. The availability for commercial devices to record panoramic videos is limited and most available options are quite expensive. Additional to this, immersive video post-production remains complex because there is no ready-made software to work with equirectangular video formats in 3d and the rendering requires a higher performance than usual HD videos. Sarmiento and Quintero developed therefore a framework for an affordable and simplified production of computer generated virtual panoramic immersive videos using open source 3D production software.

The production framework has two stages: panoramic video production and immersive visualization. In the first stage any computer-generated 3D scene can be taken as input to generate two outputs: the panoramic video and path sounds to immersive audio reproduction. In the latter, a desktop CAVE assembly is suggested to provide an immersive display available to more than one user at the time.

The panoramic view production process starts with preparing the camera settings and positions, followed by a rendering process carried out individually for each camera and a final video registration process that stitches all rendered images together. To obtain an adequate panoramic view, this framework requires covering at least 360 degrees in the horizontal line of vision. Therefore, N cameras are created with a different rotation value over the y axis from the main camera in the 3D scene.



*Image 10 eight camera rigs for 35mm lens cameras
Source: Sarmiento and Quintero 2009*

The number (N) of cameras depends on the horizontal aperture angle of the main camera. For example, a 35mm format camera with an horizontal angle of 49.3° which means that 8

virtual cameras with rotations of 45° from the main camera are needed for covering a 360° field of view.

After rendering it is necessary to stitch the views together to create a single panoramic scene in the end. The approach used in this step depends on the camera configuration used to generate the scenes. If the rendering is done based on real camera features, the output scenes will have an overlap in the y axis. The overlapping can be later adjusted manually across images in a post-production software ideally batch propagated to all cameras.

It is recommended to use a large display device is required that allows the user to navigate on a full vision field, to provide a feeling of immersion in the virtual environment. In cases where finding a big display is not affordable a desktop cave can be set up using correctly placed flat panels. For the cave implementation it is necessary to have a PC with independent video outputs (at least two).

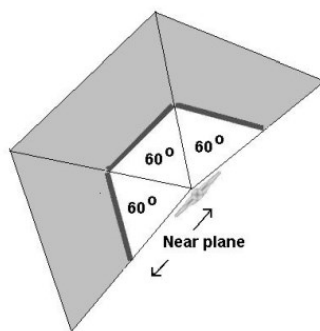


Image 11 simplified cave monitor installation
Source: Sarmiento and Quintero 2009

The discussed method represents a simple framework for producing and displaying virtual immersive videos. This technique can be useful for people with no or low budgets that are not able to afford an HMD or other expensive hardware or software (Sarmiento and Quintero, 2009).

3.4.2 DISTORTION AND LINE SEGMENT DETECTION

If a low budget and a simplified post-production is not in focus but delivering extremely realistic equirectangular images from panorama pictures a post-production algorithm based on line segment detection can be implemented. A recent method to improve the quality of the images when being projected is presented by On and Jung in the 2012 article “great circle arc detector in equirectangular images”.

Equirectangular images are starting to become ubiquitous; several map services such as Google Street View have user generated equirectangular images in their street view enabling anybody with an internet connection to “explore” places somebody else has photographed before around the world. The increasing amount of immersive image content has caused that conventional image processing techniques adapt into the panoramic format.

Nevertheless, researchers noticed that in most user generated panoramic pictures straight lines appear curved, in other words the projection in most images is overstretched in the upper and lower part of the images. To solve this problem, line segment detections for images presenting an undesired warping are proposed by researchers.

The main idea for the circle arc method is to achieve a perspective projection that preserves the line features. In order to do so it is necessary to deconstruct and project the viewing sphere onto an unit cube to apply the so-called the bilateral filter.

Line segment detection is a classical technique used in computer vision. This method first applies Canny’s edge detector (Canny, 1987) followed by the Hough transform (Ballard, 1981) these algorithms allow to detect and extract all lines that contain a number of edge points exceeding a threshold (that are curved). The curves are after segmented into small lines based on gap and length thresholds.

A line or straight region where most of its points share roughly the same image gradient angle can be defined as a line segment. Anyhow the points on a great circle arc (GCA) of the equirectangular image do not have the same the image gradient. This means that the lines extracted from the equirectangular “flat” image cannot be used as a reference gradient angle directly.

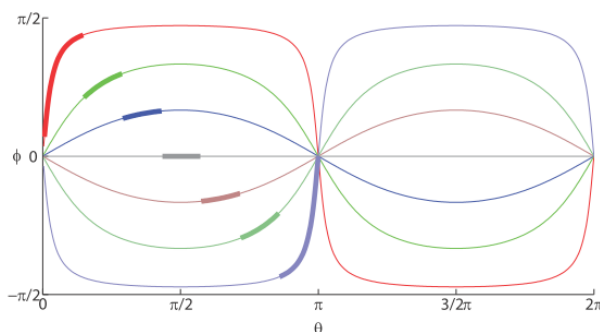


Image 12: curve segments in an equirectangular image
Source: (Ho Oh et al. 2012)

In order to be able to extract the segments correctly the researchers developed an algorithm that represented the line segments in the equirectangular image as GCAs. With this representation, the GCA-support region can be defined accordingly to the equirectangular

warping to redraw curved straight lines and achieve a more realistic image quality (Oh and Jung, 2012). The algorithm is available on the paper and can be implemented in the command line of image processing software such as photoshop.

Distortion is generally a big issue when it comes to equirectangular projections; almost all projections of a 3D scene into a wide-angle image will unavoidably result in distortion. Most projection methods either bend straight lines in the scene or have local distortion in shapes of scene objects. Carroll and Agrawala present a method that minimizes this distortion by adapting the projection to the content in the scene focusing on shape preservation.

In equirectangular images, a scene is seen from a single viewpoint located in the center of a viewing sphere. To flatten the viewing sphere, the inner surface of the sphere is mapped onto an image plane. Even so a sphere is not a flat surface and therefore a flat mapping will cause perceptual distortion.

Carroll's and Agrawala's solution attempts to minimize the most noticeable distortions in wide angle images: curving of straight lines and exaggerated deformations of objects. Their solution provides end-user with a graphical interface allowing them to mark the perceptually important lines in the image to use a reference to calculate a low-level salience measure and a face detector.

No sphere to plane projection can keep straight lines straight and preserve the shapes of objects. Therefore, Carroll and Agrawala propose a trade off between straight lines and shape preservation creating a in image between perspective and stereographic projection. The main goal isn't to keep all possible lines straight but only salient lines that appear in the image.

Commercial software alternatives for improving wide-angle images such as DxO Optics Pro (dxo.com) and Fisheye Hemi (imagnetrendsinc.com) allow the user to choose between a cylindrical or stereographic projection. The researcher's method is a combination of both producing a spatially-varying mapping.

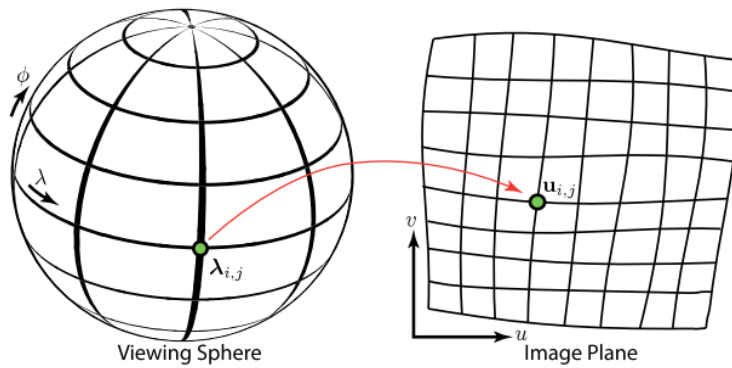
Researchers were inspired by the concept of perceptual constancy (Palmer, 1999). Constancy can be interpreted as the idea that human visual experience correlates mostly with the intrinsic properties of an object rather than with the extrinsic properties of an object's appearance in a designated perspective. For example, size constancy is the ability to know the true size of an object independent from its projected size on the distance. Based upon the perceptual constancy concept, researchers pursued following principles for wide-angle projections:

- Shape constancy (local): the shapes of objects should be maintained in the projection as they appear on the viewing sphere, for example spheres should remain circular.
- Shape constancy (global): bending global shapes should be avoided. To prevent this, linear structures in the scene should remain straight lines in the projected image.
- Orientation constancy: in case cameras are parallel oriented to vertical linear structures; vertical straight lines should remain so in the projected image.
- Perspective: the viewing sphere is projected onto a tangent plane through lines emanating from the center of the sphere.

Most commonly, independent of which type of projection is used the problem zones are usually located in the upper and lower corners of the image because when the field of view approaches 180° the stretching reaches its maximum. Researchers identified three types of projections and used them as a reference for their “mixed perspective” approach: Perspective, Mercator and Stereographic.

