

Challenges of and Recommendations for Combining 6-DOF Spatial VR-Interaction with Spherical Videos

This paper was downloaded from TechRxiv (<https://www.techrxiv.org>).

LICENSE

CC BY 4.0

SUBMISSION DATE / POSTED DATE

06-01-2020 / 08-01-2020

CITATION

Wiebel, Alexander; Keuchel, Sascha; Eckenfels, Magnus; Drengner, Jan; König, Werner (2020): Challenges of and Recommendations for Combining 6-DOF Spatial VR-Interaction with Spherical Videos. TechRxiv. Preprint. <https://doi.org/10.36227/techrxiv.11525586.v1>

DOI

[10.36227/techrxiv.11525586.v1](https://doi.org/10.36227/techrxiv.11525586.v1)

Challenges of and Recommendations for Combining 6-DOF Spatial VR-Interaction with Spherical Videos

Alexander Wiebel, Sascha Keuchel, Magnus Eckenfels, Jan Drengner, and Werner König

Abstract—Virtual reality (VR) and spherical 360 degree videos are often used to provide users with impressions of locations or events which are hard to access or, in the case of events, are only held at special times. In this paper, we first describe two examples where users were immersed in spherical videos of real events and at the same time were allowed to interact with computer-generated spatial VR content. This combination can help to make users more interested and involved in the content of the videos and thus the events, but it also poses a number of challenges resulting from the two-dimensional nature of the videos contrasting the three-dimensional nature of the interaction and the six degrees of freedom (DOF) in viewing the VR content.

The main part of this paper discusses and analyses the challenges arising from this combination, it introduces a classification of such combinations that readers can apply to their own application cases, and it provides recommendations for dealing with the different challenges where possible. The recommendations were evaluated in a public setting with over 1,000 users.

Index Terms—Virtual reality, 360 video, combination.

1 INTRODUCTION

OFTEN *digital* representations of inaccessible locations, rare cultural events and remote touristic destinations [5], [16] are used to provide users with impressions which would be hard to obtain in reality. In recent years, such digital representations have frequently either been interactive and stereoscopic virtual environments (VE) or spherical videos. Less often combinations of both have been used.

The discussion in this paper is based on an exemplary virtual reality (VR) application (or *game*) in which users were immersed in spherical monoscopic videos of real events. During their experience they were allowed to interact with additional stereoscopically rendered virtual objects. This combination has been chosen in order to make users more interested in and involved with the content of the videos and thus the events presented in the videos. During the development of the application a number of challenges resulting from the two-dimensional nature of the videos contrasting the three-dimensional nature of the interaction and the spatial content being viewed with six degrees of freedom (6-DOF) became apparent and were addressed. Thus the main contributions of this paper are

- the discussion and analysis of the challenges arising from the combination,
- the introduction of a classification of such combinations,
- a zone-based approach to describe how and why different combinations of spatial VR and monoscopic 360 degree videos work,

- recommendations for dealing with the challenges appearing in the context of the different classes of combinations,
- and an evaluation of the recommendations in a public setting with over 1,000 users.

1.1 Terminology

The term virtual reality is often used to refer to different experiences [4], [11], [15]. We will call two of them *spatial VR* and *360-video* in this paper. As this paper deals with a combination of these two variants, we want to avoid any confusion and begin with definitions describing the most important properties of both.

When talking about *spatial VR* we are referring to computer-generated and stereoscopically rendered virtual environments. The users of a *spatial VR* can interact with virtual 3D objects (sometimes called *game objects*) using spatially tracked devices (*controllers*) and move physically in the environment. The environment is presented to users using special hardware capable of head-tracking and stereo rendering, in our case head-mounted displays (HMD).

By *360-video* we are referring to pre-recorded monoscopic video where views in all scene directions are recorded at the same time. Such videos are presented to users in a way that allows them to look around in all directions from one fixed location without other possibilities for interaction. Usually, this is achieved with displays whose orientation is tracked using special sensors, in our case HMDs again. Often the video is stored in an equirectangular format and rendered on the inside of a sphere the viewer is located in [6]. The literature uses many different names to refer to *360-video*. Among them are 360-degree video [1], immersive video [1], surround video [9], omnidirectional video [14] and spherical video [8].

• All authors are with Hochschule Worms University of Applied Sciences, Germany. Corresponding author e-mail: wiebel@hs-worms.de

Manuscript received MONTH DD, 2020; revised MONTH DD, 2020.

One of the main differences of spatial VR and 360-video are the degrees of freedom of the users. While users can move and look around with 6 degrees of freedom (DOF) when experiencing spatial VR, they are limited to looking around, i. e. 3 DOF, in 360-video experience due to the fixed location of the originally recording camera. Extensions of video to 6 DOF exist [7] and are usually called volumetric videos, but unfortunately such videos are much more complex and expensive to acquire.

2 MOTIVATION: APPLICATION SCENARIO

The exemplary VR application (or *game*) which is the basis for the discussion in this paper has been developed in the context of a research project on digitalization in event marketing and destination marketing [2]. In the research project called *schaz*, a pervasive mobile game [10] (*schaz-App*) as well as the mentioned VR application (*schaz-VR*) have been developed and evaluated. Both applications were presented at *Rheinland-Pfalz-Tag 2018* (RLP-Tag) to address the research questions of the project. RLP-Tag, a festival of the German state of Rhineland-Palatinate which took place in the city *Worms* in June 2018, was the perfect platform for this research because the festival attracted over 300,000 visitors.

While *schaz-App* is mainly related to artifacts and attractions physically existing in *Worms*, the *schaz-VR* application was introduced into the concept in order to promote annual cultural events that take place in *Worms*. In particular, the goal of the VR-application was to provide the players with 360-video experiences of such cultural events from the preceding year in order to make them interested in visiting the events in the future. Thus 360-videos of four such events have been recorded: *Nibelungen-Festspiele*, a theater festival; *Worms: Jazz&Joy*, a music festival; *Backfisch-Fest*, a big wine festival and funfair; *Spectaculum*, a Renaissance fair.

