## **The study of locomotion method for panoramic videos:**

## **system usability and a sense of presence**

A.Paulauskas, T. Blazauskas, A. Skucaite, C.Cenker, M.Vasiljevas, D.Barisas, L. Paulauskas, M.Binkis, T. Valatkevičius

reiktų pažiūrėti šiuos

<https://www.mdpi.com/journal/applsci/special_issues/Innovative_Solutions_Augmented_Virtual_Reality_Applications>

<https://www.mdpi.com/journal/sustainability/special_issues/Educational_Intelligence>

<https://www.mdpi.com/journal/applsci/special_issues/3D_Vision_Virtual_Reality>

**Abstract:**

In this paper we investigate the locomotion method dedicated for virtual reality solutions that allows movement within a scene using prerendered panoramic videos. The aim of the study is to evaluate the usability of proposed locomotion method. In order to compare the proposed approach with regular implementations that use 3D scenes, the sense of presence is investigated, as well. Moreover, the study aims to answer the following questions through experiments, which is the preferable method to make a transition from one panoramic video to another, which navigational object, static lines or dynamic arrows, is preferable and more intuitive to use and which transition method has the least negative physiological consequences and is more pleasant to use. The study shows - while the usability and a sense of presence characteristics of the investigated locomotion are good enough, the regular implementation using 3D models provides better usability and a sense of presence.

**Keywords:** virtual reality, usability, sense of presence,panoramic videos, three-dimensional content.

**1. Introduction.**

[Reiktų dar aiškiau parašyti kam skirtas šitas tyrimas, kas naudotojai, kokią problemą sprendžia.]

The study is intended to evaluate the usability and the sense of presence of a locomotion method that uses panoramic video content for rendering virtual reality.

The effects caused by a sense of presence have been investigated for some time already, however, the emergence of new VR technologies, increased availability of VR devices and their prevalence. The latest technical and interaction advancements within the virtual reality (VR) field have marked a new era, not only for VR but also for VR locomotion. In this era, well-established, prevalent VR locomotion techniques are mostly used as points of comparison for benchmarking new VR locomotion designs [1].

A technique for developing an immersive walk-through system implements a rendering method for superimposing computer graphics onto a panoramic movie in an immersive walk-through system. The system is composed of a locomotion interface and an immersive spherical display [2]. Moreover, navigation is a representative task in a virtual environment (VE). The locomotion methods affect the navigation performance and can cause involuntary movements by the users, which may cause critical safety problems [3]. A defining virtual reality (VR) metric is the sense of presence, a complex, multidimensional psychophysical construct that represents how intense is the sensation of actually being there, inside the virtual environment (VE), forgetting how technology mediates the experience. The paper explores how locomotion influences presence, studying two different ways of artificial movement along with the VE: walking-in-place (through head bobbing detection) and indirect walking (through touchpad) [4].

The systematic review presented by the authors show a wide range of different locomotion techniques and each technique is characterized by different advantages and drawbacks, but classic locomotion techniques such as joystick outperformed all the proposed technique in the reviewed studies. The authors also proposed taxonomy and two types of evaluation for locomotion techniques in a virtual environment [5].

Another Room-Scale locomotion method is one of the most realistic locomotion methods used in virtual reality technologies. This is due to the natural interaction obtained through the tracking of its controllers and the head-mounted display with six degrees of freedom. However, mapping by the position between the physical and the virtual world limits the user's movement to the physical workspace provided by the corresponding device [6]. The locomotion method appeared as a potential solution to the locomotion problem in virtual reality after the emergence and the democratization of the new generation of head-mounted display systems [7].

The authors of another paper examine the effect of the amount of physical space used in the real world on one popular locomotion interface, resetting, when compared to a locomotion interface that requires minimal physical space, walking in place. The metric used to compare the two locomotion interfaces was navigation performance, specifically, the acquisition of survey knowledge [9].

A large number of virtual reality (VR) applications use teleport for locomotion. The non-continuous locomotion of teleport is suited for VR controllers and can minimize simulator sickness, but it can also reduce spatial awareness compared to continuous locomotion [8]. Virtual reality can provide innovative gaming experiences for present and future game players. However, scientific knowledge is still limited about differences between players’ experience in video games played in immersive modalities and games played in non-immersive modalities (i.e., on a desktop display). The authors provide evidence that (a) playing a video game in virtual reality was not more difficult than playing through a desktop display; (b) players showed a more intense emotional response, as assessed by self-report questionnaires and with psycho-physiological indexes (heart rate and skin conductance), after playing in virtual reality versus after playing through the desktop display; (c) the perceived sense of presence was found to be greater in virtual reality as opposed to the non-immersive condition.[10].

Moreover, it is important to choose motion tracking technology for the experiment implementation. Many authors [11, 12, 13, 14, 15, 16] discussed the motion tracking technology however, authors of the paper are focusing on the heads-up displays allowing them to assess the accuracy of the motion tracking they use, the quality of the displayed image (resolution, frame rate, viewing angle, and the number of inputs, the amount of mobility, and the freedom of movement. Other authors mentioned technologies and devices in charge of recording and tracking the user movements were the WiiMote, Wii MotionPlus and Wii Balance Board from Nintendo, Kinect from Microsoft and the PlayStation Move and Eye from Sony. All these devices use different technologies to achieve a similar goal, by means of video cameras, depth sensors, accelerometers, gyroscopes, pressure sensors, etc. Sometimes, games involving motion tracking are called active games, and the fact of playing these is often referred to as exergaming, and to date, several research studies have been conducted to explore the advantages of this practice [17].

Real-time motion tracking is a crucial issue for any AR/VR system, and there are different methods to realize the tracking performance. In marker-based motion tracking, the system needs to detect and identify the marker, and then calculate the relative pose of the observer. However, the marker needs to be stuck on or near the object of interest in advance, and sometimes it is not possible to attach the marker to certain circumstances. In addition, the marker should remain visible during the mobile AR/VR process, and the tracking is inclined to become corrupt due to the marker being out of view. Similarly, the model-based method is another typical motion tracking method for mobile AR/VR. This tracking method uses a prior model of the environment to be tracked. Usually, this prior knowledge consists of 3D models or 2D templates of the real scene. Nevertheless, the extraction of a robust tracked prior model is not always available, especially in some unorganized natural scenes. With the cost of computer vision decreasing rapidly, the visual-based markerless approach turns out to be a more attractive alternative to performing motion tracking [18].

Motion tracking and localization devices are important building blocks of motion tracking systems in a virtual reality (VR) environment [19].

In the process of evaluating the usability of two-dimensional methods for rendering virtual reality content and the sense of presence created by replacing the rendered image with different views.

The familiar 3D objects can be detected faster than their 2D pictorial representations [20]. The affordances might grant real objects higher perceptual priority [21] however, 3D objects are held to more strongly activate affordances than 2D objects [22] and intact objects have more affordances than scrambled ones, i.e. specifically, stable affordances [23], the finding that intact 3D objects, but not scrambled ones, were detected faster implies that affordances could have indeed mediated this effect.

However, panoramic imaging has important implications in robotics, [computer vision](https://www.sciencedirect.com/topics/computer-science/computer-vision) and virtual reality in the design and development of 2D/3D panoramic image capturing systems, the advancement of auto-calibration, registration and corresponding techniques, [stereo vision](https://www.sciencedirect.com/topics/computer-science/stereo-vision), [3D reconstruction](https://www.sciencedirect.com/topics/computer-science/3d-reconstruction) and image-based rendering [24].

Other authors demonstrate a system that acquires high-resolution (>3Kx480) panoramic images. These images are recorded at 30Hz frame rates and played back for later viewing. During the playback, users wear a head-mounted display (HMD) and a head-tracking device that allows them to turn their heads freely to observe the desired portions of the panoramic scene [25]. Creating an accurate and realistic virtual environment is not a task for anyone but rather for experts in 3D design and modelling. To save costs and avoid hiring experts, panorama images are often used to create realistic-looking virtual environments. These images can be captured and provided by non-experts. Panorama images are an alternative to handcrafted 3D models in many cases because they offer immersion and a scene can be captured in great detail with the touch of a button [26].

Another research is related to omni-directional stereo images and videos are popular media formats for displaying content captured from visual sensors through virtual reality headsets. By assigning a different panoramic image for each eye, 3D human vision can be simulated and the viewer is allowed rotational but not positional freedom. Incorporating positional freedom provides a viewer with six degrees of freedom which can be accomplished by using panoramic depth maps [27]. The development of a virtual reality viewer of panoramic images should consider several parameters that define the quality of the rendered image. Such parameters include resolution configurations, texture-to-objects mappings and deciding from different rendering approaches, but to select the optimal value of these parameters, visual quality analysis is required. In this work, we propose a tool integrated within the Unity editor to automate this quality assessment using different settings for the visualization of equirectangular images. We compare the texture mapping of a skybox with a procedural sphere and a cubemap using full-reference objective metrics for image quality analysis [28].

By implementing a resources review we can state that the novelty of the research is related to the usability of two-dimensional methods for rendering virtual reality content and the sense of presence created by replacing the rendered image with different views.

**2. Methods**

***2.1. Methods, Design and Settings***

**2.1.1. Locomotion method for panoramic videos**

The program uses scenes with two-dimensional dynamic content. It allows the user to move freely between the predefined viewing positions of the scriptwriter. The three-dimensional scene is rendered with panoramic stereoscopic videos generated at each preset viewing position. These clips are projected onto three-dimensional spheres in the scene, which envelop the user. As the user moves between viewing positions, he is shown the image from the panoramic video of the corresponding position, thus giving the user the illusion of six degrees of freedom. Since only the sphere is used in two-dimensional content, the detail of the scene models generated in panoramic videos no longer affects the performance of the application. However, the size and quantity of the panoramic videos do affect the performance of the system. This is especially true if the application is implemented on a mobile virtual reality system. Also, the resolution, frame rate and *bitrate* of the videos influence the inclusion of such content. The video parameters listed above have a direct impact on the size of the file, and thus of the application.