Perspective projection is considered apt for fields of view less than about 40° because distortion is low when the field of view is limited to the upper and lower areas. On the other hand, the Mercator projection is a cylindrical projection enabling to handle a complete 360° horizontal field of view. This type of projection stretches to infinity as the vertical field of view approaches 180°. The Mercator projection apt for panoramic images with big horizontal fields of view, such as images containing teams or large groups.

The last and most complex projection of all is the stereographic where the viewing sphere is projected onto a tangent plane through lines emanating from the pole opposite the point of tangency. Stereographic projections are shape oriented and tend to preserve the shapes of objects locally. Similar to perspective projection, stereographic projection presents object stretching toward the borders with a slightly better scaling than the other perspective types.



*Image 13 optimized mapping from the viewing sphere with a mixed perspective approach
Source: Carol and Agrawala 2012*

There is no “one fits all” projection across scenes. Researchers discovered that mixed projections within the same scene can help to avoid perceived distortion by end-user.

To find the right projection combination, researchers followed a content-preserving approach for projecting wide-angle photographs from the viewing sphere onto a flat image. The solution has two main components: a simple user interface that enables users to choose the most important lines that should remain straight and a weighted shape optimization procedure for generating a smooth mapping from the viewing sphere (stretched projection) to the image plane.



*Image 14 image processing example and perspective comparison
Source: Carol and Agrawala 2012*

If a scene has several straight lines in different directions, it is hard to straighten them all. perspective projection is the only mapping that can preserve all straight lines on a limited view angle (top and bottom are problematic). textures with many lines, such as a hardwood or tiled floor are harder to render without distortion.

Carroll’s and Agrawala’s method can improve noticeably the overall distortion but the north and south poles remain challenging areas due to singularity at the poles in spherical coordinates. In most scenes this was not a problem; however, scenes with textures or important objects near the poles could seem distorted (Carroll et al., 2009).

3.5 SHARING AND SHOWING 360° ONLINE

3.5.1 SHARING IMMERSIVE CONTENT ONLINE INDEPENDENTLY

3.5.1.1 WEBVR

Web VR experiments is a creative project from Google consisting of an experimental VR browser integration compatible with Firefox and chrome. This browsing experience is in an experimental and surging phase as most VR browsers. Web VR showcases a collection of creative projects curated and sponsored by Google such as animations and short immersive experiences of developers and VR artists. Web VR claims that their final goal is to make it easier for anyone to create, view, and share VR experiences.

The variety of showcased interactive VR experiences in the homepage are entertainment focused, mostly music videos and games, some of the titles of the experiences include:

string fireworks together to make them explode and collect points (game)

Join users from around the world in a musical forest (music)

Norman is an open-source Web VR tool built by an artist to create frame-by-frame animations in 3D.

The real surface of Mars. Photographed by the Curiosity Rover

In Web VR simulate how any origami crease pattern will fold in 3D

Experience a kaleidoscope of iridescent prisms and meditative soundscapes

Connect with someone by joining them in an emotive, interactive journey

Hear what it sounds like when the sky rains dog toys, ping pong balls, and more.

The Web VR API is openly available for a wide variety of VR desktop and mobile browsers including Firefox and chrome for macOS and windows.

3.5.1.2 AFRAME AND GLITCH

A-Frame is a web framework for coding and hosting VR content in the web. This framework is HTML based which means that coding is possible in standard text editors without having to install any additional software. Glitch is an online code editor where developers can code, test, deploy and host VR content for free. The A-Frame team encourages newcomers to use as

a template projects of other developers on Glitch. It is important to bear in mind that any project made on glitch could be potentially copied and used by other people on the web, as stated on glitch's about page ("Glitch - Fog Creek is now Glitch!," n.d.):

"Every app on Glitch can be remixed. With just a click, you get your own copy of the app that you can customize, personalize, extend and build on to your heart's content"

A-Frame was developed originally by the Mozilla team aiming to be an easy but powerful way to develop VR content in the web. Likewise Mozilla, A-Frame is an independent open source project which has grown to be one of the biggest VR communities over the years. The biggest advantage that A-Frame has is that it is HTML based which makes it easier for most amateurs to get started. HTML is an understandable mark-up language with an extensive documentation available online.

However, A-Frame seems only easy on the surface: A-Frame is at its core an entity-component framework that provides a declarative, extensible, and structure similar to three.js. The compatibility of A-Frame is broad, it works on standard desktop and smartphones because it is performance oriented for VR on the web and HTML based. A-Frame enables to load graphic content faster uses the browsers DOM but its elements don't use resources of the browser layout engine ("A-Frame Introduction," n.d.).

3.5.1.3 SUPERMEDIUM: A FULLY VR BROWSER

Supermedium is a browser specifically built for the VR Internet. With Supermedium users are not able to browse 2D sites like Facebook or Twitter but browse fully 3D, interactive, VR applications, instead. VR sites are built similarly to native VR applications and can display any kind of content. Most common content types vary across mini-games, musical trips, creation tools, art installations, journalistic pieces, modelled environments and even dance parties. Supermedium focuses on being available for mainstream headsets that have positional tracking and controllers such as the HTC Vive and Oculus Rift.

The content that can be browsed in Supermedium is published using standard Web technologies. As a developer it is not necessary to use or refer to Supermedium when building a VR site. VR sites built on the Web that could be opened in other browsers that support VR can be opened using Supermedium providing end users with mainstream headsets a better VR browsing experience ("SUPERMEDIUM - virtual reality browser," n.d.).

3.5.1.4 BANDWIDTH AND DEVICE LIMITATIONS

Streaming or watching VR experiences uses up a lot of resources and bandwidth, this can represent a problem when aiming to reach great audiences and/or mobile users. A considerable amount of papers has been published around this issue. One of the most solution-oriented methods is presented in the paper “Gaze-aware streaming solutions for the next generation of mobile VR experiences” by Lungaro et al. in 2017. The proposed method aims to deliver high visual quality in real time, by rendering only the area where the user is looking in HD while lowering the quality in the remaining scene. This reduces the overall bandwidth requirements for playback of VR video experiences while maintaining high levels of user perceived quality.

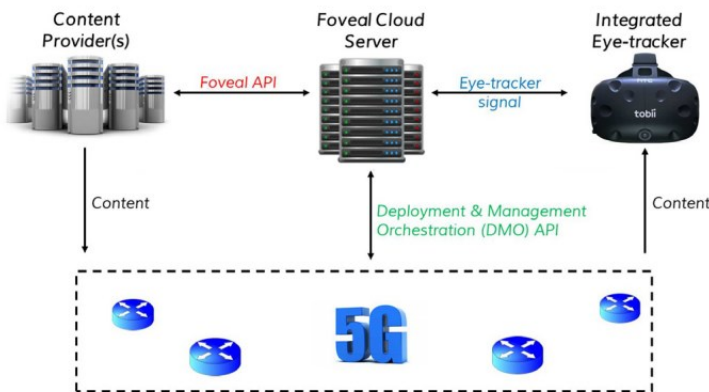


Image 15 end-to-end architecture for selective VR rendering
Source:(Lungaro et al., 2018)

A common issue of video streaming is the vast amount of bandwidth required to support high quality videos. For example, YouTube guidelines recommend bitrates in the range of 35-85 Mbps but, the average internet speed the top ten countries in the world can stream videos with a bitrate of 18.7-28.6 Mbps meaning that an acceptable user experience is highly unlikely for most viewers. Instead of “throwing more bandwidth at the problem” the approach of Lungaro et al. proposes to focus on following parameters to improve render efficiency:

“Eye-tracker latency (EL) time between the capture of an eye image and the availability of its gaze position in the client device.

Network latency (RTT) - this is the network Round-Trip Time (RTT), referring to the time between when a new eye-gaze sample is sent to the

foveal server and when the corresponding updated content is received at the user side.

Processing latency (PL) - time spent to decode and process the received content and to submit it for rendering in HMD.

Display latency (DL) - time required to render and output to the user HMD a content frame.(Lungaro et al., 2018, p. 1536)”

Videos are split into tiles to be dynamically assigned to one of two regions: the foreground region in high quality and the background region in lower quality. Resolution declines by about 50% every 2.5° from the centre up to 30° and up from this point quality declines increasingly. The foreground region is located at the centre of the gaze point (looking direction) detected by the eye-tracker (Lungaro et al., 2018).

A similar research paper suggests splitting render resources between a server and a device for mobile users. The paper “*Furion: Engineering High-Quality Immersive Virtual Reality on Today’s Mobile Devices*” analyses the feasibility of the usage of high-quality VR apps on mobile devices such as smartphones. Mainstream HMS such as Oculus Rift and HTC Vive offer high quality VR streaming via a bulky HDMI cable. This cable limits users’ mobility and represents therefore a potential safety hazard to the user that could trip by stepping on the cable. Wireless VR is therefore the future.

Researchers determined that the potential quality of experience (QoE) for high- quality VR applications on mobile hardware and wireless networks via local rendering or offloading is 10 times lower than an acceptable QoE on desktop pcs. Next-generation wireless networks (e.g., 5G) will not bring an immediate solution because factors such as power limitation and the high processing resources needed will remain.