2.1 Pre-Test Version of the VR Application

A first version of the VR application was presented during the *Day of German Unity* celebrations 2017 in the city of Mainz, which can be seen as a kind of pre-test. More than 200 guests used the application and feedback of 59 of them has been collected by a questionnaire. In this first version, there was one small VR level in which the players were informed playfully about the upcoming *Rheinland-Pfalz-Tag*. At a key position in the game, the players were able to trigger different 360-videos by grabbing provided game objects with a game controller in their hand. In particular, a standard 6 DOF (degrees of freedom) controller was used. The special game objects reflected the themes of the corresponding 360-video and activated them. For example, players were able to start a 360-video of the music festival by grabbing a microphone, or start a 360-video of the Renaissance fair by interacting with a sword.

While watching the 360-video, the players still held the game object in their hands. At that time they were not able to trigger any interactions with that game object (Figure 1, left image). However, observation of the players as well as the collected feedback showed that, despite the otherwise very positive feedback, players would have liked more



Fig. 1. Difference between two versions of *Spectaculum* 360-video. The image on the left side demonstrates the situation of the prototype. Here, players had a sword in their hand, no interaction could be performed except for swinging the sword. The image on the right side illustrates the new situation. Players can use their sword to hit the shield of their opponent. Positive feedback by rendering splinters and adding bonus time has been implemented.

interaction in the video and thus left it soon because it became “boring” for them. This behavior of the players was undesirable, because the most important goal of the application was to promote the tourist attractions and events of the city to the interested players and not only to entertain them. With this target in mind, we decided to provide more interactivity and thus also more immersion in the 360-video levels for the final version of the application.

2.2 Final Version of the Application

For the final version of the application, called *schaz-VR*, the concept to let guests play spatial VR levels leading them into the 360-videos has been retained (see Figure. 2). This approach allowed us to teach the players how to control a certain game object before they entered the video level where they would be able to use the same objects for interaction. This strategy can be best described as letting players “experiment with, learn and apply” interaction possibilities, and was implemented for each 360-video level. An overview of the level structure can be found in Figure 2.

After a short interactive tutorial in the *start room* (Figure 2, top), each guest enters the game in the so-called *treasure room* (Figure 2, center), which can be considered a base camp. The players can experiment with a variety of game objects, learn to orient themselves in the virtual environment and discover the key elements that can lead them to another spatial VR level by opening a portal in the middle of the treasure room (Figure 4, left). The key objects are primarily used to open the portal to the next level, but they are also the tools the player needs to use to complete certain tasks in the upcoming spatial VR Level. Defeating an animated 3D knight with a sword would be an example for such a task (Figure 2, *Spectaculum* level). In addition, each key object serves as a metaphor for the related cultural event we want to present to the player.

In the 360-video level players can use the key element to interact with the video environment, like fighting another knight, this one being a 2D protagonist in the 360-video (Figure 1). To motivate players to interact in the 360-video levels, a reward system has been implemented. The more interaction is measured, the longer the player is allowed to

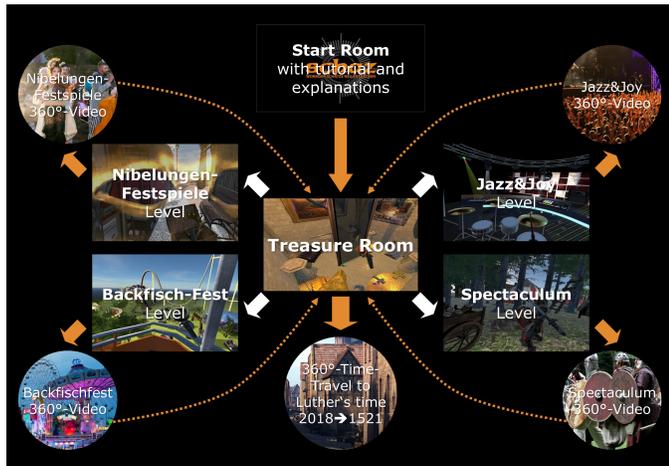


Fig. 2. Overview of structure of example application *schaz-VR*. Boxes represent spatial VR levels, circles represent 360-video levels with interactive spatial VR parts. Arrows indicate possible paths through the levels of the application. Each path starting from the treasure room is dedicated to a single cultural event taking place in the city *Worms*.

spend time in the 360-video level. An example of this is visible in the right image of Figure 1 as “+3 Sek. Bonuszeit” rewarding the player with an extra time of three seconds.

In summary, *schaz-VR* leads players into environments where 360-videos are combined with interactive spatial VR content, and this combination is used to motivate them to observe the 360-videos for a longer time. A believable combination, such that spatial VR contents seem natural in the 360-video and still are recognized as interactive elements, is not trivial to achieve. The existing challenges and possible solutions are discussed in the rest of the paper.

2.3 Evaluation Setting

The discussion in this paper, is based on data obtained by presenting *schaz-VR* during the three days of the RLP-Tag festival. Over 1,100 guests used the application at eight VR booths located at 5 different places spread over *Worms*. Feedback has been obtained from 476 of the guests using a questionnaire and from 1,111 users via log data produced by the application [3]. Of the 476 questionnaires 124 were incomplete, leaving 334 questionnaires contributing to the final evaluation.

3 CHALLENGES: TWO PRACTICAL EXAMPLES

To illustrate the problem we are addressing and to explain the different challenges, we describe two practical examples where problems arise when combining spatial VR content and 360-videos.