The minimum requirements for panoramic video on mobile virtual reality platforms are resolution - 3840x1920 pixels; frame rate - 30 or 60 frames per second; bit rate - 25-60 Mbps; size - up to 10 GB; duration - up to 30 minutes. However, the resolution of the panorama is not the same as the resolution that the user sees in the panorama. Depending on the field of view of the virtual reality helmet, the resolution seen by the user varies. The visible resolution can be calculated from the percentage of 360 degrees that the field of view of the VR helmet is. For example, if the field of view is 120 degrees, the user will only see 33%, so at the minimum recommended resolution for the content, the user will only see an image with a resolution of 1267x633 pixels. In order for the user to see an image with a resolution of 2K, the panoramic resolution needs to be 6K. Higher apparent resolution means greater detail and less blending of the image. However, a higher resolution also requires a higher bit rate for a high-quality image. If the bit rate is too low, artefacts appear in the moving image and degrade the quality of the content.

Two-dimensional content uses more than one video. Therefore, the number of video clips, and therefore the overall size of the program, depends on the nature of the content and the number of viewing positions and the total duration of the content as foreseen by the scriptwriter. The size and preparation time of such content depends on the equipment used to generate the panoramas, the number of panoramas and the parameters (Table 1).

**Table** **1.** Generation time and file size for a single frame at different resolutions in PNG format.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Resolution | 3840x3840 p | 4096x4096 p | 5760x5760 p | 8192x8192 p |
| Generation time | 45 s. | 50 s. | 94 s. | 185 s. |
| Size | 21 MB | 24 MB | 47 MB | 93 MB |

At each viewing position, a full-length video file of the scene is generated. The length depends on the specifics of the content. For content depicting a scene in a dynamic environment, it may have a looping cycle and a lower overall duration, thus saving on file size. If it is a scripted scene that cannot be subject to a repeat cycle, the duration of the clips and the file size will be higher. Before generating the video, a sequence of duration frames of the scene in PNG format is generated. The size of the sequence of frames depends on the frame rate of the video to be generated. The sequence of frames is then used to generate a video in MP4 format at the specified bit rate. This is a time-consuming process.

The size of video files depends on the duration and bit rate and remains the same as resolution and frame rate change. The quality of the content is determined by the resolution, but the computational time increases with it. The higher the resolution, the higher the bit rate it requires to avoid image crumbling in dynamic content.

Based on the known duration of the scene, the planned number of viewing positions, the frame rate and the resolution, it is possible to calculate how long it will take to prepare the content to be displayed. The calculation is made according to formula (1):

; (1)

where *tKSG* is the generation time in seconds for sequences of frames of viewing positions; *nPoz* is the number of viewing positions; *tVid* is the duration of the video in seconds; *kDaz* is the frame rate of the video, and *tKG* is the duration of the generation time in seconds for one frame*.*

For example, for a scene with six viewing positions and a scene duration of 1 minute, the minimum recommended resolution of 3840x3840 would be required to generate a video at 30 frames per second:

6 x 60 x 30 x 45 = 486 000 s = 135 hours = 5,625 days.

And if the frame rate is 60 frames per second, you will need:

6 x 60 x 60 x 45 = 972 000 s = 270 hours = 11.25 days.

Only the first part of content preparation - the generation of frame sequences - has a high time cost. The time needed to generate frame sequences for minimum (3840x3840) and recommended (8192x8192) resolutions differs by a factor of almost 5 (Table 2).

**Table 2.** Generation times for one-minute, 30 fps, six-viewpoint panorama sequences.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Frame resolution | 3840x3840 | 4096x4096 | 5760x5760 | 8192x8192 |
| Generation time | 5,625 days | 6,25 days | 11,75 days | 23,125 days |

Based on the bitrate used to generate the videos and the length of the video, it is possible to calculate how much space the files will take up. The calculations are based on formula (2):

; (2)

where *dVid* is the total file size of all video files with viewing positions; *nPoz* is the number of viewing positions; *kP* is the bandwidth; *tVid* is the video duration in seconds.

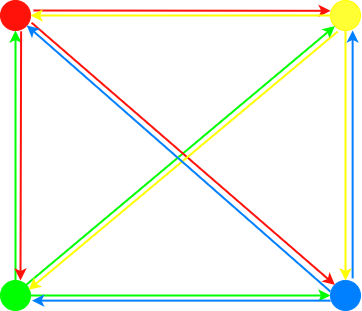
For example, if bandwidths of 25, 60 and 100 Mbps are used for generation, the total size of the video clips calculated according to formula (2) will be:

6 x 25 x 60/8 = 1125 Mb = 1.125 Gb for a bit rate of 25 Mb/s

6 x 60 x 60/8 = 2700 Mb = 2.7 Gb for a bit rate of 60 Mb/s

6 x 100 x 60/8 = 4500 Mb = 4.5 Gb for a bit rate of 100 Mb/s

When transforming between viewing positions using clipping or merging methods, only the videos in the viewing positions are generated. However, when transforming by the video method, it is additionally necessary to generate video transforms of the specified duration *t* between all viewing positions, at *t* time points and in all directions. High bit throughput is particularly important in highly dynamic content, where not only the objects around the camera move but also the camera itself. Therefore, a video of transformations must be generated with a higher bit rate than a video of viewing positions (fig. 1).



**Figure 1.** Example of the layout of the four viewing positions, with the directions of transformation from each position illustrated

The maximum number of changes of video positions if each position is accessible to all the others, can be calculated according to formula (3):

; (3)

where *nMaxVid* is the maximum number of video position changes; *nPos* is the number of viewing positions.

For example, if a scene lasts one minute, there are 6 viewing positions between which you can move freely in all directions, and the transformation lasts one second, you will need to generate a total of 1800 one-second videos. The number and size of the transformation videos depend on the number of transformation directions, the duration of the transformations and the total duration of the scene. For the example case, the total file size of the transformations is:

1800 \* 25 / 8 = 5625 Mb = 5.625 Gb for a bandwidth of 25 Mbps;

1800 \* 60 / 8 = 13500 Mb = 13.5 Gb for a bandwidth of 60 Mbps;

1800 \* 100 / 8 = 22500 Mb = 22.5 Gb at 100 Mbps.

This large amount of transformation videos increases the overall size of the programm significantly. In order not to overload the system, the number of preview positions, the distances between them and the number of positions each of them can reach must be optimised. Optimisation can be achieved by reducing the number of transformations or the bandwidth. If we reduce the number of transformations, we reduce the possibilities of movement, so that engagement suffers, but not image quality. If we want to keep a wide range of movement, it is worth reducing the bit rate, but this can lead to artefacts in the transformations and a drop in image quality, and thus in inclusiveness.

**Methods of image modification**

In the study, two-dimensional scenes update the image shown to the user as they move between viewing positions. In order to simulate six degrees of the freedom movement, the change of the image should look as natural as possible and cause as little discomfort to the user as possible. Also, when the image is changed, it continues from the same point in time, the image is not jerky during the change, and a seamless representation of the scene is maintained. The application implements three methods of changing the image, simulating different naturalness of the transformation: clipping, blending and sequential, using the video clip.

**Changing videos by cropping**

Replacement by the principle of deforestation is a common and frequently encountered practice. When approaching the next viewing position, it changes the displayed panorama to another one. This method of changing the view ensures a smooth refresh of the image, continuing from the point of crossing.

When the user reaches a viewing threshold, the time point of the video is captured and the image is paused. The player is then given a new video to play, decodes it, loads it into the time-lapse of the previous video captured at the time of the crossing, and plays it. When a new video is specified, the last frame of the previously displayed video is displayed during loading. This results in decoding and loading time glitches during each transformation, which spoils the integrity of the scene. This can be solved by using more than one player (fig. 2).

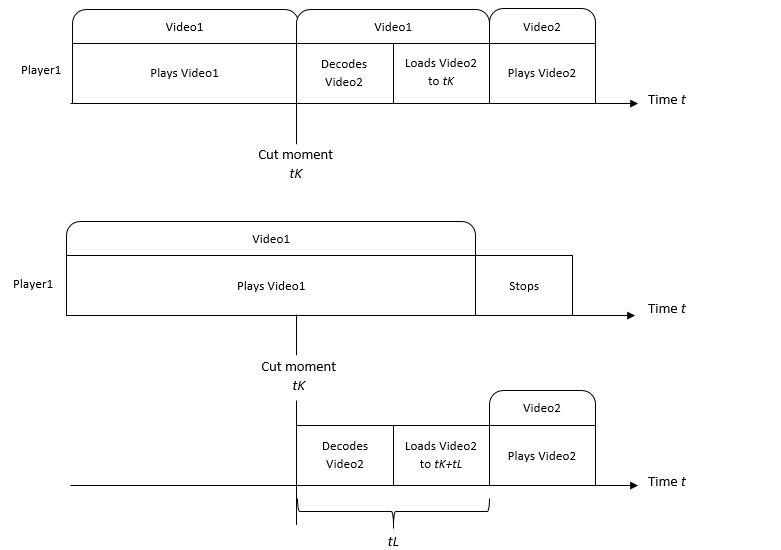


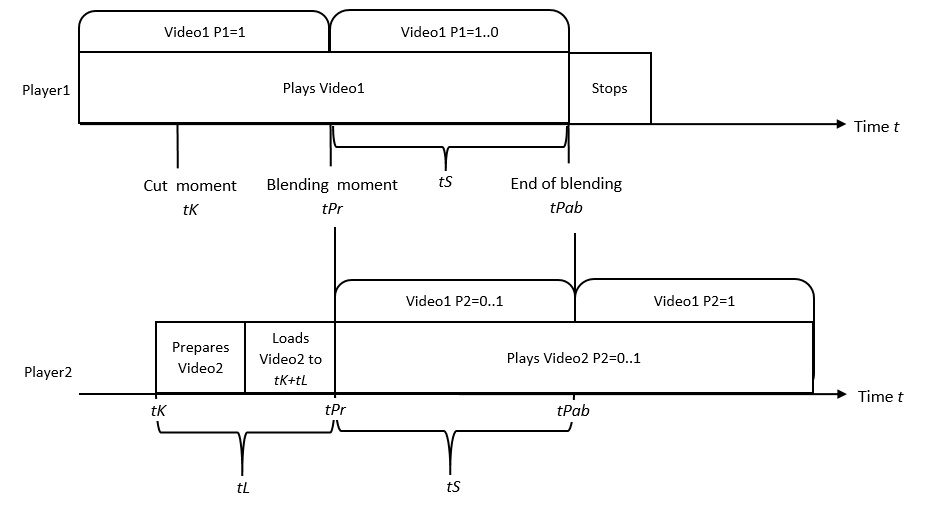
Figure. 2. Results in decoding and loading time.