Researchers created as a solution a VR framework called Furion for optimizing VR for mobile devices in wireless networks. Furion basically splits rendering workload between a wireless device and an external server to save up CPU resources on the end-device side. the framework offers video compression along with use of panoramic frames and parallel decoding on multiple cores on the phone to ensure a seamless VR experience in a smartphone connected to WIFI. Furion is available to any experienced developer using Unity and can be implemented to improve any VR experience for mobile users. To enable Furion the developer must add the Furion prefabs in the scene and configure app-specific parameters (such as scene area limits) as shown below (Lai et al., 2017):

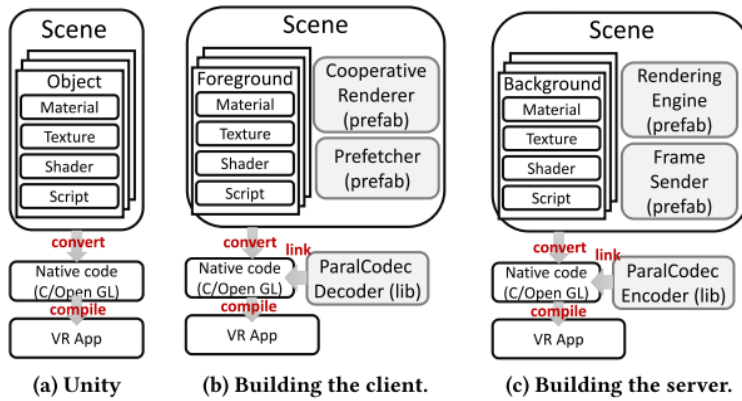


Image 16 system architecture in Unity - how to optimize VR apps in Unity with Furion
Source: (Lai et al., 2017)

3.5.2 HARDWARE

3.5.2.1 AFFORDABILITY AND AVAILABILITY



Chart 1 Global virtual reality video gaming revenue 2016, by region
Source: SuperData Research; IGN

VR hardware such as HMDs and controllers and VR ready devices are rather expensive devices which represents an economic limitation for potential VR audience members in developing countries. As shown on the chart above the amount of potential audience members varies across rich and poorer countries. The VR market is currently bringing the most revenue from first world countries, particularly from north America with the highest VR game revenue per capita. As a content creator this knowledge might be useful in order to determine if creating VR content designed exclusively for certain VR devices or only available in VR browsers will be accesible by the target audience. VR is still as of today a niche market in an experimental phase. Nevertheless, if the content creator knows how to make VR experiences fit into the mainstream devices of today the limitations discussed above will not

matter anymore. The big price tag might scare an unknown number of potential viewers off, but the interest in experiencing VR with affordable devices prevails amongst potential users.

A website article titled “*Everything You Need to Enjoy VR On Mobile in 2018*” states that anybody can start enjoying VR experiences for as little as \$10 by converting a smartphone to a VR device. A simple paper or plastic headset adaptable to any smartphone can be purchased for a relatively affordable price between 10–45\$.

to ensure high picture quality it is recommended using a smartphone with a resolution not lower than 1920 x 1080 pixels. Another important asset that the device should have is the G-sensor or gyroscope that allows the device to map its orientation in three-dimensional spaces. The gyroscope enables the user to look around in the 3d world; most VR games and applications can run on smartphones without a gyroscope, but the end-user will not be able to look around in the scene (Kuryliak, 2018).

3.5.2.2 HMD HEAD MOUNTED DISPLAYS

The table below is a price and feature comparison across the mainstream HMDs available in the market. The price range is wide and only the 3 cheapest options are wireless (a smartphone is required to use them). Non-wireless HMDs have on the other hands advantages when it comes to performance and quality because they don’t rely on device battery life and use CPU resources from a PC or console allowing them offer faster and better experiences.

Table 1 HMD price and feature comparison
Source: (“VR Brillen Vergleich,” n.d.)

Name	The Rift	HTC Vive	Play- station VR	Gear VR	Daydream View	Cardboard
Manufacturer	Oculus VR	HTC	Sony	Samsung	Google	Google
Display	2 x 90mm OLED	2 x OLED	5,7-inch RGB OLED			
Resolution	2160 x 1200	2160 x 1200	1920 x 1080	up to 2960 x 1440		
Pixel depth	456 ppi		386 ppi			
Framerate	90 Hz	90 Hz	120 Hz	60 Hz		
field of view (diagonal)	approx. 100°	approx. 110°	approx. 100°	approx. 101°	approx. 90°	approx. 100°
weight	470 g	555 g	610 g	approx. 280 g	approx. 220 g	< 90 g
Release date	01.03.2016	01.04.2016	01.10.2016	01.09.2016	01.11.2016	01.06.2014
special features	fabric cover	Room Size mapping	Only for PlayStation	includes controller	includes controller	cheapest alternative
Price in EUR	699	699	399	130	65	10

4 AUTHOR'S PROJECT: THE 360 EXPLANT VIDEO FOR THE DESY HAMBURG

4.1 PROJECT STORY

The Author was involved in a project where a 360° explanatory video for scientists working in the DESY Hamburg and the Helmholtz Center was created by using 3D animation software. The video was created to be shown in a conference of the organisation to draw guests attention to an investigative project. In the video actual 3d data of scans of a 2mm metal implant after being 3 months inside of a lab rat could be seen and compared across time. These 3d scans are used by scientists to figure out how to optimize the magnesium alloy properties in terms of degradation speed and stability.

The Helmholtz Center Geesthacht has several institutes working on materials sciences and oceanic science. For this particular project researchers from the Institute of Biomaterials and Institute of Materials Physics division worked together with the University of Lund, Sweden. The overall research aim of the project is to develop alternative biodegradable implant material for medical applications.

The implants are introduced into rats to test states of degradation after 1, 2 and 3 months. These metallic implants are very small screws (2mm length) of magnesium alloys. After implantation the metal decays within the body over a certain period of time, ideally until the broken bone is fully healed, resulting in a full body absorption and an intact bone. Researchers are finding out which magnesium alloy has optimal absorption time. This is beneficial since no second surgery would be necessary to remove or change an implant.

Variables such as the loss of original volume, the amount of newly formed bone and the area in contact between bone and screw are measured afterwards. In order to perform exact measurements in small implants the collection of 3D image data is performed on a special X-ray scan in Hamburg at the DESY facilities. After interpreting the 3d data researchers are able to optimize the recipe and production process of implant materials.

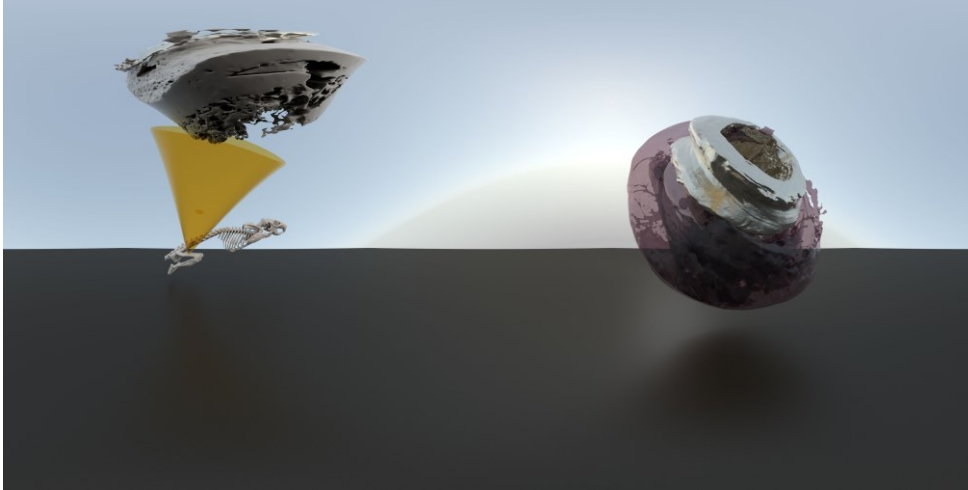


Image 17 equirectangular render from MODO
 Source: Self-Made

In the following chapters the production process of the explanatory video will be described.

4.2 PRODUCTION WORKFLOWS

4.2.1 MODO 3D EQUIRECTANGULAR RENDERING

The author was provided with 3d raw data from the x-ray scans to create the explanatory video with the goal to compare and show the differences of implant decay/absorption across time. The 3D animation and modelling software MODO was used to design a sleek scene with appropriate textures and materials for the screw, bone and the parts of the screw that were assimilated by the body across time.