3.1 Virtual Floor

In a spatial VR application we usually have the situation, that the player will probably notice a virtual floor, a rendered ground to ‘stand on’. In the simplest case this can only be a cube or a plane surface, but it becomes more realistic if textures are applied to decorate the surface. The virtual floor provokes the feeling that the spatial distance between head and floor is about the distance the player knows from

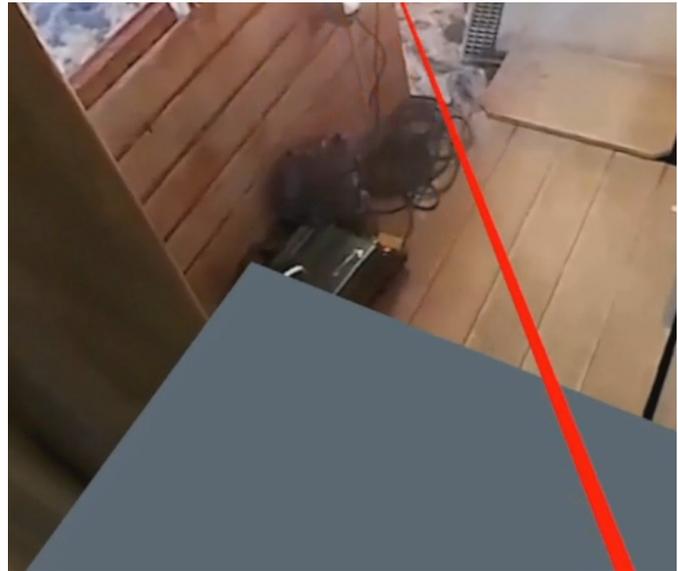


Fig. 3. A computer generated virtual floor (part of rectangle in lower part of image) rendered in 360-video of train waggon. The floor occludes parts of the scene and seems to float high above the floor in the video due motion parallax.

real life. If the visual impressions of the player do not match the real distance, this can lead to VR sickness [12].

When presenting a 360-video using an HMD, the camera/player is usually positioned inside a sphere and the pre-recorded 360-video is rendered onto the inner surface of the sphere. There are also other practicable approaches [1], e.g. cubical or pyramidal projections, but in the case of our *schaz-VR* application, we used the spherical projection. Inside the sphere, the point of view is approximately the center of the sphere. The sphere itself is usually quite large in order to avoid distortions of the 360-video which are often observed when moving (translation) instead of only rotating the head in small 360-video spheres.

During application testing in this setting, some users felt as if they were flying in the middle of the scene. Most people did not mind, but some complained about a feeling of discomfort. In search of a solution, we first tried to render a simple plane as virtual floor beneath the player in the sphere. But that disturbed the feeling of space and realism, because parts of the video, which should normally be visible, were now hidden and invisible to the player. Above all, the player perceives that in reality there could never be a floor, at this point and so the created scene feels wrong.

After that experience, we tried to create a connection between the rendered floor and a floor inside the 360-video, in our case the stage of a concert at *Worms: Jazz&Joy*. We rendered a virtual stage, that exactly fit to the edges of the stage in the 360-video, and thus represented an extension of the stage on which the player can stand. Unfortunately, although the negative feeling regarding flying in the air had been reduced, we had created a new issue, a collision with the sphere and the floor. The issue was not the contact between the floor and the sphere, but the resulting change in the player’s spatial perception. What normally seemed to be huge and wide now appeared much too small because



Fig. 4. Left: Player with chalice in treasure room; portal to spatial VR level of Backfischfest is open. Players can learn to use the chalice in the VR level. Right: In the 360-video, players can collect grapes appearing in front of them using the chalice.

the floor provided depth cues revealing the actual size of the video sphere. Edges of the stage, which are normally long and straight, now appeared to be short and curved in the 360-video. The illusion of being at a big concert could no longer be held up.

3.2 Behavior of Liquids

As second example, we consider the behavior of liquid in our setting combining spatial VR and 360-video. Liquid will always flow towards where the gravity has its centre and will stop if it is blocked by an obstacle. In one of our spatial VR levels, we let the player fill up an empty chalice with wine. The wine itself, flows constantly and endlessly out of a jug. At the moment when the wine touches the ground it is sucked up by the ground’s material. At least this is the impression the player should have. In fact, we destroy those water particles for performance reasons. This fact will become relevant as we discuss the use of liquids in the 360-video setting.

As part of the idea to involve players in the video content, we also implemented liquid in one of our 360-videos. In one video scene, in which visitors enjoy a wine tasting while having a ride in a Ferris wheel, grapes pop up frequently around the player (Figure 4, right image). After successfully catching grapes with a chalice, they start to transform into liquid and run into the chalice. That behavior is exactly what we would describe as a working physical and also spatial relationship between the liquid and the chalice. Nevertheless, it might happen that the player decides to pour out the wine out of the chalice. The result would be, that liquid would fall below the player into the sphere used to present the 360-video, and at the end fall completely out of the sphere or lie at the bottom of the sphere on top of the video. This would be a state that does never occur in real life and thus destroys the illusion of the 360-video.

3.3 Common Observations

By considering the two above examples it is obvious that the special setting of combined spatial VR and spherical 360-video rendering where players are able to look around in 6 DOF needs to adhere to special rules in order to result in presentations which are believable for the players. As the

derivation and description of such rules results in a new model for describing the setting, we first discuss related ways and theories that have been used to combine 3D content with 360-video.

4 RELATED WORK

In the past, many different approaches to create convincing combinations of 3D content with 360-videos have been presented. Unfortunately, most of them deal with settings where users can move their head only in 3 DOF, i. e. the virtual camera can only be rotated. Thus these approaches do not need to deal with problems resulting from translations of the camera in general and motion parallax in particular.

A recent example of the mentioned approaches is the MR360 approach [13] which aims at the seamless integration and compositing of computer generated 3D content into 360-videos. For this purpose lighting information for the computer generated objects is derived from the 360-video. This information is both used to shade these objects and to produce plausible lighting interaction of the 3D object within the 2D scene of the 360-video, e. g. realistic shadows of the 3D object. A seamless integration as achieved by MR360 is not possible in our case because in our case the virtual camera is not held in a fixed location (due to the players’ option to move with 6 DOF) thus revealing the true 3D position of the computer generated objects by motion parallax.