When using a second player, the first player does not stop playing when you move to a different viewing position, but at the same time, a new video is being prepared in the second player. The preparation and loading take *tL* of time, so it is necessary to allow for the full preparation time and to load the new video in the *tK+tL* time instant, and only then start playing and render the second player with the new video. The decoding and loading time of the video depends on the processor used for the computation of the virtual reality system. It may also vary minimally. This time can be set to a fixed time, after observations of the application's performance on the specific virtual reality system. However, this value will be approximate and will need to be reset when using a different system or a different computer for the computation of the renderer. Alternatively, each time a transformation *tL* is performed during the lifetime of the application, it can be set to the average of the two preceding and measured transformations *tL*. The automatic calculation of *tL by* the software results in a greater overlap of the frames of the transformations, making the transformations smoother.

Figure STYLEREF 1 \s . SEQ Pav. \\* ARABIC \s 1 4. The process of changing the video to be rendered by clipping using one player (top) and two players (bottom)

**Normal blending**

Image fusion-based image replacement is common and often found in virtual environments. As a new viewing position is approached, *tS* changes the pixel transparency values of the displayed and the new image within a predetermined time. This method uses two players and plays the video from both players during the resizing process (see Figure 3.5). The fusion method is smoother than the clipping method and potentially less disorienting, as the user sees the overlapping pixels during the fusion of the images, and thus the change in their position before this happens. However, this method requires a higher performance of the computing equipment, since two players re-running during the replacement (fig. 3).



**Figure 3.** The process of altering the video image using the fusion method.

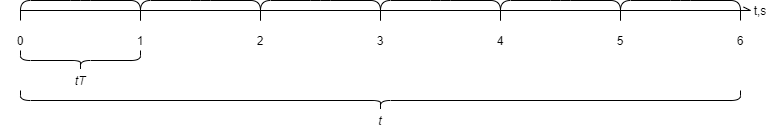
The very short duration of this transformation means that fusion occurs very quickly. In this case, it can go unnoticed by the user and produce the same effect as the cut-off method. However, two players playing at the same time may affect the performance of the system.

**Sequential transformation using video**

The new transformation proposed in this work is performed by playing a pre-generated video of a certain duration, in which the camera changes position between the same viewing positions that the user is moving between at that moment. During the transformation, the video simulates the sequential movement of the user in space.

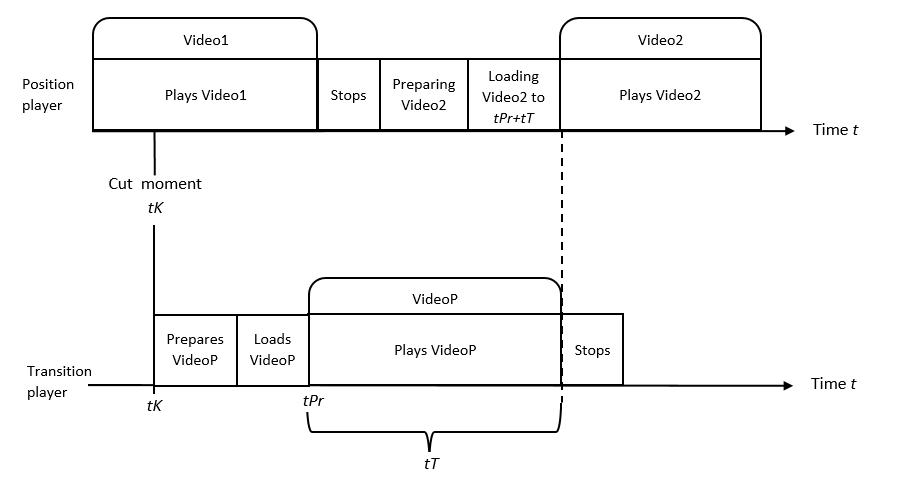
The scene duration *t* is divided into segments of transformation duration *tT* that follow each other (fig. 4). At each transition between viewing positions in both directions, *t/tT* transition videos are generated, each occurring at time *k\*tT*. As a very large number of transformed videos can be generated, the overall size of the application increases significantly, which can create difficulties when using limited-memory mobile virtual reality renderers.

**Figure STYLEREF 1 \s . SEQ Pav. \\* ARABIC \s 1 6:** Segmentation of the video duration used during the change of viewing position

When simulating user movement with a viral video, it is important to take into account the durations of the generated transformations as well as the distances between viewing positions. As the user moves through space, the time between positions should be matched as closely as possible to the duration of the transformation. It is also important that the distance covered during the transformation is the same as in virtual reality and that vestibular dissonance is minimised (fig. 4).

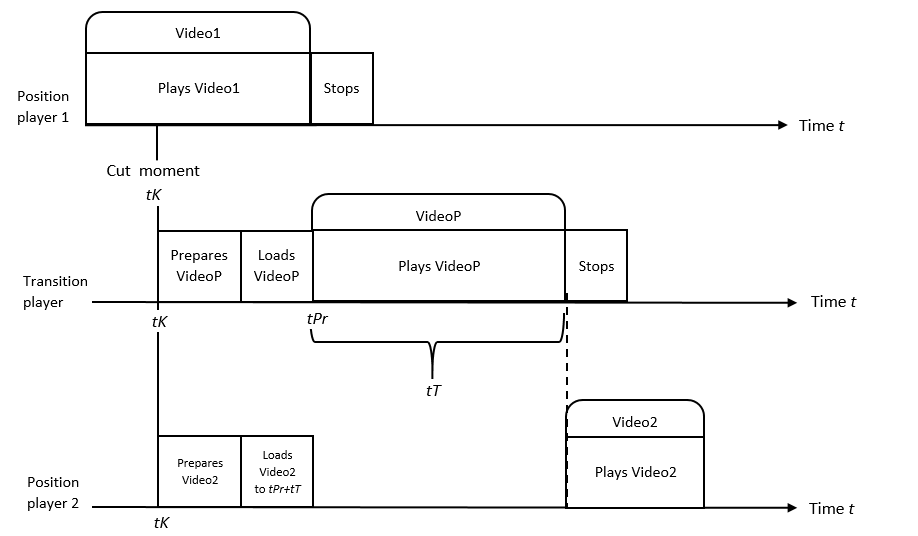
**Figure 4**. Distance covered during the transformation.

Two players are used during the transformation. One plays the video of the viewing positions and the other plays the recordings of the transformations. When a transformation is captured, it is rounded up to the start of the next transformation and the transformation is loaded in the player dedicated to the transformation. As the surface plays, the next video at the next viewing position is loaded in the next player, showing the continued image at the exact moment after the end of the surface. It is important that the decoding and loading of the video take less time than the transformed video. Otherwise, the new position video will not be able to be loaded during the transformation, and the last frame of the previous position may be returned after the transformation is finished. This jumping of the video between positions reduces the inclusion and integrity of the displayed video (fig. 5).



**Figure 5. The** process of changing the display of a video using two video player components.

The transformation can also be realised using three players. In this case, not only is the video of the transformation loaded when crossing the video change threshold but also the video of the next viewing position is loaded simultaneously on a separate player. The advantage of this implementation is that the loading of the second position video becomes independent of the duration of the transformation video. No matter how small the video is, the second video will always have time to load. However, it is necessary to have another video player, which means additional resource usage. This implementation has been explored, but as the number of players has been shown to significantly reduce the stability of the system, this approach has been abandoned for experimentation (fig. 6).



**Figure 6.** Theprocess of changing the display of a video using the three video player components

This work proposes a novel method for detailed three-dimensional content rendering, allowing movement between spatially positioned viewpoints that render stereoscopic panoramic videos generated in a three-dimensional scene. Changing the viewing position during the proposed transformation displays a transformation video, generated in advance in the three-dimensional scene and selected from a list, simulating a sequential and immediate movement in three-dimensional space.

It is hypothesized that this newly proposed method of rendering three-dimensional dynamic scenes allows the creation of high-fidelity dynamic virtual reality content suitable for use on simple devices with limited capacity, over and above three-dimensional content of the same level of detail, without compromising the sense of presence the content creates in the environment. The proposed content maintains a greater sense of presence in the environment than conventional methods of transformation between panoramic videos.

To confirm or refute the hypothesis, experimental studies is carried out allowing a group of subjects to test and then evaluate the proposed method of rendering virtual reality content with conventional methods, comparing the usability of the content, the sense of presence, the adverse physiological effects and the impact on the performance of the rendering equipment. For this purpose, software implementing conventional methods for rendering and navigating three-dimensional and two-dimensional content, as well as a new proposed method, are used.