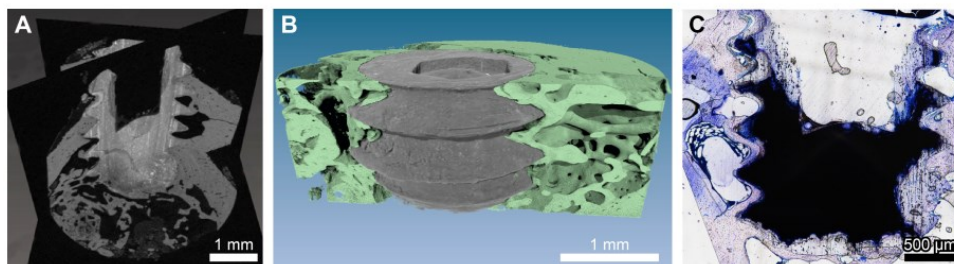


Image 18 Tomography (raw 3d data) of the screw inserted in a rat's femur
 Source: DESY Hamburg

The tomography data had to be first simplified (amount of vertices and polygons was reduced) to use less resources while animating and rendering the scene. The reduction was performed in MODO using the auto retopology function.

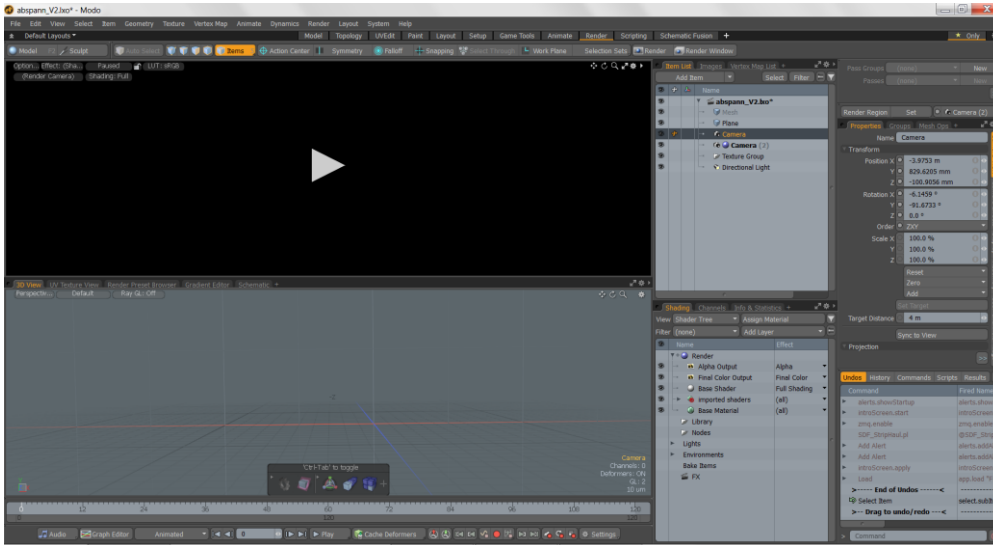


Image 19 Modo UI
Source: self-made

The camera settings were modified to render an equirectangular projection of the scenes in a HQ video frame size and rate (25 fps) as showed in the screenshot below.

- projection type: Spherical VR
- Focal length: 8mm
- Final render output needs to be 2160p
- Top-Bottom stereoscopic with total frame size 3840 x 2160

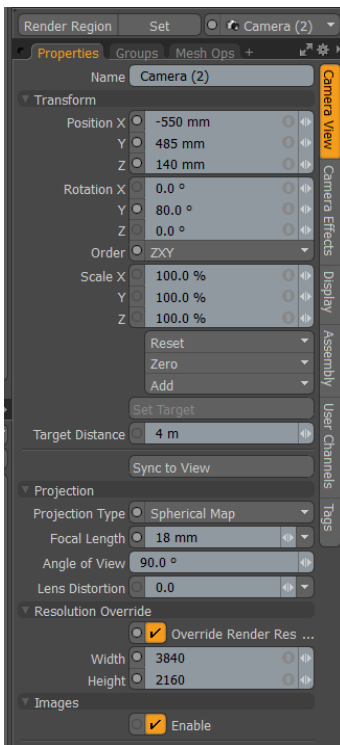
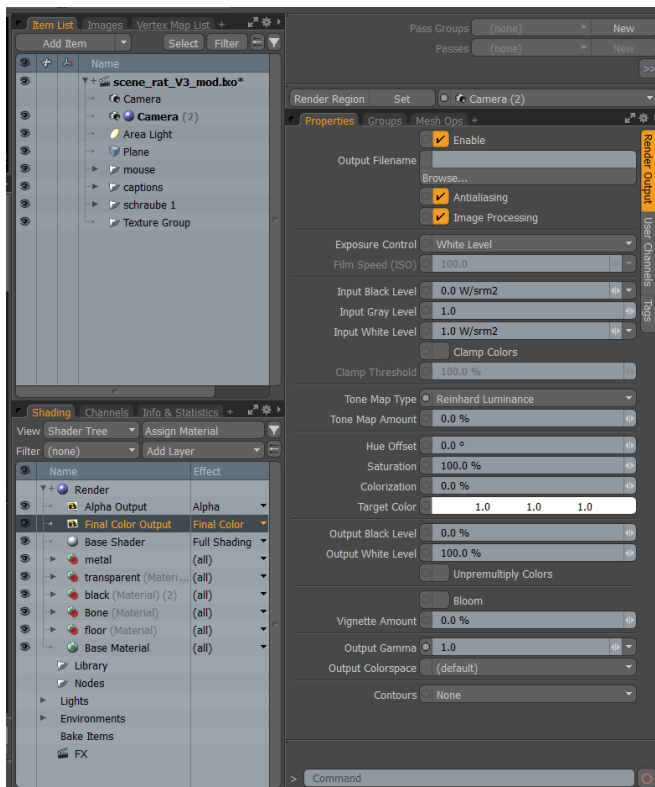


Image 20 Camera Settings in MODO
Source: self-made



*Image 21 Environment and Material settings in MODO
Source: self-made*

After setting up the perspective (for viewing) and 360-degree camera the environment, lighting and textures were configured as shown above. One area light was added extra to sunlight to improve visibility of the explant in all possible angles.

4.2.2 360° VIDEO RENDERING AND UPLOAD TO YOUTUBE

Once the animation and rendering are complete and all the frames are rendered accordingly, the video production can start. Static elements such as the author credits and a banner with logos were rendered as a single scene to save rendering resources. Only animated objects had to be rendered fully which resulted in 600 HQ frames which were rendered over a period of 3 days using the resources of 5 pcs sponsored and administrated by the DESY team. Once the Rendering was complete, one video the single frames were imported as frames, edited, compressed and rendered as a video using adobe premiere pro.

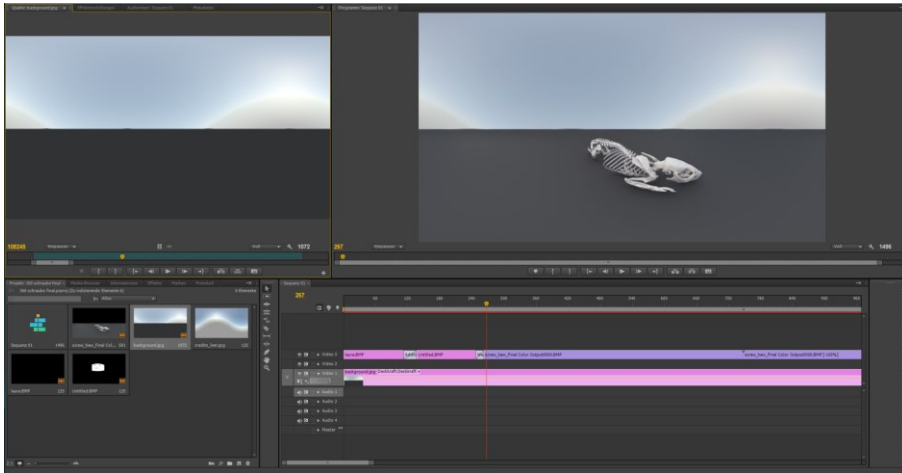


Image 22 Adobe Premiere pro UI
Source: self-made

Below are the the layers used in the video: the ground with the skeleton (animations), the background for the banner and credits, the background without floor, the credits and the banner. The video is a combination of juxtaposed static elements and animation