Argyriou et al. [1] also discuss a system integrating 3D objects with 360-video for a 3 DOF setting. Their work, however, is more focused on the interaction in order to engage users. They discuss a number challenges posed by the setting and also present recommendations on how to deal with the challenges. Only one of their recommendations is also relevant for our setting: “[...] the appearance of the 3D objects incorporated in it [the video] should be as realistic as possible presenting a consistent and natural environment”. This recommendation should be considered in addition to the ones we discuss below. We did not adhere to this recommendation because this paper focuses on the challenges of the 6 DOF case.

Furthermore, there are papers providing guidelines for designing interactive 360-video applications. Saarinen et al. [14], for example, provide hints for recording 360-videos (viewpoint, objects close to camera), hints regarding the types of content lending themselves for good presentation in 360-videos, hints regarding prominence and visibility of interactive objects and paths, and recommendations for transitions between 360-videos.

5 THEORETICAL FOUNDATIONS OF THE DISCUSSION

Our discussion of the challenges of combining spatial VR and 360-video is based on two conceptual models. The first model is a typology of relationships between the different elements of the virtual environment (Figure 7), the second model is a division of the sphere presenting the video into three zones (Figure 8). We will introduce the idea of these models in the following subsections.



Fig. 5. Left: A balloon, is being started in the camera zone. It will cross the intermediate zone and end up in the video zone later. Right: Skyrocket explodes in the video zone. Both, the rocket and the balloon do not interfere with the 360-video.



Fig. 6. Left: A flat 3D object (an arrow) in the video zone which is seamlessly integrated into the video. It does not spatially stand out from the video. Right: Smoke particles in the intermediate zone and also in the camera zone that blur the scene.

5.1 Relationships

Scenes combining 360-video with spatial VR usually consist of three different classes of objects:

- the player (can also be seen as the camera),
- the 360-video, which is rendered on the surface of a sphere around the camera or player, and
- the stereoscopically rendered 3D objects inside the sphere.

These objects can be related to each other in different ways. In our model there are three possible relationships between the objects in the scene:

- spatial,
- physical, and
- semantic.

Not all kinds of objects can have all types of relationships to all other kinds of objects. Combining the object types with the types of the relationships, creates the matrix shown in Figure 7. In this matrix, a cross (⊗) represents the fact that a certain kind of relationship is not possible or meaningful and a checkmark (✓) represents the fact that a certain kind of relationship is meaningful and can exist.

5.1.1 Spatial Relationships

Spatial relationships can be described as the positions of all existing 3D objects in the computer-simulated environment relative to each other.

	player - video	player - 3D	video - 3D
spatial	✓	✓	✓
physical	⊗	✓	✓
semantic	⊗	⊗	✓

Fig. 7. Matrix showing possible relationships (rows) between different types of elements (columns) of the scene.

Most of these objects are probably the rendered 3D objects inside the sphere. This kind of 3D objects can exist in infinite quantities and appear anywhere in the sphere at any time.

As already pointed out in above, another object is the player itself. Just like in any other VR application where the player experiences the scene from an ego perspective, one can identify the player with the virtual camera that renders the scene seen by the player. This object exists exactly once in the scene. Although the optimal position of the camera is in the center of the 360-video sphere, players can physically move in real space and thus inside the virtual environment. If they do, the virtual camera moves always with them. This implies two possible relations. First, the camera could interfere with the video, e. g. by clipping parts of the video sphere, or by producing a distortion of the video when the camera gets close to the sphere. This is an unnatural and thus undesirable scenario. Second, players can interact with the game objects around them, which is probably a desired behavior.

Finally, the largest object in the scene is the 360-video sphere. It also exists only one time. In addition to the player, the sphere can also interfere with the other 3D objects. This can be a desired behavior or, on the other hand, lead to disturbances. More details will be provided in the following sections.

5.1.2 Physical Relationships

In principle, physical relationships can exist between all 3D objects. However, the player should not be able to influence the sphere physically (see Figure 7). A desired behavior, however, can be a physical relationship between the player and the game objects. By manipulating and interacting with the respective game objects, physical effects can be part of the game experience, e. g. when throwing objects or pouring liquids. But it is precisely this influence that makes it possible for them to change their position, size or speed and thus spatially interfere with the sphere displaying the video.

5.1.3 Semantic Coherences

Semantic relations only come to bear at one point in the matrix of relationships (see Figure 7). This relation arises at the moment when a rendered 3D object in the scene matches the content of a part of a video or another rendered 3D object. This could be, for example, a 3D sword that matches an anvil in the video or a floor that can be seen as an extension of a stage. The spatial position of the rendered 3D

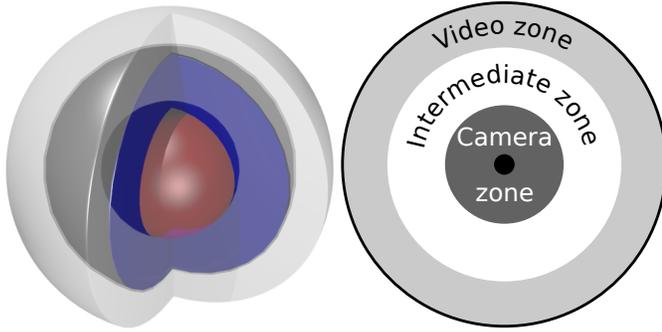


Fig. 8. Different zones divide the 360-video sphere into three parts.

object is irrelevant for the relation. This means that if there is a 3D object at a location where the viewer cannot connect the 3D object to similar content, the relation still exists. In order for a semantic relation to be relevant, however, it is inevitable that spatial proximity *and* similarity is established between the similar contents. It can therefore be stated at this point that we can only speak of a functioning semantic relations if perception can meaningfully link the existing contents with one another. The basics of this theory are known as *Gestalt Laws* [17].