In this work, experiments are carried out to compare three-dimensional and two-dimensional virtual reality methods for content representation. The aim of the experiments is to evaluate whether the proposed two-dimensional content rendering method is the most pleasant to use, also, costs less computational resources for the system and can compete with three-dimensional content rendering. Therefore, the experiments in the study evaluate the usability, performance and sense of the presence of each implemented two-dimensional virtual reality content rendering method. At the same time, these properties are also evaluated for the three-dimensional content. During the experiments, the system collects performance data and, after performing certain tasks in the virtual environment, subjective surveys are completed to assess the usability of the rendering method and the user's sense of presence in the environment.

The following experiments are carried out:

1. System usability and sense of presence experiment - subjectively assessing the ease of use of the methods, the symptoms of cyber-sickness induced by the methods and the sense of presence induced by the methods in the virtual space.
2. Stereoscopic Imaging Subjective Acceptance Tendency Study - subjectively assessing the acceptability tendency of using stereoscopic imaging over monoscopic imaging to create the illusion of a three-dimensional image, in a sample of subjects taking part in an experiment.
3. Performance experiment - objectively assessing the impact of different levels of detail in three-dimensional and two-dimensional virtual reality content on the frame rate and stability of the rendered frames.

***2.3. Ethical Considerations***

No personal data was collected. When providing cases all details are anonymized even though their solutions have been publicly presented a number of times and there are various publicly available presentations.

***2.4. Data Collection***

The research data was collected during the research implementation process and experiment by the researchers, whose reflections and the account of the process is also an important source of data used in the paper. During the implementation, researchers administered questionnaires. Other relevant details are needed to adjust the working progress if needed. All the collected data was anonymized by one of the researchers by applying statistical analysis methods.

***2.5. Instrument Used***

## The theoretical framework was built related to the systems usability and sense. Moreover, detailed three-dimensional content was discussed in rendering, allowing movement between spatially positioned viewpoints that render stereoscopic panoramic videos generated in a three-dimensional scene. Changing the viewing position during the proposed transformation displays a transformation video, generated in advance in the three-dimensional scene and selected from a list, simulating a sequential and immediate movement in three-dimensional space. The experiment was implemented to get the conclusions of the System usability and sense of presence study

***2.7. Data Analysis***

The data used were anonymized and analysed to make a design in the evaluation processes. The framework of the research was based on two main focuses, namely, two-dimensional and three-dimensional images

**3. Experiment results**

The aim of the study is to evaluate the usability of two-dimensional methods for rendering virtual reality content and the sense of presence created by replacing the rendered image with different views. The study aims to answer the following questions through experiments:

1. Which is the easiest method to use for changing two-dimensional content images?
2. Which navigational object, static lines or dynamic arrows, is preferable and more intuitive and easier to use?
3. Which method of changing the two-dimensional content of an image has the least negative physiological consequences and is more comfortable to use?
4. Does the immediate two-dimensional content replacement method using video create a stronger sense of presence in the environment than other two-dimensional content representation methods?

|  |
| --- |
| Figure 7. Location of viewing positions at shorter (0.52m) and longer (1m) distances |

The content presented to the user in the trial is a virtual reality three-dimensional scene with a single moving object (Fig. 8). In the two-dimensional version of the content, the scene has six and twelve viewing positions. From each position, the user can move to two adjacent positions. In these positions, stereoscopic top-bottom panoramic videos are generated in the three-dimensional scene, with the right eye at the top and the left eye at the bottom. The distance between the left and right cameras is specified at the time of generation as 0,65 mm. The distances between the viewing positions are constant and the viewing positions are arranged in a hexagonal and a dihedral pattern. In a scene with 6 viewing positions, the distances between them are 1 m and in a scene with 12 viewing positions, the distances between them are 0,52 m (Fig. 7). Panoramic videos of viewing positions meet the minimum requirements discussed. The videos are generated at a resolution of 3840x3840 pixels at a frame rate of 30 frames per second. The duration of the videos generated is 30 seconds at all viewing positions. The animation in the scene is 30 seconds long. The animation is in the same position in the first and last frames. Therefore, the videos can be repeated without causing interference. The videos are generated at 12-bit bandwidth. The videos used for the transitions between viewing positions are generated by moving the camera from one position to another at a constant rate, every second, throughout the animation (720 videos in total for a 12-position configuration, 360 videos for a 6-position configuration).



|  |
| --- |
| Figure 8.The scene used in the virtual reality experiments. |

The duration of one transformation is 1 second. The parameters used in the configuration file during the rendering of two-dimensional content are shown in Table 3. Their values were determined during testing in order to optimise the experience of moving in space, reduce the time between movement and application response, and make the change of viewing position as natural as possible.

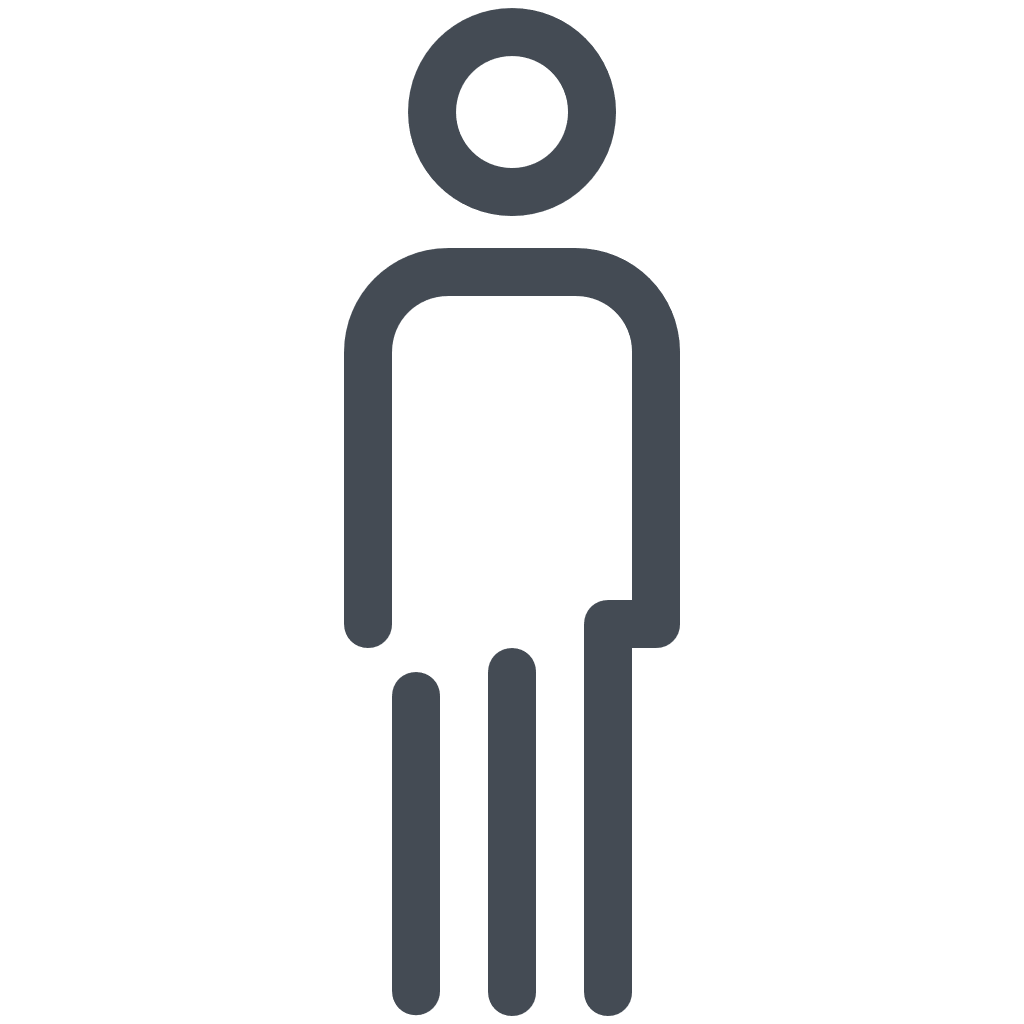
Table 3. Parameters used in two-dimensional content related to image modification.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | For the cut-off method | | For the splicing method | | For the sequential approach | |
| 6 pos. | 12 pos. | 6 pos. | 12 pos. | 6 pos. | 12 pos. |
| Image change threshold distance multiplier | 0,5 | 0,7 | 0,5 | 0,7 | 0,3 | 0,3 |
| Road width (m) | 0,6 | 0,6 | 0,6 | 0,6 | 0,4 | 0,4 |
| Splicing time(s) |  | | 0,2 | |  | |

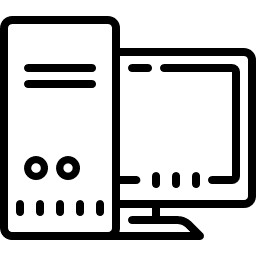
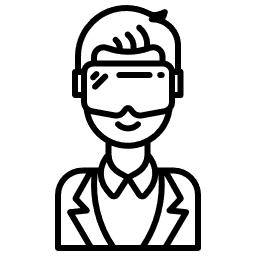
The app renders virtual reality content using the Unity game engine. The total size of the application is 4,36 Gb, of which 2,81 Gb are videos. It requires a computer and a virtual reality helmet, or a virtual reality helmet with an integrated processor for computing. The application runs on a personal computer with an Intel Core i7-8700K 3,7 GHz CPU, 8 Gb of accelerated memory and an NVIDIA GeForce GTX 1070 video processor. A Dell Vizor Windows Mixed Reality virtual reality helmet is connected to the computer, which sends information about changes in the user's position and displays the information received. Control of the application, by changing scenes or objects, is performed using a keyboard connected to a PC. The viewing positions are arranged in 2a space of 2 m. A 2,5 m2 area has been prepared for free navigation, taking into account the user's possible deviation from the intended path (Fig. 9).

2 m.

2 m.