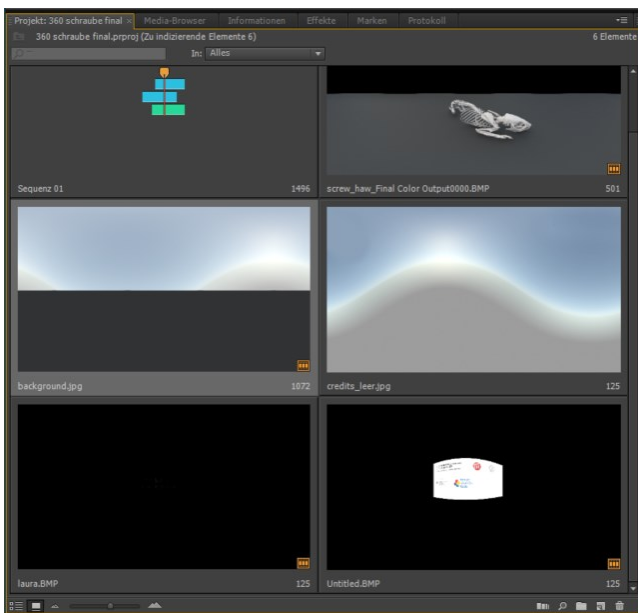
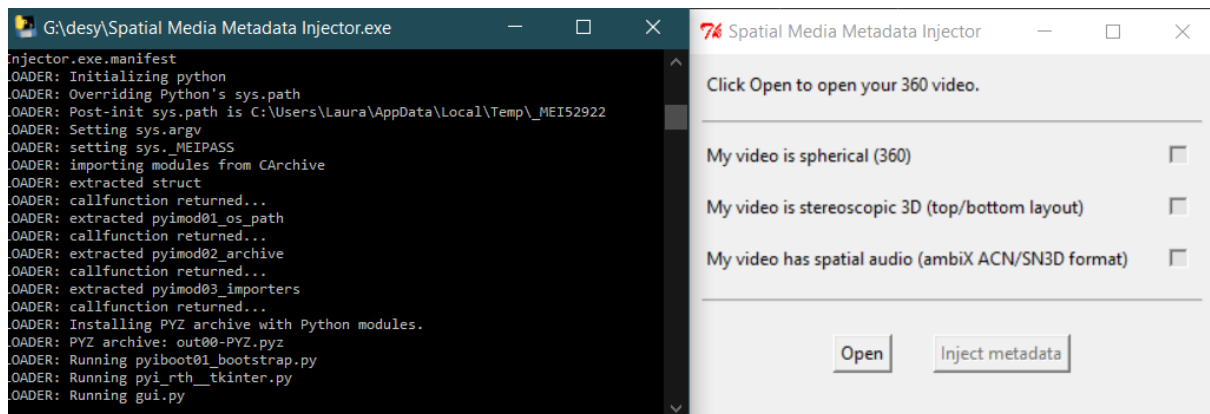


Image 23 Separately rendered objects imported to video library in Premiere Pro
Source: self-made

The video credits, institute logos and rendered scenes were brought together in PremierePro. To make the video apt for loop playing the animation was repeated twice once forward and once backwards and the end frame looked exactly like the start frame. The video was exported once in the highest quality possible for YouTube with following configuration:

- Resolution: 1920 x 1080
- Frame rate: 29.97

- Format: MP4
- Codec: H.264



*Image 24 Spatial metadata injector interface
Source: self-made*

After rendering the video, the metadata of the mp4 file was changed using a free metadata injector recommended by YouTube. This step is necessary so that the video is identified correctly as a 360 degree on any device or on most online video streaming services (“YouTube Hilfe 360° Videos hochladen,” n.d.). For this case the first option “my video is spherical” was selected because the video was rendered as a flat equirectangular projection.

A semi-final version of the video was uploaded to YouTube and shared for a last review before showing it to the public. It was very important that the video was in a format that could adapt and be streamed or played on all HMD and most smartphones and pcs. The video was showed the first time in a conference of the Helmholtz Center Geesthacht in July 2017 using a PS4 with a PlayStation VR in the pre-installed media player. The video and visualisations were later shared by the scientists and seen by event visitors and employees of the institute a few months after the event in their own devices. The video reached up to 66 views between July and the present day most likely of employees of the Helmholtz Center and event visitors in YouTube as shown in Chart 2.

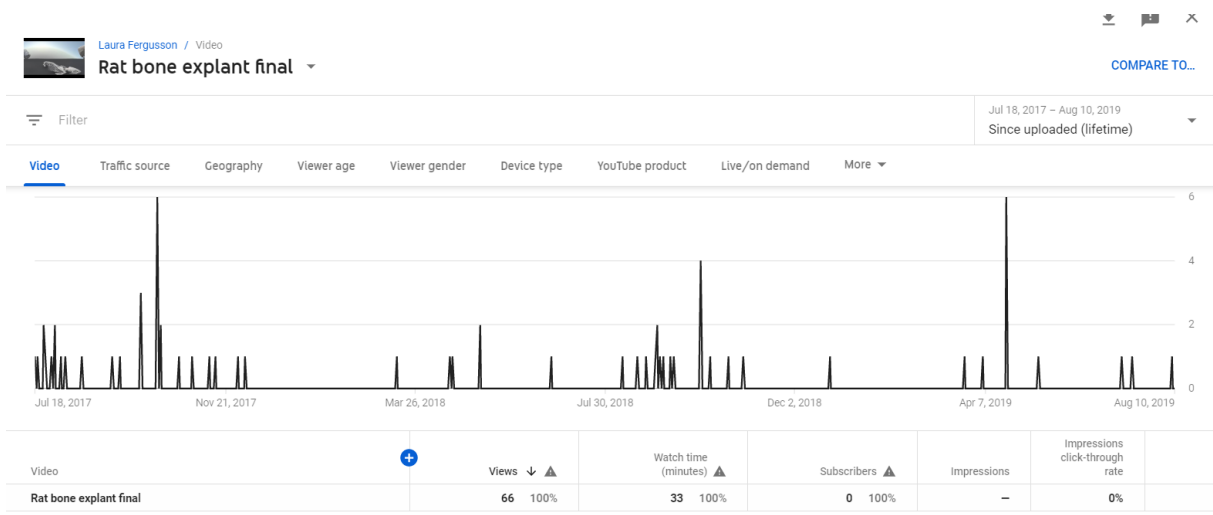


Chart 2 Final video stats in YouTube (views from July 2017 until August 2019)
Source: YouTube

This might not seem as a lot of views for a YouTube video but these are views of people that were really interested in the project (they actively searched for it in YouTube after watching it in an event in a HMD or colleagues shared with them the link). The views in the HMD are not counted on these stats because the video was played locally in the PS4 instead of using the YouTube version with the aim to see the interest of people in the topic after the event. As a content creator viewers here might be potential clients for creating similar projects.

4.2.3 STATIC IMMERSIVE VISUALIZATIONS

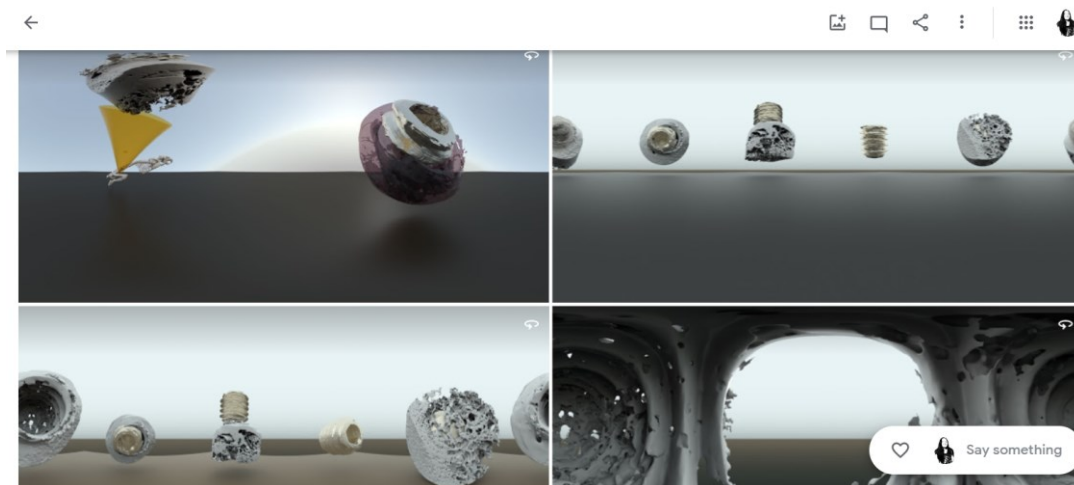
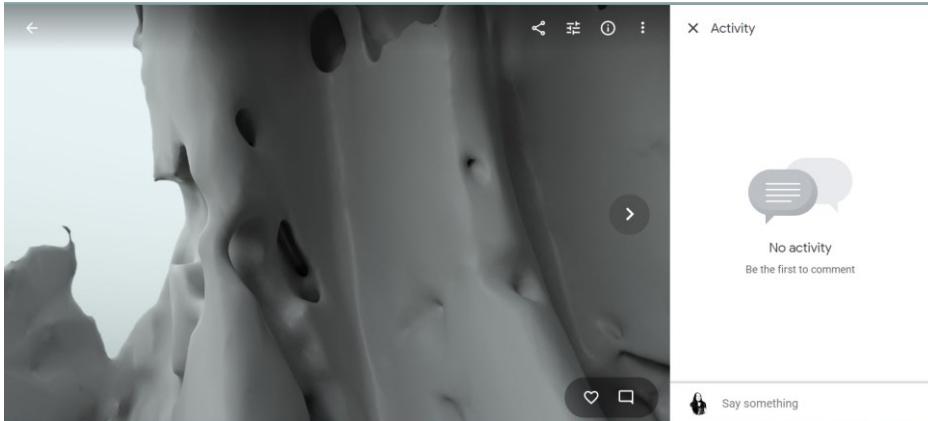


Image 25 Google Photos 360-degree preview gallery
Source: self-made

Tools such as Google Photos can host and display 360-degree images for free online which means that anybody with a PC, modern smartphones or headsets can get an idea of the work in

progress of the immersive experience. Static visualisations can be very practical to save up rendering resources and to gather feedback: instead of rendering a scene for hours, single frames in higher quality can be quickly presented to the people involved in the project to discuss the videos look and feel before rendering the whole animation.