5.2 Zones

In order to model the relative positions of the different 3D objects in the scene, we divided the space within the sphere into three different zones: the *video zone* is located close to the sphere displaying the video, the *camera zone* refers to the area around the player or camera and is typically located close to the center of the sphere, and the *intermediate zone* which is the area not covered by the two other zones (see Figure 8). There is no way to exactly determine where the borders of these zones should be, because they have a merely conceptual nature. Together with the above mentioned relationships between objects in the scene, the zones can be used to describe settings where the combination of spatial VR and 360-video seems to work, i.e. where it produces a believable environment, and why it works in these settings.

5.2.1 Video zone

First and foremost, the video zone is about the relationship between the 3D objects and the rendered video scene and how their relationship is being perceived by the player inside the camera zone. Thus it is critical at which position within the video zone the 3D objects appear. The 3D objects can be very close to the inner surface of the sphere and thus at maximum distance from the player. But they can also get closer to the intermediate zone and thus closer to the player. Since the video zone has a smooth transition to the intermediate zone, no direct boundary can be defined. The size of the zone also depends on the total size of the video sphere. The bigger the video sphere is, the further away the rendered 3D objects will be perceived by the player. In this zone, the player perceives a semantic relationship as proper or improper, depending on how close to or far away from the video sphere we position the 3D object in the foreground of the video scene.

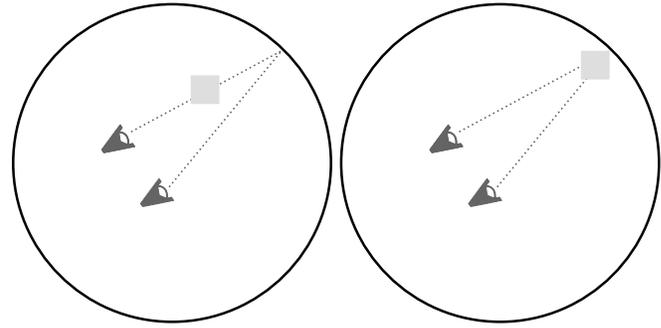


Fig. 9. One reason for problems when combining spherical 360-video and spatial 3D content is the difference in stereo parallax or motion parallax. *Left:* Viewing a cube from different perspectives moves it in front of the sphere while the contents in the 360-video do not change their position to each other. *Right:* The effect is reduced if the 3D object is close to the sphere (video zone) and vanishes if the 3D object is flat and lying on the surface of the sphere.

5.2.2 Intermediate zone

The intermediate zone is the space between the camera zone and the video zone. If rendered 3D objects are positioned in this zone, the players can perceive motion parallax between the 3D object and the video (see Figure 9) when changing their own position and thus get a feeling for spatial relations inside the sphere. If there is no 3D object between the player and the video, it is harder to notice the distance. The more the 3D object shifts towards the video zone, the more the semantic relationship plays a role as it is perceived more strongly. The further it shifts towards the camera zone, the less the semantic relationship plays a role, since the spatial distance is too large (no proximity [17]) for our perception to recognize 3D object and video content as belonging together and thus a semantic relationship as existing or true.

5.2.3 Camera zone

The camera zone is the area, where the player has the ability to move around and manipulate the environment. This zone is suitable for working with physical and spatial relations between the player and the game objects. The player should always stay in this zone to avoid collisions of the camera with the sphere. 3D objects in this zone are located maximally far away from the video sphere. This makes it difficult for the player to create a semantic relation between the 3D object and video content.

6 APPLYING THE THEORY

In order to get a better understanding for the application of the theory, we proceed on the basis of several different examples.

6.1 Forging 3D Metal Bar in Video Fire

First, we present an application case that did not work well.

In this scenario, that can also be seen in accompanying video in the supplemental material, the player was able to pick up a metal bar and hold it “into” a fire to make it glow. The fire was inside the video zone and a part of the video itself. The metal bar was located in the camera zone because the player was holding it and was a 3D rendered, tangible



Fig. 10. Left: Approach to create a semantic relationship between a game object and 360-video content. A 3D iron bar is held towards the fire in the video. Right: 3D chain mounted at ceiling of train wagon crosses all zones without creating issues with motion parallax.

game element (see Figure 10 left). The length of the metal bar was such, that it reached into the intermediate zone when the player held the bar towards the fire. In addition to the fire, a virtual flame was implemented in the intermediate zone. This was intended to intensify the impression of the fire. However, any motion of the player or the metal rod resulted in a strong motion parallax, so the distance between the rod and the fire in the video was interpreted as strange or incorrect by the players.

6.2 Sword in Hand against Shield in Video

An example of a semantic relation appearing very natural is also part of the final *schaz-VR* application. As already mentioned in subsection 2.2, the players could fight with a sword in their hand against a knight in the video. The interaction was to attack the knight's shield with the sword and as a reaction the application rendered virtual splinters flying towards the player. The video accompanying this paper shows this part of the application as well. In this scenario, the player's sword is always inside the camera zone and does not penetrate the intermediate zone or the video zone. The shield is located in the video zone as it is part of the video itself. At the moment of the hit, the splinters are rendered within the video zone for a short moment (see Figure 1 right). The splinters are actually a particle effect, which gives the impression that the splinters are moving apart in all directions.

The semantic relation thus exists between a part of the video and the splinters. Both are located in the video zone. There is also a relation between the sword and the shield, but this is perceived as correct by the player, because the sword is with the player and the shield is with the opponent. Both face each other at some distance. Furthermore, the sword never leaves the camera zone and therefore the player cannot estimate the distance to the video zone.

6.3 Collecting Grapes

In the example from section 3.2 where players collect grapes (see Figure 4 right) and the resulting wine flows into the chalice, there is also a semantic relation. The liquid, in this case the wine, belongs to the chalice. There are also semantic relations between the wine glasses in the video and the chalice in the player's hand. But due to the spatial proximity

of the grapes to the player, the semantic relation between the chalice and the rendered grapes is perceived as stronger than the relation to the contents of the video.