Observer



Participant

|  |
| --- |
| Fig 9. Space, observer and participant positions used in the experiment |

The experiment involves presenting a dynamic scene to participants using a virtual reality headset. The scene is rendered using three-dimensional and two-dimensional methods. If the scene is rendered in two dimensions, one of the three methods for changing the viewing positions is used. The experiments is performed in quadruplicate, with at least a 10-minute break between experiments. Each time, data is collected for a specific rendering method:

1. Normal two-dimensional, by changing the viewing positions by cropping;
2. Normal two-dimensional by changing the viewing positions by blending;
3. The proposed new two-dimensional, changing viewing positions sequentially, using video clips;
4. Ordinary three-dimensional.

Due to the pandemic, additional conditions were imposed on the experiment. The experiment was carried out in a closed room with ventilation throughout the experiment. Two people are present in the room at the same time: a participant and an observer who monitors the experiment. After each experiment, the empty room is ventilated for an additional 10 minutes and all surfaces, including the virtual reality helmet and keyboard, are disinfected. Participants are allowed to enter the room wearing masks only and disinfecting their hands with hand sanitisers on entering and leaving the room.

As the experiment assesses symptoms of cyber-sickness, participants are asked about their current state of well-being before the experiment. Poor well-being before the experiment can lead to even more severe ailments during the experiment. However, all participants responded that they did not feel any health problems or ailments before the experiment. Participants are warned that they may stop the experiment at any time if they experience very strong symptoms of cyber-sickness.

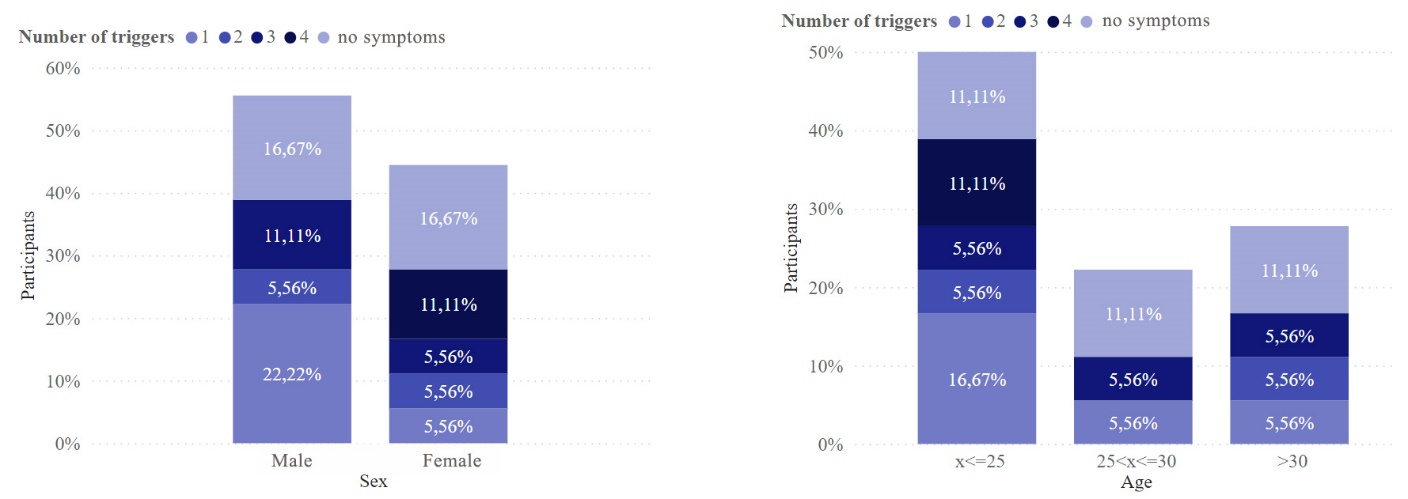
Experiment protocol:

1. The experiment is carried out with a 2 free movement area of 2.5 m;
2. The computer to which the virtual reality device is connected is placed at the centre of one side of the free space where the experiment will take place;
3. The virtual reality device is plugged into the computer, calibrated, and the device's usage area is drawn to cover the entire area mentioned in point 1;
4. The participant is introduced to the task to be performed in the virtual space;
5. The participant takes a position in front of the computer, facing it, 0.5 m away;
6. With the help of an observer, a virtual reality headset is placed on the participant's head;
7. The participant adjusts the position of the viewfinder to see a sharp image in the intended Windows Mixed Reality environment;
8. The observer runs the software developed for the experiments on a computer;
9. The observer triggers a two-dimensional, stereoscopic representation of the content for the participant, in which the viewing position is changed by cropping;
10. The participant is instructed to turn around on the spot and familiarise himself with the surroundings;
11. Enables a configuration with 6 viewing positions arranged in a circle around a moving object in the scene every 1 m;
12. The participant is activated to view the road navigation object;
13. The participant is instructed to use a navigation object to circle around the moving object, changing the viewing position 6 times (from 1st to 2nd, from 2nd to 3rd, ..., from 6th to 1st);
14. The participant is activated to see the arrows on the navigation object;
15. The participant is instructed to use a navigation object to circle around the moving object, changing the viewing position 6 times, in the opposite direction to the previous one (from 1st to 6th, from 6th to 5th, ..., from 2nd to 1st);
16. Enables a configuration with 12 viewing positions arranged in a circle around a moving object in the scene every 0.52 m;
17. When the road navigation object is activated, the participant is instructed to use the object to circle around the moving object, changing the viewing position 12 times (from 1st to 2nd, from 2nd to 3rd, ..., from 12th to 1st);
18. When the arrow navigation object is activated, the participant is instructed to use the object to circle around the moving object, changing the viewing position 12 times, in the opposite direction to the previous one (from 1st to 12th, from 12th to 11th, ..., from 2nd to 1st);
19. The observer stops the application on the computer and helps the participant take off the VR headset;
20. The participant completes the SSQ, IPQ and SUS surveys on the Google Forms, using the identification code provided, to assess his/her experience;
21. After at least 10 minutes, points 5 to 8 are performed;
22. The observer triggers a two-dimensional, stereoscopic representation of the content for the participant, in which the viewing position is changed by blending;
23. Points 10-21;
24. The observer triggers a two-dimensional, stereoscopic display of the content for the participant, in which the viewing position is changed by the proposed video-based method;
25. Points 10-21;
26. The observer activates the participant's usual three-dimensional dynamic content with low detail;
27. The participant walks twice in different directions around a dynamic object on stage;
28. Point 20 is done.

Totally 18 responders took part in the experiment: 8 women and 10 men. The age of the participants varied from 20 to 48 years old. Out of the 18 people, one English-speaking person participated in the experiment. Most of the participants can be classified in the age group under 25 years old - 50%, 22.22% in the age group between 25 and 30 years old, 27.78% in the age group over 30 years old. It should be noted that during the experiment with two-dimensional content, where the image is replaced by video, one participant stopped the experiment as soon as it started due to a very sudden and severe onset of cyber-sickness symptoms. As this participant did not complete the experiment, the evaluation of his video approach was not included in the overall evaluation of the results.

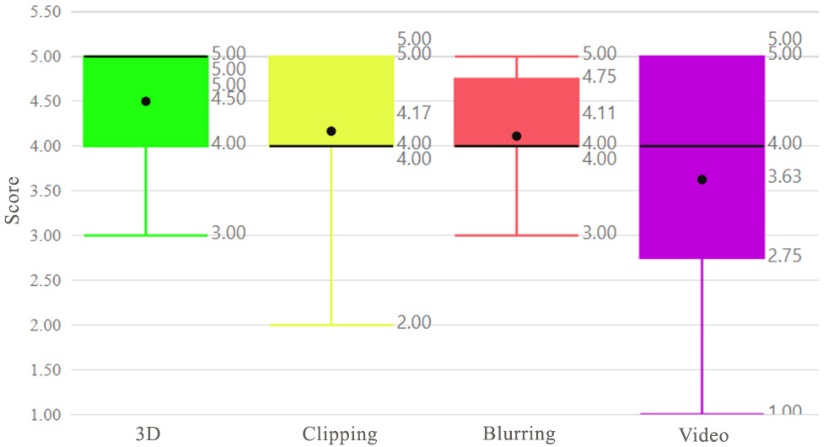
|  |  |  |
| --- | --- | --- |
| |  |  | | --- | --- | | Figure 10. VR usage by gender | Figure 11. VR usage by age | |

The baseline survey on VR usage habits showed that 44.44% of respondents use VR occasionally, 33.33% almost never, 5.56% (one respondent) never and 16.67% often. As many as 30% of the male respondents reported using virtual reality frequently and only 30% of the male respondents reported using virtual reality almost never. In contrast, as many as 50% of the women surveyed said they hardly ever or never use virtual reality and none of the women surveyed use virtual reality often (Fig. 10). Looking at VR usage habits by age of users, the most frequent users are under 30 years old and the least frequent users are over 30 years old (Fig. 11). The data show that men and people under 30 are more likely to use virtual reality.



|  |  |
| --- | --- |
| Figure 12. Distribution of susceptibility to cyber-diseases by gender and number of pathogens | Figure 13. Distribution of susceptibility to cyber-disease by age and number of pathogens |

To assess the susceptibility of the participants to motion sickness, they were asked whether they experienced symptoms when using one or more of the triggers (car, bus, train, boat, carousel). There is a tendency that the more triggers a participant uses, the more susceptible he/she is to motion sickness and thus to cybersickness. 33.33% of the participants answered that they do not experience any symptoms when using the triggers normally. 27.78% of the participants started to feel symptoms with one of the triggers, 11.11% with two triggers, 16.67% with three triggers and 11.11% with four triggers. Women are the most susceptible to motion sickness (Fig. 12). Although the same number of men and women reported not experiencing symptoms of motion sickness, as many as 22.22% of men reported experiencing symptoms from only 1 agent, compared to 5.56% of women. Among the participants in the experiment, 16.67% were men who reported feeling symptoms from 2-3 triggers and 22.23% were women who reported feeling symptoms from 2-4 triggers. It can also be noted that all the most sensitive (i.e. experiencing symptoms from 4 triggers) participants to motion-induced nausea are younger than 25 years of age (Fig. 13).