*Image 26 comment function for google photo gallery
Source: self-made*

the inside of the bone opened on google images as 360° panorama here is where teams can give structured feedback on the rendering phase.

For this project high definition static visualisations were generated and uploaded to google images to have easily accessible sneak-peeks accessible on any device enabling anybody with an URL to the album to comment or like different creative approaches and storylines. The process to generate these images is similar to the one generating the video frames for a 360 video with the minor difference that only single frames are rendered in a bigger size and that the aspect ratio should be 2:1 (e.g. 4096 x 2048).

4.3 REFERENCE USE CASES AND SCENARIOS FOR IMMERSIVE MEDIA

To develop a better storyline for the DESY project the author first looked at successful VR projects and research in journalistic, entertaining and scientific fields to get inspired. In the following 3 chapters a summary of the most inspiring use cases is discussed.

4.3.1 JOURNALISTIC VR

Immersive journalism is a content creation concept in which viewers can have first-person experiences of the events or situation described in news stories. Immersive journalism is fundamentally a medium to take viewers into a virtually recreated scenario representing the news story.

journalism aims to create a connection between the audience and the news story. Immersing the audience in the story is an ideal way to achieve this. Current approaches to immersive journalism are usually “avatar experiences” framing histories in immersive systems. In an avatar experience could for example viewers experience being in an interrogation room with real time interrogators in an offshore prison. The experience seems real except for the bodies of the people interacting: the body of the viewer or any other connected person is usually replaced by an avatar or a virtual body. In some cases, depending on the extent of body tracking, the movements of the virtual body will be same as the movements on the viewers body.

The viewer can enter the story in different roles: instead of as oneself as a non-interacting external viewer, or as a character in the news story. Role-playing demands direct interaction with the environment and other viewers or actors in the scene, a “first person” experience in this case does not mean the viewer is the subject of the story.

The role-playing format has been used in exhibitions at the Newseum (Interactive News Museum in Washington, DC). Here, viewers are in a journalist role where they for example must interview people resulting into an insight into the issue. In most virtual scenarios, users have a rather lower virtual presence which in the worst-case scenario can lead to disengagement or withdrawal from the virtual world.

Place illusion (PI) plausibility (Psi) and virtual body ownership (VBO) are the three main aspects that should be considered as important to generate high quality immersive journalism. Place illusion is the sense of being in the place currently being seen. Total immersion happens when the viewer does not differentiate between the real or virtual presence, as researchers described it:

“the sensation of being and operating at a remote or virtual place; and plausibility—the illusion that what is happening is really happening”

As for VBO recent research in cognitive neuroscience in this field has evidenced that the brain has a high degree of flexibility of the representation of the body in humans. Inducing virtual illusions of body distortion can be easily “believed” by viewers as long as the sense of ownership of an entire virtual body is cohesive to the real body movements. This can be only achieved by having good Hard- and software: head and body tracking, gyroscope and magnetic sensors are responsible for mapping movements of the body.

Plausibility is particularly relevant when it comes to media experiences because it is concerned with the dynamics of events and the story being told. How likely is the situation in

the eyes of the viewer to happen in real life? are there events that personally relate to the viewer?

The ever-growing amount of information available on media outlets today is a possible reason for people's indifference to topics involving human suffering. A potential key role of immersive journalism is to regain the audience's emotional involvement in current events. Virtual reality systems are therefore an effective medium to deliver engaging first-person experiences of stories aiming to maximise and spread social responsibility amongst individuals.

4.3.2 ENTERTAINING PURPOSES

VR experiences have been implemented since the mid-90s but remained up until the last years an isolated market due to lacking availability of hardware (plug and play) and software solutions for developers and end-users. One of the first massively shown VR experiences about Aladdin's story was exhibited in Disneyland in 1995. The creators of this experience documented and analysed the results of their work as discussed below.

Disney Imagineering developed a VR attraction where guests were invited to fly a magic carpet through a virtual world based on the animated film "Aladdin." Researcher focused on using VR as a new medium to tell stories. The VR experience was observed for a period of fourteen months where controlled experiments took place documenting the reactions of about 45,000 guests.

The main finding is that in a good quality VR experience, men and women of all ages are immersed in the illusion that they are in a different place. In VR, as in all media, content is important. Most guests are not impressed with the technology for its own sake; the majority cares about the graphics and the story being told. The experience can be improved by simply telling the participant a "background story" and by giving a task to perform while being immersed.



Image 27 Physical setup of the Aladdin experience
Source: Pausch et al. 1996

The HMD display and set-up implemented was built by the researchers, the HMD rendered a single horizontally wide graphical window. Additionally, a combination of stereo ambient sound, with eight channels of localized sound were implemented (binaural sound) the soundtrack had background voices, animals, and other background noises.

Following main findings were discovered by researchers:

1. Guests focused on what there was happening in the virtual world instead of in the hardware. Most guests did not care how VR works, they focused on what they felt instead.
2. Several guests reported the sensation of being “in” the scene: this made the experience for some overwhelming, some were so absorbed that they had trouble answering questions early in their flights.
3. Guests asked for a goal. If not given a specific goal, guests would ask “What should I be doing?”
4. Guests needed a background story. We found that giving as much context as possible about the scene helped reduce the severity of the transition from the real to the virtual environment.
5. Guests tend to explore fast and seek new spaces. Most guests don’t spend time studying in detail a scene.

6. Guests did not turn their heads often. Either due to the fact that they were piloting a vehicle, or because they were not used to being able to turn their heads when wearing a HMD.

VR must be personally experienced, pilots' observers enjoyed the show and grasped that something fascinating was going on the pilots' side, but it was clear that VR seen from another perspective seems much less interesting than what it is (Pausch et al., 1996).

4.3.3 SCIENTIFIC PURPOSES

VR is a medium well suited for behavioural research, as one interesting study from Cornell university food evidences. For this food tasting experiment, scientists implemented Headsets and 360° videos to find out how the environment where the food is eaten influences the perceived food taste. Scientists aimed to find out how people's perception of real food can be altered by their surroundings.

About 50 participants were asked to have a meal with HMDs on their heads. The participants saw 3 virtual scenes: a pleasant park bench, a cow barn and a sensory booth. Participants had to eat one piece of cheese per virtual scene and were asked after which one of the cheeses had a better taste. The participants were unaware that the cheese samples were identical and rated the flavour intensity of the blue cheese significantly higher in the cow barn scene than in the other two scenes (Stelick et al., 2018).

"We consume foods in surroundings that can spill over into our perceptions of the food" as stated by Dando. After yielding clear results researcher Dando concluded that a VR testing setting is convenient, and more flexible than building physical environments for test subjects: *"This research validates that virtual reality can be used as it provides an immersive environment for testing"* ("Eating with your eyes," n.d.).

Another inspiring use case for of VR scientific implementation is implemented in the Helen F. Graham Cancer Center in Delaware to distract cancer patients during chemotherapy so that they experience less pain. Nearly 200 patients have tried VR during chemo sessions and reported feeling less anxiety and pain during and after the sessions. The initiative started when one of the doctor's tried this method to alleviate pain of one patient that was suffering during chemotherapy.

The VR scenes for chemo patients are pleasant environments such as: Bavarian Alps, Ireland landscape, Australian beach preserve, winter forest, autumn forest and lush meadow. This simple but effective idea brought national recognition to the centre: the centre was awarded with the Magnet award from the American Nurses Credentialing Association including a

\$50,000 prize. The was used to create a virtual reality program for chemotherapy patients in other treatment centers (“Delaware patients use virtual reality to get through chemotherapy,” n.d.).