Furthermore, a physical relation can be observed in this example. As already pointed out before (Section 3.2) the wine flows in a physically correct way into the chalice. The physical relation therefore refers to the liquid and the chalice. If, however, the player pours the wine out of the chalice, the physical relation changes and now should exist between the wine and the video sphere. This relation could lead to an issue, since there is neither soil nor any other obstacle that prevents the wine from falling onto the inner surface of the sphere. To avoid this, we have implemented a solution that dissolves the liquid shortly after falling. This behavior can be observed in the supplementary video.

6.4 Creating a Floor in the Camera Zone

In section 3.1, we have mentioned our experiments with virtual floors. We will take a closer look at another example of this experiment here. In the 360-video of the *Nibelungen-Festspiele* we had a scene that plays in a train. The train compartment is narrow and full of the actors' utensils. The feeling of narrowness increased the feeling of depth downwards, i.e. towards the video sphere, so we decided to implement a virtual floor to reduce the feeling of flying in space. The floor was rendered directly below the player and was completely inside the camera zone. However, since no semantic relation to the video can be established within the camera zone, the floor floats like a foreign body in the middle of the room and disturbs the perception of the scenery. In addition, parts of the scene were no longer visible due to masking and this disturbed the player while watching the video. Another big issue here is the physical relation between the ground and the sphere. It can never happen that the ground one is standing on floats over another visible ground without it falling down following gravity. In addition to these issues, the strong parallax has an unfavorable effect and increases the feeling that something is wrong. The motion parallax and the blocking of the 360-video can be observed in the accompanying video and in Figure 3.

6.5 Inflate Balloon and Let it Fly

At the end of the *Backfisch-Fest* video the players could inflate 3D balloons (see Figure 5 left). For this purpose they had a virtual balloon machine at their controller which inflated them. One after the other, the balloons flew into the sky, exploded and let confetti rain from the sky. This worked very well and was great fun for the players.

The balloon was inflated in the camera zone, crossed the intermediate zone and exploded in the video zone. The confetti behaved similar to the splinters in the *Spectaculum* video, so it fell towards the player. Although the balloons crossed the video zone and intermediate zone, the players did not notice any unnatural motion parallax, which is due to the fact that the semantic counterpart (the sky) has no prominent texture or structures. Additionally, the distance of the balloon to the player is quite large when entering the video zone which results in very small translations of the balloon when moving the head. Thus the motion parallax effect is nearly invisibly small. In other words, the player

does not experience any difficulties with the motion parallax. Additionally, due to the round shape of the balloon, there are only few geometric features that could have been supported motion parallax effects.

6.6 A Fog Machine

In the 360-video Jazz&Joy the players had the possibility to create different visual effects with a magic wand. They could shoot rockets (fireworks), draw in the air and create smoke, similar to a fog machine (see Figure 6 right). The case of the rockets is similar to the example with the balloons discussed above. Launched in the camera zone, they fly through the intermediate zone and explode in the video zone. At this point we would like to look at the fog. It is created at the end of the magic wand as a particle effect and spreads quickly in the player's field of view. The semantic counterpart to the fog was the fog machine on stage in the 360-video. When it produced smoke in the video, and the players used their wand, it gave an almost perfect feeling of immersion, as the smoke from the machine and the wand were credibly mixed together. Although there was smoke in the camera zone and in the intermediate zone, the players could not notice any parallax due to the tiny size of the particles. The density of the particles also caused a blurring of the video and the area around the player, so that distances could hardly be estimated anymore.

6.7 A Chain with a Handle

There is a second type of spatial VR and 360-video combination located in the train setting of Nibelungen-Festspiele: To enhance the feeling of being in a train wagon and to address the spatial perception some more, a virtual handle hanging from the roof and extending into the camera zone has been implemented. The handle itself is an interactive game element and can be pulled down with the controller. As positive feedback, horn signal can be heard.

So that the handle does not simply float in the air, and thus is physically behaved correctly, a chain extends from the handle to the inner surface of the 360-video sphere. This results in an object which is located in all three zones simultaneously. Two relations are created in this situation. The first one is a semantic relation between the handle and the chain. This semantic relation prevents the player from trying to create a relation between the handle and the 360-video. Thus, the parallax between the handle and the 360-video no longer plays a role.

Secondly, we have created a semantic relation between the chain and the 360-video, as the chain goes to a point in the 360-video where the ceiling of the wagon can be seen (see Figure 10 right).

The fact that the chain has no gaps between the individual chain links makes it hard for the player to estimate the length of the chain. In addition no strong motion parallax can occur within the chain, because it is a single object, not consisting of loose parts, which must first be combined by perception into a semantic whole. See Figure 10, image right. Furthermore, there is no other 3D object in the sphere that could support the spatial perception. Finally, the shape of the video sphere is similar to that of the wagons ceiling thus rendering the connection between video and chain hanging

from it spatially sensible. Nevertheless, this is not a perfect solution, as players would consider the length of the chain to be too long if they looked at the chain for a long time.

6.8 Icons for Navigation

The Nibelungen-Festspiele performance has been recorded with four cameras at four different locations. The *schaz-VR* application allows to navigate between the recordings of the different cameras. For this purpose we used various icons in the game that were placed in the immediate vicinity of the inside of the video sphere and thus inside the video zone. The icons were very flat and small and fitted into the 360-video accordingly (see Figure 6 left). Due to the small size and flatness of the icons, and due to their spatial proximity to the 360-video no motion parallax is perceivable and thus the fusion of the icons with the video works well. This is an interesting example, that shows that it is not mandatory to create a semantic relation between the 360-video and a 3D object in order to join different contents, but that this can also be achieved via spatial proximity alone.