|  |
| --- |
| Figure 14: Distribution of direct ratings of the virtual reality experience with different content rendering methods (1-very bad, 5-very good). |

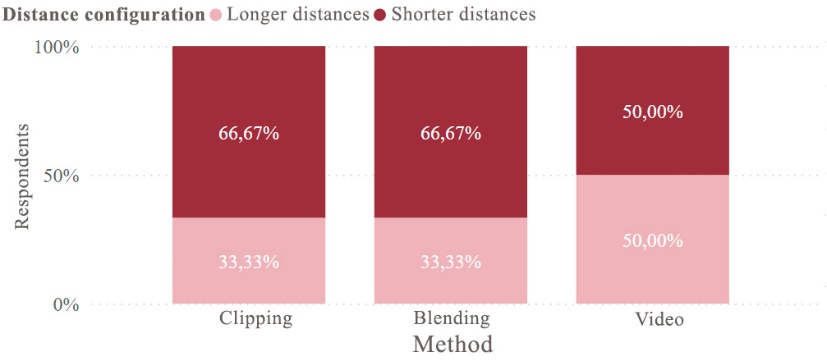
Each time a task was performed in virtual reality with a particular content rendering method, users rated how much they enjoyed the experience. From the distribution of ratings in the graph (Fig. 14), no two-dimensional method was rated better than a three-dimensional method, although the latter had a much lower level of detail. This may have been influenced by the unrestricted movement in 6 degrees of freedom. Among the two-dimensional methods of content representation, the clipping and the blending methods scored similarly. However, although a higher proportion of respondents gave the bevel method the highest rating (5) than the fusion method, the lowest rating for the bevel method is lower than the lowest rating for the fusion method. The fusion method was the most consistently rated method and the method using transformation videos was the least consistently rated method. The majority of the scores for this method range between 2.75 and 5.

When looking at the mean estimates for the different groups of participants in the experiment, there was a small correlation between participants' gender, age, susceptibility to motion sickness and the rating (see 4.2Table). On average, women scored better on all methods than men. Women also rated the decapitation method better than the fusion method. In contrast, men rated the two methods equally. This may be because in this sample of respondents, women have a higher susceptibility to cybersickness than men. Both men and women on average rated the method using video overlay the lowest. Looking at the estimates by age of the respondents, younger participants in the experiment rated the clipping method the best of the two-dimensional methods, while respondents aged 25-30 rated all the two-dimensional methods of displaying content very similarly, with the clipping method and the blending method scoring slightly better. However, respondents over 30 years of age have a higher preference for the blending method than for the cropping method. No bias was observed when examining the image estimates by respondents' susceptibility to cyber-sickness, and a larger amount of data is needed to assess possible correlations.

Table 4. Means of the direct estimates of the methods for different categories of respondents

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Category/method | 3D | Cutting off | Linking | Video |
| Women | 4,75 | 4,25 | 4,13 | 4,00 |
| Men | 4,25 | 4,10 | 4,10 | 3,33 |
| <=25 m. | 4,63 | 4,22 | 4,00 | 3,22 |
| 25< and <=30 m. | 4,00 | 3,75 | 3,75 | 3,50 |
| >30 | 4,75 | 4,40 | 4,60 | 5,00 |
| 1 causative agent | 4,50 | 4,60 | 4,60 | 2,40 |
| 2 causative agents | 4,00 | 2,50 | 3,50 | 4,00 |
| 3 causative agents | 4,33 | 3,67 | 4,00 | 3,67 |
| 4 causative agents | 5,00 | 5,00 | 3,50 | 4,00 |
| Symptoms do not appear | 4,50 | 4,33 | 4,17 | 4,60 |
| Overall | 4,50 | 4,17 | 4,11 | 3,63 |

For two-dimensional content, the experiment evaluated the distance between viewing positions for each method of changing the view. The graph (Fig. 15) shows a trend that the majority of respondents (66.67%) preferred smaller distances between viewing positions in the clipping and blending methods. However, for a consistent change of viewing positions using video, the distribution of distance preference is uniform.



|  |
| --- |
| Figure 15. Distribution of distance choices for each two-dimensional content method |

From the distance ratings for the two-dimensional content mapping methods presented in Table 5, can be seen that short distances are more favourably rated than long distances in the truncation and blending methods, with the smallest differences in their ratings. However, in the video-based transformations, there is only a slight difference between the mean estimates of short and long distances, but the standard deviation shows that these estimates are the most different. The truncation and fusion methods give priority to shorter distances between image changes, possibly because in this case the user perceives a discrepancy between his/her movement and the visible still image for a shorter time. In the video, the priority for distances is low, possibly because the motion is not only performed but also seen when changing position.

Table 5. Estimates and standard deviations for long and short-distance estimators for two-dimensional content image conversion methods.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Method | Average for long distances | Standing deviation for long distances | Average for short distances | Short-distance stand. deviation |
| Cutting off | 3,61 | 0,83 | 4,00 | 1,05 |
| Linking | 3,44 | 0,83 | 4,11 | 0,94 |
| Video | 3,25 | 1,20 | 3,38 | 1,27 |

The distribution of the distance estimates used by this method over a larger amplitude is likely to depend on the individual respondent's movement speed and its correspondence with the movement speed of the camera in the video.

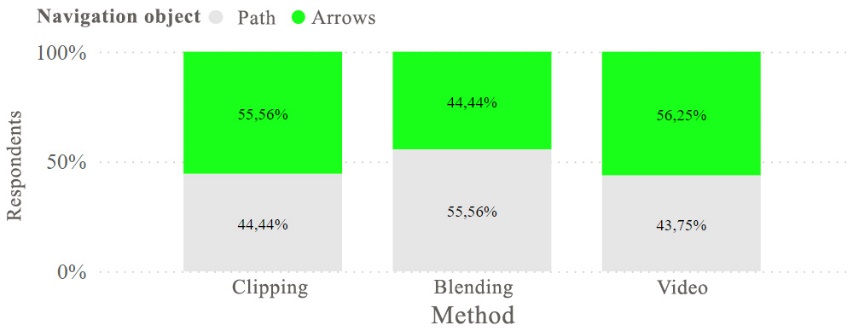


Figure 16. Distribution of navigation object selection in two-dimensional content image conversion methods

The choice of the navigation aids, static path and dynamic arrows, is in a similar proportion for the different methods of repositioning two-dimensional content (Fig. 16). However, it can be distinguished that the blending method preferred the static object more often. As the image is blurred when moving between positions, seeing not only the direction but also the distance to be travelled to reach a clear image potentially reduces the incidence of cyber-sickness and makes users feel more confident. In contrast, methods of cropping and changing the view of the video were more likely to prioritise the dynamic navigational object. Moving in a static line without video refresh during video clipping may be more likely to lead to cyber-sickness due to image-motion dissonance. In contrast, when using a dynamic object - an arrow - and seeing the direction but not the distance, the dissonance is potentially less. The reason for the increased popularity of the pointer in other methods may be the same. The transformation shows a moving image, while the arrows only show the direction of movement, causing less dissonance between the coincidence of the movement of the depicted objects. In the case of a road, it only moves when the user moves, so if the user stops at the start of a dynamic image change, so does the displayed road. This creates dissonance between the movement of the rendered images.

The best scoring road navigation object was in the fusion method and the worst in the video overlay method (Table 5). Arrows scored best in the cut-off method and worst in the video method. Although the latter method has on average the worst estimates of the auxiliary navigation objects, the high standard deviation suggests that the distribution of estimates of navigation objects for this method is very wide. As video is generally the worst-rated method for changing the viewing position, the estimation of navigational objects in this method may also have suffered as a result.

Table 5. Estimates and standard deviations of navigational objects used in two-dimensional content image conversion methods

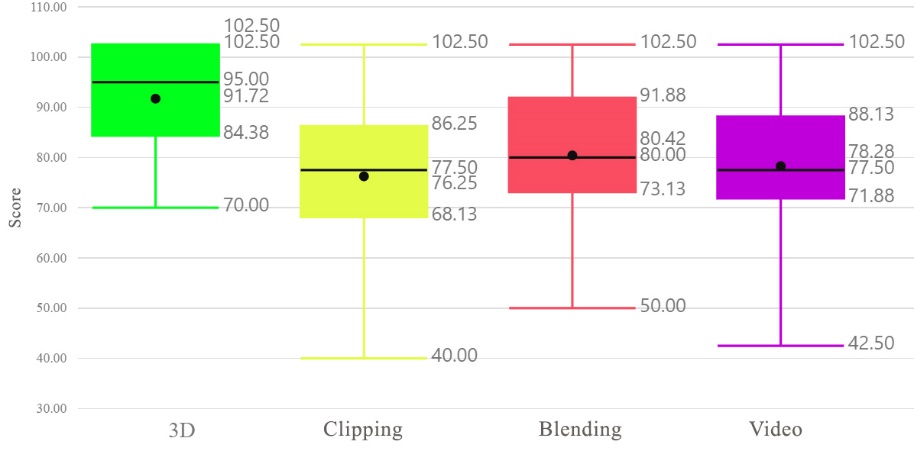
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Method | Average road object estimate | Stand. deviation of the road object estimate | Average of the estimates for the indicator object | Standard deviation of the indicator object estimate |
| Cutting off | 3,83 | 0,96 | 4,00 | 1,00 |
| Linking | 4,11 | 0,94 | 3,78 | 0,97 |
| Video | 3,06 | 1,34 | 3,63 | 1,05 |

For the usability estimates collected from the experiments, a 95% confidence interval is calculated for each content representation method (Table 6). As the ranges of estimates for the two-dimensional content representation methods overlap, it is concluded that there is a 95 % probability that there is no statistically significant difference between the SUS estimates of the clipping, fusion and video methods. They all fall within the good usability range (68-80.3). However, there is a 95% confidence that there is a statistically significant difference between two-dimensional and three-dimensional content representation, as the confidence intervals do not overlap. Also, the usability of three-dimensional content is 95% likely to fall into the category of excellent usability.