5 OUTLOOK: FUTURE (PERVASIVE) USES OF IMMERSIVE MEDIA CONTENT

5.1 MASSIVE COMMERCIALIZATION OF HARDWARE FOR IMMERSIVE EXPERIENCES

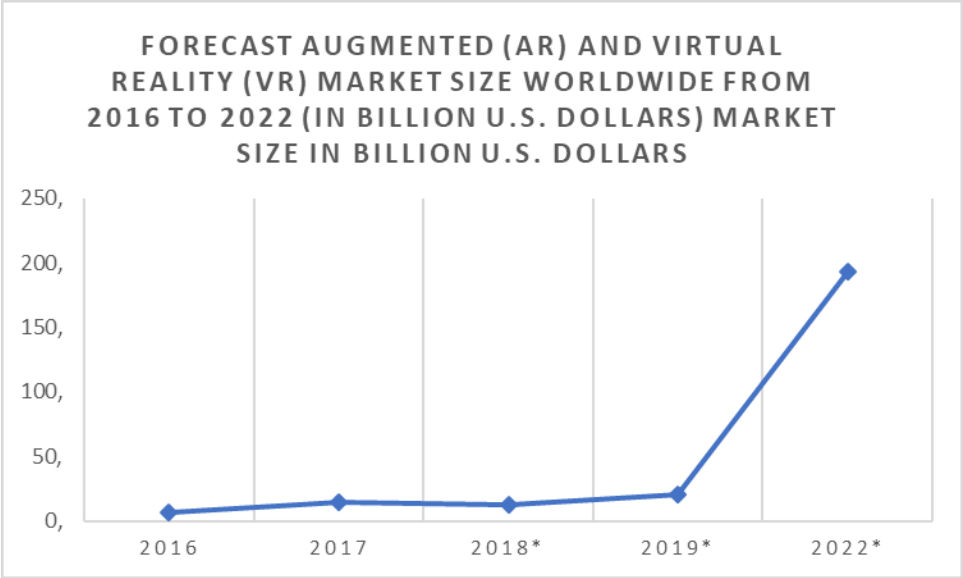


Chart 3 Projected size of the augmented and virtual reality market 2016-2022
Source: IDC, Statista estimates

As evidenced on Chart 3 the forecast for a pervasive use and revenue for VR and AR is on the rise. As mentioned on chapter 3 and VR technologies and devices are on their way to a higher affordability, usability and availability for mobile users. On the other hand, VR user generated content, experience design and browsing are on its initial experimental phase where larger groups of amateurs and developers can find a lot of resources and possibilities online to generate content. VR and AR are on the rise.

Some people have already known this for a while and are ensuring a revenue model before the technology has reached the stage it will reach in a near future. A patent filed the first time in the year 2000 now owned by Google describes a method and apparatus for generating a customized dynamic virtual reality environment.

The patented VR environment is accessible through the TV and uses a virtual personal assistant (artificial intelligence) to communicate with the end user. Information obtained from the conversation will be interpreted into a request for information, services or products. The main goal is to provide an easy to use virtual reality environment allowing users that are not familiarized with PCs or users that are not able to read to be able to interact with the service.

The artificial intelligence assistant receives commands through a speech recognition device and adapts with time to each individual user in order to provide customized guidance to the user. The intelligent assistant imitates human features and behaviours according to the preferences of its user and has a human avatar.

the Internet today is a chaotic and overflowed with useless information accessible only to people with access to end-devices such as pcs and smartphones. The virtual environment will bypass the computer/device dependency for browsing the web. Additionally, Web pages are not designed for a person, but in the best-case scenario for a particular group.

The Internet will unleash its real potential when its enabled to be used as mass medium accessible through voice interaction and AI. By using AI, creating an environment focused on its users' expectations and wishes possible because the data gathering for anticipating individual information, Services, entertainment, and transactional needs in real life becomes seamless.

“A primary feature of the virtual environment of this invention is that it should, to the possible extent, make the technology disappear. Currently, users must directly deal with three artifacts of technology—operating systems, user interfaces and browsers—before they even get to deal with applications that could possibly affect their lives. Today's operating system/user interface/browser combinations are not truly intuitive nor user friendly.” (Gaos, 2008, p. 23)

The customized virtual environment has been patented for 19 years as of today. It is as well briefly stated on the patent that users still have rights and own the data stored on their environments, an explanation on how the data is protected and where it is sent to remains unclear. For VR content creators such an invention would open up several content creation opportunities, that in not so very far away future could be designed without needing an end device as known today (Gaos, 2008).

5.2 USER GENERATED VR CONTENT AND ONLINE COMMUNITIES

5.2.1 GOOGLE STREET VIEW

One of the first social websites that is still currently being used and hosts 360° panoramas in maps for large audiences is Google Earth and google maps. Google maps relies heavily on user generated content to gather information about places and encourages users to take pictures, rate places and even register as a business owner in a designated location. For

content creators this offers a platform with a large audience to showcase geotagged content that might trigger a conversation with other users. It is important to bear in mind that the social component and openness (anybody can comment or upload) in google maps might also open a discussion beyond the content (panoramic images) itself but around the context, language and culture in which they were published. A recent study investigated how user's perspective of the world and territory influences the quality and type of the content uploaded to google maps

“this study argues for ways that can users recontextualize and subvert ‘master representations’ of visual media within the authorial structure rather than re-authoring the existing software and structures.” (Farman, 2010, p. 869)

Google earth is since 2005 one of the many geographic information systems (GIS) available online for free. The researchers posed themselves the question if any colonialism could be present in the ‘neutral’ technology of Google Earth. By drawing a comparison between this popular GIS and colonial history of cartography researchers discuss the cultural implications the software is subject to.

Google bought the originally called Earth Viewer and owned by Keyhole, Inc. in 2006 and renamed it after to google earth. On the free version of Google Earth, the average zoom in major cities such as London is significantly better than in other less populated areas and other continents such as Africa the quality of imagery is significantly less.

The integration of user generated Panoramic images (equirectangular projections) at street level were added in mid-2008. Some locations have significantly more content than others. For example, there is a flying tour through the Grand Canyon created based on topographic information is available and a ‘flight simulator’ to dive into big cities with skyscrapers.

the ‘Google Earth Community’ is the user network that collaborates and generates content. Community members communicate in a spatial Bulletin Board System (BBS) which has been there almost since the beginning of google earth. Anybody belonging to the community can contribute placemarks with information about a location or create ‘overlays’ that offer an update such as a new or modified path of Cyclone Nargis and the affected areas in Myanmar. These overlays can be downloaded and implemented by any end user. Allowing people to contribute information shapes the map opening a debate about where the borders of a location or a certain route could be.

User generated maps cannot be regarded as an objective visualization of a territory because the content is heavily influenced with the culture of the society that creates them. Maps are used because they are thought to be reliable and usable. A Map depicting connected to

'reality' is essential to our everyday navigation of lived space which can be implemented from planning your way to work to something as deadly as to map out targets for missile attacks into Israel.

“The main problem with this expectation of maps and GIS ‘representing reality’ is that it assumes such representations are neutral and outside of cultural interpretation.” (Farman, 2010, p. 874)

Platforms like Google Earth for the good or the bad enable content creators to show the ways cultures inhabit mixed reality spaces encouraging a seamless collaboration between the existing landscape and digital interface. Content creators should question the ways that these interfaces (digital and material) culturally situate their own methodology in order to promote or discourage others to start a critical discussion. Location aware programs where geotagged data is published to wide audiences will continue to change the way users define and inhabit space as embodied interactors (Farman, 2010).

5.2.2 VR SOCIAL MEDIA: A PARALLEL REALITY?

5.2.2.1 STARAMBA SPACES

VR worlds preparing to welcome large audience members are in development and will be available soon, one of them is called Staramba.spaces. The creators own a company specialized in creating realistic 3d avatars of famous people for VR and games. Staramba aims to create the first ultra-realistic virtual reality social network with services and experiences built around the real world's mainstream celebrities, athletes, and musicians.

Staramba.spaces will be accessible in the near future (no exact date is provided) to any user owning a HMD. The spaces or VR worlds will be places where users can be immersed in differing virtual worlds where they can meet friends, walk around, go shopping, chatting, party, watch movies, go to events and mingle with strangers (“Staramba – Touch the Future - 3D Instagram,” n.d.).

To create users' realistic avatars the company offers a body scanning service with a special full body scanner. The “3D Instagram” body scanner is a high-resolution 3D scanner which has been already used to scan and create real looking avatars of athletes, musicians, celebrities for Hollywood movies and video games.

It is as well planned that users will be able buy virtual items by using the VR worlds currency called STARAMBA.token. With this currency users will be able to buy virtually anything they

can already buy in real life, including real estate. The token price will start between \$0.07—\$0.10 during the initial phase and rise to \$1 when the spaces are opened to the public. Tokens are already available for sale to anybody and early access to the VR world is currently being given only to investors (Staramba, 2018).

5.2.2.2 VR COFFEE SHOP TO WATCH TV

Another player in VR social media is a French start up called Molotov, this company is on its way to become the leading platform to stream TV in France. Users are able to watch TV on almost any device with one account. Molotov releasing soon a VR app that will enable users to watch TV using a HMD.

Nowadays the huge amount of content and information in Netflix and similar libraries allows users to decide what to stream whenever wherever they want. Streamers have left behind the “classical” way of watching TV where choices and devices were limited. Molotov aims to bring back the social component of watching TV together with others by allowing users to watch TV in a VR environment with others together. This kind of way of watching TV usually comes up in public viewing of football matches, election nights and other important events. Users will have an added value to their watching tv experiences having side conversations and jokes with

The VR experience will be sold under the name Molotov Together in a virtual coffee shop as a multiplayer experience. Molotov currently has around 7 million users in France from which 1.2 million users stream something daily as of today by themselves and maybe virtually together in a near future too (“Molotov creates a VR coffee shop to watch TV together,” n.d.).