7 RECOMMENDATIONS

In this section, we provide a number of recommendations for combining spatial VR content with 360-video. The recommendations are derived from the discussion of the different examples in section 6. To distinguish a recommendation from its derivation and justification we set the recommendation in italics.

If one part of a semantic relation is 360-video content, all spatial VR parts of the relation should be within the video zone and in immediate vicinity of its semantic counterparts. Human perception will try to interpret the semantically matching parts as a whole, even if they are in different zones. If the distance between them is too large, the entire entity will be considered untrustworthy, because of the motion parallax between them. This recommendation can be negatively derived from example 'Forging 3D Metal Bar in Video Fire' (Section 6.1). A well working example for a spatial relation and thus a positive derivation of the recommendation is the example 'Icons for Navigation' (Section 6.8). This claim is supported by the fact that users frequently used the option to navigate between camera perspectives (median of 5 and mean of 5.09 camera changes per user, $n = 482$).

A 3D object within the camera zone should not reach into another zone. This recommendation has been derived from example 'Forging 3D Metal Bar in Video Fire' (Section 6.1). When a part of the 3D object comes closer to the 360-video sphere, perception attempts to establish a semantic relationship between the video and the 3D object, but the motion parallax destroys the sense of integrity between the two parts of the semantic relation. An exception or maybe better an example of how to work around this issue can be found in example 'A Chain with a Handle' (Section 6.7). Here we demonstrated how it is possible to implement a credible 3D object, that reaches from the video zone up to the camera zone in a special case.

3D objects with a semantic counterpart in the video zone should not leave the spatial proximity of their video counterpart when being animated. This recommendation has been derived

from example 'Sword in Hand against Shield in Video' (Section 6.2). The splinters moving from the shield towards the camera zone are small and remain in the video zone during the animation. Thus the player has no possibility to detect a motion parallax. This recommendation can also be derived from example 'A Fog machine' (Section 6.6), where we point out the way animated rockets work, and from example 'Inflate Balloon and Let it Fly' (Section 6.5). Both, the balloons and the rockets, explode in the video zone and release particle effects for a short time. This recommendation is supported by the fact that users extensively used the sword, rockets, fog and balloons:

-
- Rockets: median of 167 and mean of 247.44 rockets per user, $n = 501$).
-

There is no data on available on the usage of fog.

Semantic relations between 3D objects in the camera zone can be perceived as working, if suitable counterparts are within the same zone. This will work, even if there are also suitable counterparts in the video zone. The spatial distance of the 3D objects to the player must be close, otherwise the relation to the video could be perceived stronger and the discussed issues of semantic relations between the 360-video and 3D content can arise. This recommendation has been derived from example 'Behavior of Liquids' (Section 3.2) and shows that although there are semantically fitting counterparts in the video zone, the players are not trying to interact with the video, but with the grapes seen in their immediate environment. Since their attention is focused on the relation between the grapes and their chalice, they do not notice any parallax effect between the grapes and the 360-video. This recommendation can also be derived from example 'A Chain with a Handle', where a semantic relation between a chain and a handle has been created. This relation worked well because the relation between the chain and the handle is perceived as stronger than the relation between the handle and the 360-video.

Semantic relations between 3D objects and video content or 3D objects and the player will only work very poorly in the intermediate zone. For game objects in the In the intermediate zone, the distance between the player and the object is too large for an interaction. Additionally, the distance of the object to the video is too large to establish a working semantic relation to the video content without causing issues in perception due to the wrong motion parallax effects. This recommendation has been derived from example 'Forging 3D Metal Bar in Video Fire'. By rendering the fire in the intermediate zone, the player could not create a semantic relation between the iron bar and the fire, or between the fire and the video, without the motion parallax destroying the impression of integrity.

3D objects must behave physically correctly to be considered real, however, physical relations between the 360-video and 3D objects should be avoided to prevent interference, e. g. intersection or collision, with the sphere. This recommendation has been derived from example 'Behavior of Liquids' (Section 3.2) and shows that the liquid after it has fallen out of the chalice falls physically correctly downwards, but this would result in the liquid falling far below the players as the sphere

surface is at a large distance from the player. To prevent this, we implemented an invisible layer in the middle of the sphere that prematurely dissolves the particles. This recommendation can also be derived from example 'Inflate Balloon and let it fly' (Section 6.5). Here, we let the balloon explode before it could interfere with the 360-video.

Covering the 360-video by semantically unrelated 3D objects in the intermediate zone or in the camera zone disturbs the player while watching the video. This recommendation can be derived from example 'Creating a Floor in the Camera Zone' and the similar example described in section 3.1.

3D objects in the intermediate zone and the 360-video can have working semantic relations. If the content of the video has no prominent features facing to the player and thus does not reveal any information about its depth, 3D Objects in the intermediate zone and the 360-video can have a working semantic relation. An open sky with a balloon would be a good example for that. We described this in 'Inflate Balloon and let it fly' (Section 6.5). Another way how semantic relations between the 360-video and the intermediate zone work can be derived from the example 'A Chain with a Handle'. Here we had the situation, that a 3D object passed all three zones as one continuous object, so the player had only few cues possibly revealing the length of the chain. In general, one can state that it is easier to create semantic relations between video and 3D objects when the 3D objects have few geometric features that reveal depth information with respect to its semantic counterpart.

Dense particle effects such as smoke can help merge the video and 3D environment because the environment becomes blurred. This blurry environment makes it harder for the player to obtain precise information about the depth of the space and thus the semantic relation is more likely to be perceived as real. This recommendation can be derived from the example 'A Fog Machine' (Section 6.6).

8 CONCLUSION

In this work we described the difficulties and challenges that arise when joining spherical 2D videos and 3D content. The goal was to derive recommendations based on the experience from an application case where such a combination has been implemented.