Table 6. Totally 95% confidence intervals for SUS estimates for all content mapping methods

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Method | Lower limit of the mean | Average | Upper limit of the average | Standard deviation |
| Cutting off | 69,4 | 76,3 | 83,2 | 15,9 |
| Linking | 73,9 | 80,4 | 86,9 | 14,7 |
| Video | 70,6 | 78,3 | 86 | 15,7 |
| Three-dimensional | 86,8 | 91,7 | 96,6 | 10,7 |

The chart (Fig. 17) gives an overview of the distribution of usability estimates according to the SUS survey. The three-dimensional method of displaying content is clearly the best-rated. The majority of the scores are distributed between 84.3 and 102.5, indicating that the usability of this method is excellent. The usability of the blending and video methods is distributed between good (73.13) and excellent (91.88). Although the mean score (78.28) for the content representation method using video for transformations is slightly lower than the mean score (80.42) for the fusion method. The method of changing the viewing position using clipping scored the lowest in this category. The 25th percentile of the estimate is on the borderline of moderate usability. Overall, the usability of all three two-dimensional virtual reality methods is good. In contrast, the usability of the fusion method ranges between good and excellent.



|  |
| --- |
| Figure 17. Distribution of system usability estimates for each content display method |

Figure 18. Distribution of overall cybersickness symptom sensory ratings for different methods of representing virtual reality content.

The evaluation of the usability of virtual reality content rendering techniques also looks at the expression of symptoms of cybersickness. By scoring the symptoms according to a set of formulas described and an overall symptom score is calculated. A higher score means a more strongly expressed symptom. The distribution of the total score can be seen in the graph (Figure 18). The lowest score distribution is observed in three-dimensional content (mean 3.74) and the highest in a two-dimensional content rendering method that uses image cropping to change the viewing position (mean 90.54). Viewing position change realised by blending and viewing position change realised by a video are similarly weighted. The merge method scores slightly more favourably, with a smaller range of estimates and a mean score of 20.99. In contrast, the video method has a slightly wider range of estimates, with a mean of 25.95 points. The same trend in the distribution of method estimates can also be observed for each of the cyber-disease components individually.

The SSQ questionnaire was used to assess the symptoms of cyber-sickness that affect the usability of the methods. For the sample of data collected during the experiment, the internal consistency of nausea, oculomotor impairment and disorientation domains was calculated using Cronbach's alpha (Table 7). The coefficients obtained are greater than 0.8, indicating good internal consistency of the SSQ scale for this data.

Table 7. Cronbach's alpha coefficients of the SSQ scale items for the total sample of experimental data.

|  |  |  |
| --- | --- | --- |
| SSQ dedicated | Cronbach's alpha | Data sample size |
| Nausea | 0,810 | 68 |
| Oculomotor disorders | 0,863 | 68 |
| Disorientation | 0,893 | 68 |

When looking at the symptoms of cyber-sickness caused by the different mapping methods according to the gender of the participants, it was observed that women rated cyber-sickness symptoms more strongly than men across all methods (Figure 19). Among women, the strongest symptoms are caused by video-mediated changing of viewing positions, while among men they are caused by clipping-mediated changing of viewing positions. Both men and women reported feeling the least symptoms with a fusion-realized change of viewing position. Men rated the symptoms caused by the video-assisted method of changing the viewing position very similarly to those caused by the fusion-assisted method. Given the previous finding that women are more susceptible to cybersickness than men in the sample of respondents in this experiment, we can assume that the video method is not more unpleasant for those more susceptible to cyber-sickness. In the absence of highly pronounced sensitivity, individuals experience more pronounced symptoms using the cutoff method.

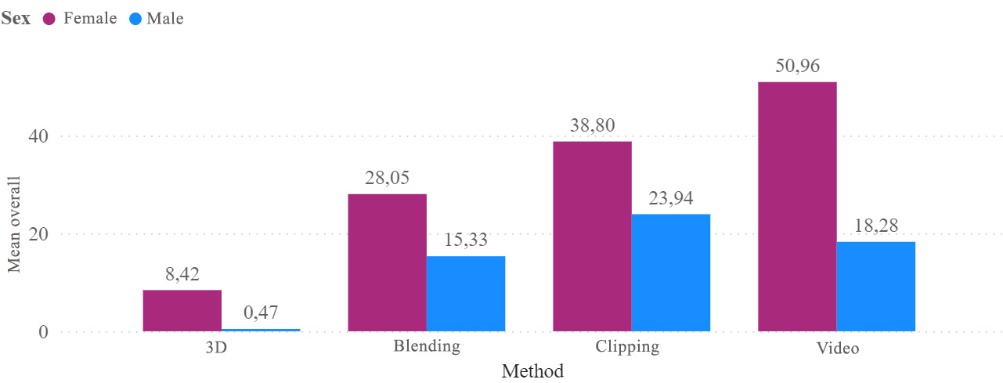


Figure 19. Distribution of the mean overall cyber-symptom score by gender of respondents.

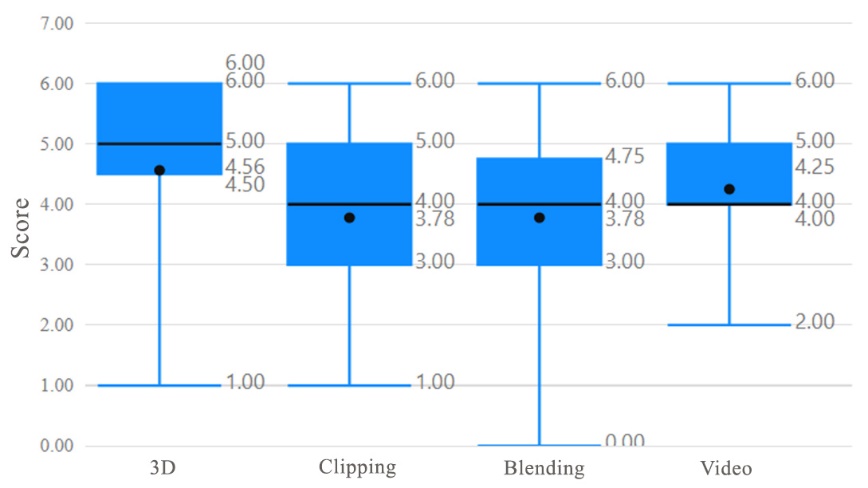
In the case of a content display using the cropping method, the image is suddenly refreshed during the change of image, while the change of position is visually invisible, resulting in a strong dissonance between what the user sees and what he does. In contrast, with the video method, the change is not only felt but also seen. Therefore, there is less dissonance between what the user sees and the movements they make. However, this is the case when the user's movement coincides with that of the visible image. If the user stops moving when the video exchange starts, there is a dissonance between the visible image and the perceived movement, and symptoms of cyber-sickness are observed.

The IPQ survey was used to assess the overall feeling of being in the environment.

Table 8. Cronbach's alpha coefficients of the IPQ scale items for the experimental data samples

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Criteria | Spatial judgement | Get involved | Reality | IPQ |
| Full programme | 0,748 | 0,763 | 0,64 | 0,861 |
| Three-dimensional rendering | 0,842 | 0,803 | 0,759 | 0,916 |
| Two-dimensional, by cutting off | 0,485 | 0,647 | 0,616 | 0,769 |
| Two-dimensional, merging | 0,826 | 0,643 | 0,693 | 0,846 |
| Two-dimensional, displaying a video | 0,619 | 0,808 | 0,232 | 0,729 |
| Sample of variables | 5 | 4 | 4 | 14 |

The internal consistency of this survey for the sample of data collected during the experiment is adequate for each mapping method, with coefficients > 0.5 for a sample of fewer than 10 variables and > 0.7 for a sample of more variables measured. From the coefficients calculated in Table 8, it can be seen that almost all coefficients indicate internal consistency. The only IPQ variable for which the internal consistency is too low is the realism induced by the proposed two-dimensional method.

****

|  |
| --- |
| Figure 20. Sense of presence ratings for different content representation methods |

The distribution of the data shown in the graph (Fig. 20) shows that three-dimensional content in virtual reality produces on average the highest sense of co-presence in space (4.56). Among the methods of displaying two-dimensional content, the video evokes the greatest sense of presence. It has the lowest range of estimates and the highest mean (4.25). The lowest scoring method in this category is the blending method, with the cropping method scoring slightly better.

When it comes to the components of the sense of being in virtual space (engagement, spatial perception of presence, realism), three-dimensional content scored the highest on average in all categories. Among the two-dimensional methods, engagement scored on average highest for the fusion method, while spatial awareness and realism scored on average higher for the proposed transformation method using video.

Table 9. Statistics of the Sense of Presence survey estimates for content with different rendering methods

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | Average | Stand. Error | Stand. Deviation | Variation |
| Video surface | Get involved | 12,94 | 0,755 | 3,021 | 9,129 |
| Spatial judgement | 17,69 | 0,902 | 3,610 | 13,029 |
| Reality | 10,69 | 0,700 | 2,798 | 7,829 |
| Topping by cutting off | Get involved | 11,83 | 0,825 | 3,502 | 12,265 |
| Spatial judgement | 16,72 | 0,803 | 3,409 | 11,624 |
| Reality | 10,11 | 0,449 | 1,906 | 3,634 |
| Surface by blending | Get involved | 13,44 | 0,525 | 2,228 | 4,967 |
| Spatial judgement | 17,61 | 0,714 | 3,031 | 9,193 |
| Reality | 10,66 | 0,681 | 2,890 | 8,353 |
| Three-dimensional content | Get involved | 14,35 | 0,821 | 3,285 | 10,796 |
| Spatial judgement | 19,43 | 1,004 | 4,016 | 16,129 |
| Reality | 11,56 | 0,741 | 2,965 | 8,796 |

The sense of spatial presence seems to be most favoured in the three-dimensional mapping method (Fig. 21). Among the two-dimensional content representation methods, video again scores highest. Although the range of most of the fusion estimates is noticeably lower than for video, the averages for both are very similar (17.61 and 17.69 respectively). The lowest scoring method in this category is the clipping method, although its average (16.72) is not far behind the latter two.