VR might seem as an escape to another reality to some, an imaginary universe that a content creator and game developer built to distract people from real life. Viewers choose to be immersed in these virtual worlds that can be one the positive side extremely interesting but on the down side highly absorbing. However, humans are social animals by nature, in search of a community they feel they belong to. As technology enables people to connect in different ways the search for community might be possible for some in a virtual reality. Both existing user generated content sites with immersive media and the VR massive experiences to come have a common focus: enabling community and user interaction. Social VR its on its early stages and will potentially be a widely used medium to have a “second life” filled with virtual experiences and real relationships in artificial worlds.

6 BIBLIOGRAPHY

6.1 IMAGE INDEX

Image 1 - google lens mobile interface identifies text on an object Source: self-made.....	5
Image 2 - object identification in google lens mobile app Source: Self-made	5
Image 3 - Start of the Virtual experience where the user is part of a virtual audience Source: (Ladeira, Marsden, and Green 2011)	7
Image 4 - Interactive Exhibitions in Museums: Three drops (left) and Boundary Functions (Right) Source: Snibble and Raffle, 2009	10
Image 5 - System Architecture of EasyTour Source: (Mazzoleni et al. 2006)	14
Image 6 - Equirectangular grid with Up, Back, Left, Forwards, Right, and Down Source ("Drawing a Spherical Panorama" 2016)	15
Image 7 viewer image projection onto a perpendicular plane Source: Sandnes and Huang 2016	16
Image 8 plane to panorama mapping Source: Sandnes and Huang 2016.....	17
Image 9 square room example illustration before and after translation Source: Sandnes and Huang 2016	17
Image 10 eight camera rig for 35mm lens cameras Source: Sarmiento and Quintero 2009..	18
Image 11 simplified cave monitor installation Source: Sarmiento and Quintero 2009	19
Image 12: curve segments in an equirectangular image Source: (Ho Oh et al. 2012)	20
Image 13 optimized mapping from the viewing sphere with a mixed perspective approach Source: Carol and Agrawala 2012	23
Image 14 image processing example and perspective comparison Source: Carol and Agrawala 2012	23
Image 15 end-to-end architecture for selective VR rendering Source:(Lungaro et al. 2018)..	26
Image 16 system architecture in Unity - how to optimize VR apps in Unity with Furion Source:(Lai et al. 2017).....	28
Image 17 equirectangular render from MODO Source: Self-Made	31

Image 18 Tomography (raw 3d data) of the screw inserted in a rats femur Source: DESY Hamburg	31
Image 19 Modo UI Source: self-made	32
Image 20 Camera Settings in MODO Source: self-made.....	32
Image 21 Environment and Material settings in MODO Source: self-made	33
Image 22 Adobe Premiere pro UI Source: self-made	34
Image 23 Separately rendered objects imported to video library in Premiere Pro Source: self-made	34
Image 24 Spatial metadata injector interface Source: self-made	35
Image 25 Google Photos 360 degree preview gallery Source: self-made	36
Image 26 comment function for google photo gallery Source: self-made	37
Image 27 Physical setup of the Aladdin experience Source: Pausch et al. 1996	40

6.2 LITERATURE AND WEBSITES

- A-Frame Introduction [WWW Document], n.d. URL <https://aframe.io/docs/o.8.0/introduction/> (accessed 12.20.18).
- Anderson, D., 2011. The British Museum.
- Araujo, A., 2017. Guidelines for drawing immersive panoramas in equirectangular perspective, in: Proceedings of the 8th International Conference on Digital Arts. pp. 93–99.
- Ballard, D.H., 1981. Generalizing the Hough transform to detect arbitrary shapes. Pattern recognition 13, 111–122.
- Canny, J., 1987. A computational approach to edge detection, in: Readings in Computer Vision. Elsevier, pp. 184–203.
- Carroll, R., Agrawal, M., Agarwala, A., 2009. Optimizing content-preserving projections for wide-angle images, in: ACM Transactions on Graphics (TOG). p. 43.
- Delaware patients use virtual reality to get through chemotherapy, n.d. . WHYY. URL <https://whyy.org/articles/delaware-patients-use-virtual-reality-to-get-through-chemotherapy/> (accessed 12.21.18).
- Drawing a Spherical Panorama, 2016. . D. M. Swart. URL <https://dmswart.com/2016/06/28/drawing-a-panorama/> (accessed 4.20.19).
- Eating with your eyes: Virtual reality can alter taste [WWW Document], n.d. . Cornell Chronicle. URL <http://news.cornell.edu/stories/2018/09/eating-your-eyes-virtual-reality-can-alter-taste> (accessed 12.14.18).
- Farman, J., 2010. Mapping the digital empire: Google Earth and the process of postmodern cartography. New Media & Society 12, 869–888.
- Gaos, M., 2008. Method and apparatus for delivering a virtual reality environment. Google Patents.

- Glitch - Fog Creek is now Glitch! [WWW Document], n.d. . About Glitch. URL <http://about-glitch.glitch.me/about/> (accessed 12.20.18).
- Irvine, K., Category: #Design, F.D.A., October 31, C.P. on, 2017., n.d. XR: VR, AR, MR—What's the Difference? | Viget [WWW Document]. <https://viget.com>. URL <https://www.viget.com/articles/xr-vr-ar-mr-whats-the-difference/> (accessed 11.20.18).
- Kuryliak, O., 2018. Everything You Need to Enjoy VR On Mobile in 2018 [WWW Document]. Virtual Reality Pop. URL <https://virtualrealitypop.com/everything-you-need-to-enjoy-vr-on-mobile-in-2018-dca39a004385> (accessed 11.20.18).
- Ladeira, I., Marsden, G., Green, L., 2011. Designing interactive storytelling: A virtual environment for personal experience narratives, in: IFIP Conference on Human-Computer Interaction. pp. 430–437.
- Lai, Z., Hu, Y.C., Cui, Y., Sun, L., Dai, N., 2017. Furion: Engineering High-Quality Immersive Virtual Reality on Today's Mobile Devices, in: Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking. pp. 409–421.
- Lungaro, P., Sjöberg, R., Valero, A.J.F., Mittal, A., Tollmar, K., 2018. Gaze-aware streaming solutions for the next generation of mobile VR experiences. *IEEE transactions on visualization and computer graphics* 24, 1535–1544.
- Mazzoleni, P., Valtolina, S., Franzoni, S., Mussio, P., Bertino, E., 2006. Towards a contextualized access to the cultural heritage world using 360 Panoramic Images., in: SEKE. pp. 416–419.
- Merleau-Ponty, M., 2013. *Phenomenology of perception*. Routledge.
- Michel, G., 2008. *Le café du cinéma Sauvenière*, 360°.
- Molotov creates a VR coffee shop to watch TV together, n.d. . TechCrunch. URL <http://social.techcrunch.com/2018/12/11/molotov-creates-a-vr-coffee-shop-to-watch-tv-together/> (accessed 12.14.18).
- Oh, S.H., Jung, S.K., 2012. A Great Circle Arc Detector in Equirectangular Images., in: VISAPP (1). pp. 346–351.
- Palmer, S.E., 1999. *Vision science: Photons to phenomenology*. MIT press.
- Pausch, R., Snoddy, J., Taylor, R., Watson, S., Haseltine, E., 1996. Disney's Aladdin: first steps toward storytelling in virtual reality, in: Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques. pp. 193–203.
- Sandnes, F.E., Huang, Y.-P., 2016. Translating the viewing position in single equirectangular panoramic images, in: Systems, Man, and Cybernetics (SMC), 2016 IEEE International Conference On. pp. 000389–000394.
- Sarmiento, W.J., Quintero, C., 2009. Panoramic Immersive Videos-3D Production and Visualization Framework., in: SIGMAP. pp. 173–177.
- Snibbe, S.S., Raffle, H.S., 2009. Social Immersive Media
Pursuing Best Practices for Multi-user Interactive Camera/projector Exhibits. *CHI 2009 New Media Experiences* 2 2, 1449–1456.
- Staramba, 2018. What is STARAMBA all about? Staramba. URL <https://medium.com/@Staramba/what-is-staramba-all-about-17b0ea5072aa> (accessed 12.21.18).
- Staramba – Touch the Future - 3D Instagram [WWW Document], n.d. URL <https://company.staramba.com/3d-instagram> (accessed 12.21.18).
- Stelick, A., Penano, A.G., Riak, A.C., Dando, R., 2018. Dynamic context sensory testing—A proof of concept study bringing virtual reality to the sensory booth. *Journal of food science* 83, 2047–2051.
- SUPERMEDIUM - virtual reality browser [WWW Document], n.d. . SUPERMEDIUM - virtual reality browser. URL <http://supermedium.strikingly.com/> (accessed 12.20.18).
- Thomas, F., Johnston, O., n.d. *Disney Animation: The Illusion of Life*. 1981. Abbeville Press.
- VR Brillen Vergleich, n.d. . VR·Nerds. URL <https://www.vrnerds.de/vr-brillen-vergleich/> (accessed 12.14.18).

YouTube Hilfe 360° Videos hochladen [WWW Document], n.d. URL
<https://support.google.com/youtube/answer/6178631?hl=de> (accessed 8.10.19).