To ensure, the reader can get a complete and deep understanding of the recommendations, we first provided information on the scope of the application case by introducing the motivation for the application case and details regarding the implemented *schaz-VR* software. Based on this application we extracted appropriate examples, to go deeper into the subject matter. Subsequently, we described a zone model, in order to explain where and how combining spherical videos and spatial VR work well and where not, and introduced a relation-based notion, which was also a conceptual model of how VR content and spherical videos can be related in a spatial, physical or a semantic way.

By applying the models to the examples of the application case, we were able to derive a number of recommendations, which represent the actual core of our work. The recommendations resulted from the experience during development of the software as well from feedback obtained during its presentation.

We are aware of the fact that we did not create a complete manual with which everyone can create own applications without having issues. But we believe that we have created a theoretical framework which can be used by VR practitioners to assess whether their planned combination of spatial VR and 360-video will be convincing or not.

ACKNOWLEDGMENTS

We would like to thank all members of the *Research Group User Experience & Visualization (UX-Vis)* at Hochschule Worms University of Applied Sciences for their support. This project was partially funded by the *Kultur und Veranstaltungen GmbH* of the city Worms. The authors would like to acknowledge the help of *HTC Europe Co.,Ltd.* who provided the Vive VR headsets. The authors would also like acknowledge the help of *Mediamarkt Worms* who provided the VR-ready graphics workstations.

REFERENCES

- [1] L. Argyriou, D. Economou, V. Bouki, and I. Doumanis. Engaging immersive video consumers: Challenges regarding 360-degree gamified video applications. In *2016 15th International Conference on Ubiquitous Computing and Communications and 2016 International Symposium on Cyberspace and Security (IUCC-CSS)*, pp. 145–152, Dec 2016. doi: 10.1109/IUCC-CSS.2016.028
- [2] J. Drengner. *Events und Marke*, chap. Das Potential von Veranstaltungs-Apps zur Beeinflussung des Brand Meanings von Destinationsmarken am Beispiel des Rheinland-Pfalz-Tages, pp. 159–182. Springer Gabler, 2018. doi: 10.1007/978-3-658-23285-6_7
- [3] M. Eckenfels. Telepräsenz in der virtuellen Realität: Anwendungskonzepte und deren Evaluation am Beispiel von schazVR. Bachelor’s thesis, Hochschule Worms University of Applied Sciences, 2018.
- [4] J. Goldman and J. Falcone. Virtual reality doesn’t mean what you think it means. CBS Interactive Inc., March 2016. <https://www.cnet.com/news/virtual-reality-terminology-vr-vs-ar-vs-360-video/>.
- [5] J. Hakulinen, T. Keskinen, V. Mäkelä, S. Saarinen, and M. Turunen. Omnidirectional video in museums – authentic, immersive and entertaining. In A. D. Cheok, M. Inami, and T. Romão, eds., *Advances in Computer Entertainment Technology*, pp. 567–587. Springer International Publishing, Cham, 2018. doi: 10.1007/978-3-319-76270-8_39
- [6] M. Hosseini and V. Swaminathan. Adaptive 360 VR video streaming: Divide and conquer. In *Multimedia (ISM), 2016 IEEE International Symposium on*, pp. 107–110. IEEE, 2016. doi: 10.1109/ISM.2016.0028
- [7] J. Huang, Z. Chen, D. Ceylan, and H. Jin. 6-dof VR videos with a single 360-camera. In *2017 IEEE Virtual Reality (VR)*, pp. 37–44, March 2017. doi: 10.1109/VR.2017.7892229
- [8] F. Nielsen. High resolution full spherical videos. In *Proceedings. International Conference on Information Technology: Coding and Computing*, pp. 260–267, April 2002. doi: 10.1109/ITCC.2002.1000397
- [9] F. Nielsen. Surround video: a multihead camera approach. *The Visual Computer*, 21(1-2):92–103, 2005. doi: 10.1007/s00371-004-0273-z
- [10] J. Paavilainen, H. Korhonen, H. Saarenpää, and J. Holopainen. Player perception of context information utilization in pervasive mobile games. In *Proceedings of DiGRA 2009: Breaking New Ground - Innovation in Games, Play, Practice and Theory*, pp. 1–8, 2009.
- [11] F. Qian, L. Ji, B. Han, and V. Gopalakrishnan. Optimizing 360 video delivery over cellular networks. In *Proceedings of the 5th Workshop on All Things Cellular: Operations, Applications and Challenges*, pp. 1–6. ACM, 2016. doi: 10.1145/2980055.2980056
- [12] L. Rebenitsch. Managing cybersickness in virtual reality. *XRDS*, 22(1):46–51, 2015. doi: 10.1145/2810054
- [13] T. Rhee, L. Petikam, B. Allen, and A. Chalmers. MR360: Mixed reality rendering for 360° panoramic videos. *IEEE Transactions on Visualization and Computer Graphics*, 23(4):1379–1388, April 2017. doi: 10.1109/TVCG.2017.2657178
- [14] S. Saarinen, V. Mäkelä, P. Kallioniemi, J. Hakulinen, and M. Turunen. Guidelines for designing interactive omnidirectional video applications. In R. Bernhaupt, G. Dalvi, A. Joshi, D. K. Balkrishan, J. O’Neill, and M. Winckler, eds., *Human-Computer Interaction – INTERACT 2017*, pp. 263–272. Springer International Publishing, Cham, 2017. doi: 10.1007/978-3-319-68059-0_17
- [15] J. Steuer. Defining virtual reality: Dimensions determining telepresence. *Journal of Communication*, 42(4):73–93, 1992. doi: 10.1111/j.1460-2466.1992.tb00812.x
- [16] I. P. Tussyadiah, D. Wang, T. H. Jung, and M. tom Dieck. Virtual reality, presence, and attitude change: Empirical evidence from tourism. *Tourism Management*, 66:140 – 154, 2018. doi: 10.1016/j.tourman.2017.12.003
- [17] C. Ware. *Information visualization: perception for design*. Elsevier, 2012.