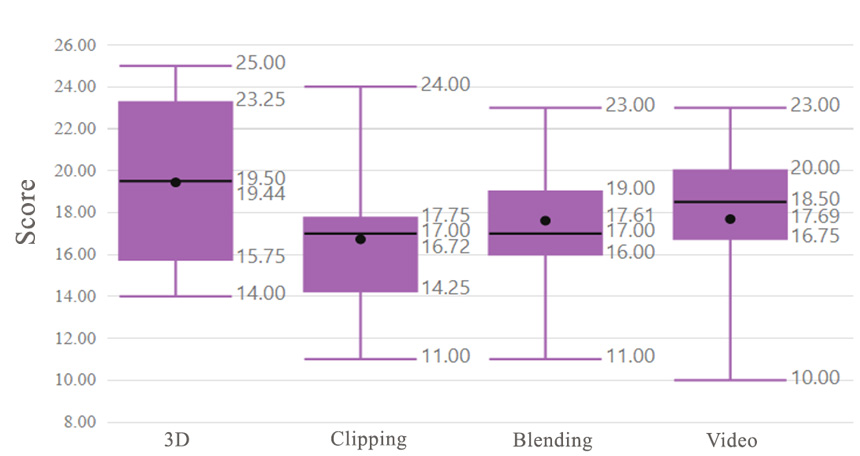


Figure 21. Spatial sense of presence ratings for different content display methods

The engagement chart shows that the three-dimensional mapping method scores best (Figure 22). Among the two-dimensional methods, in this case, fusion has the highest score range (12-15) and the highest mean (13.44). The video has a wider range, indicating more distinct estimates. Although the distribution of ratings is more stable, the cut-off has lower ratings and therefore a lower mean than the video (11.83 and 12.94 respectively). The blending method scored the highest, possibly because, during the display of the content, the user not only sees the overlapping visual destination when changing position but also controls the brightness of the destination when moving towards it.

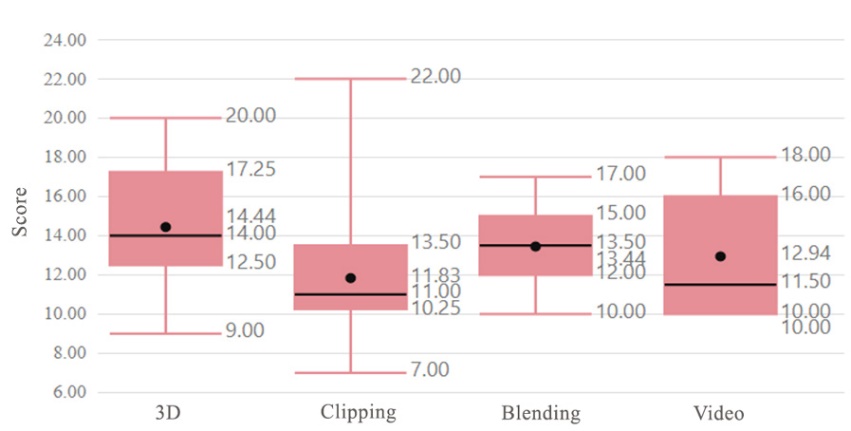


Figure 22. Engagement ratings for different content display methods

The graph for the assessment of the sense of realism shows that the two- and three-dimensional distributions of method scores are very similar (Figure 23). The most notable method is the beheading method, which has the most stable distribution of estimates, albeit in the lowest range of estimates. This method has the lowest mean realism (10.11). This can be explained by the absence of a consistent transformation, which has a noticeably large impact on the perception of realism. The video method scores best in this category, where the realised change in viewing positions is consistent and immediate.

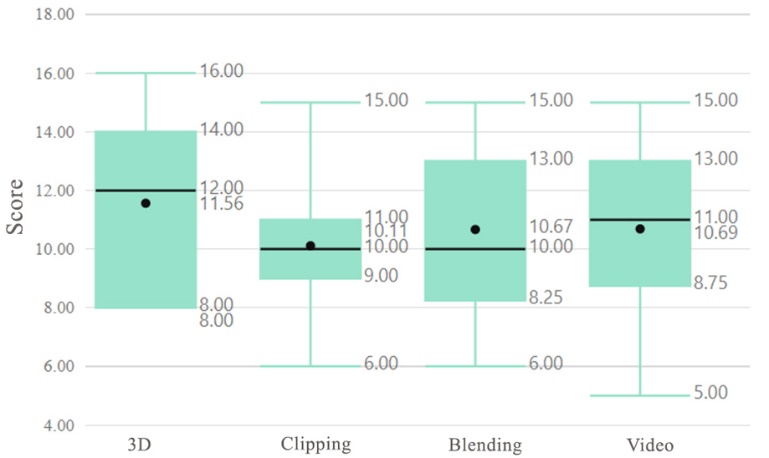


Figure 23. Ratings of the sense of realism for different content representation methods

The experiment found that all two-dimensional methods of content representation are usable. In this area, the fusion method slightly outperforms the video and clipping methods. The video and clipping methods prefer a dynamic navigation method, while the fusion method prefers a static one. Two-dimensional methods using fusion and video transformations are the least likely to cause symptoms of cybersickness. However, the fusion method produces fewer symptoms in individuals more susceptible to vestibular disorders, while the video method produces the most severe symptoms, and is therefore not suitable for use in individuals more susceptible to cyber-sickness. In terms of the sense of presence in the virtual environment induced by the different methods, the video-over-video method is superior to other two-dimensional methods of displaying content, as it induces a greater sense of co-presence in the environment, realism and spatial presence.

**4. Conclusions**

The analysis of the devices has identified devices that allow 6 degrees of freedom. The analysis of the devices according to the qualitative parameters identified has led to a generic list of devices that are suitable for the field of application. For the experimental studies, it was decided to use Oculus Quest 2 and Windows Mixed Reality devices. The choice of a virtual reality system must take into account the autonomy and mobility of the virtual reality system, with an emphasis on mobility.

Virtual reality content creation methods were compared in terms of the type of content being created, the degree of freedom of movement, constraints, and additional equipment used. Based on the results, a map of virtual reality content creation methods was drawn up. For the research, it was decided to choose the creation of video content using three-dimensional simulation in order to compare the results obtained with traditional methods implementing 6-degrees-of-freedom movement.

A literature review found that the use of virtual reality can lead to negative symptoms in terms of well-being - cyber-sickness. Symptoms are caused by three factors: the parameters of the virtual reality equipment, the dynamism of the content and the individual characteristics of the user. In order to reduce negative symptoms for the user, it is recommended to choose VR equipment with a stable rendering latency. In addition, the content should minimise the discrepancy between the information from the vestibular apparatus and the apparent motion, which is the cause of the symptoms. The individual characteristics of users that make them susceptible to cyber-sickness must be assessed before the content is presented to determine whether they are the main source of symptoms.

Taking into account the granularity and immersive property of three-dimensional scene content, and the property of lower rendering cost of video content, a methodology is proposed that combines modelling and video production. The methodology enables constrained movement in 6 degrees of freedom in prescribed directions between viewing positions. Each change of the viewing position is a video recording in which the camera position is changed in predefined steps. The proposed methodology is thus assumed to reduce the computational cost of detailed content representation by simulating 6 degrees of the freedom movement.

The software faced the problem of seamless transformation between different videos, as a certain amount of time is needed to prepare and play a video. To overcome the problem, the system was designed in such a way that the required videos are loaded in advance before the transformations take place. Multiple players are utilised to play them. In this way, a smooth video transformation was ensured.

Experimental studies assessing the usability, inclusiveness and performance of the proposed methodology have shown:

1. According to the survey, the video method image gives a lower sense of presence than the three-dimensional scene model, which provides 6 degrees of freedom. This, therefore, refutes the hypothesis that the proposed method can maintain the same inclusiveness as a three-dimensional scene;
2. When comparing the proposed transformation with methods that do not have a consistent transformation between viewing positions, overall presence, spatial judgement and realism are better evaluated in the video method (see Section 4.1).
3. The experiments showed an above-average estimate of the usability of the video method (estimate: 78.28), but no statistically significant difference between its usability and that of conventional two-dimensional mapping methods. (see section 4.1)
4. The usability of the proposed method suffers from adverse physiological effects compared to three-dimensional content representation. The post-usage symptom ratings for cyber-sickness are 25.95 for the proposed method and 3.74 for three-dimensional imaging, respectively;
5. There has been a trend that more susceptible users, who are more susceptible to symptoms of cyber-sickness, are less likely to use the proposed method than other two-dimensional content representation methods.
6. When assessing the impact of three-dimensional scenes and interactive video on the performance of mobile systems, it was found that only the video method realises detailed content at a stable rate of 72 frames per second.
7. The study of the two proposed navigation objects (dynamic and static) in the video did not show an advantage of one or the other method. Also, no advantage was found for the two proposed distances between viewing positions.
8. The study found that women and younger age groups prefer monoscopic imaging. In the group of respondents studied, there was a tendency for those with higher sensitivity not to choose stereoscopic imaging, which causes unwanted symptoms.

**Author Contributions**

Conceptualization, A.P.; methodology, A.P., A.S.; software, A.P. and A.S., T.B.; data curation, A.P., A.S, M.V.; writing — original draft preparation, A.P., A.S., M.B, C.C. D.B, L.P; writing — review and editing, T.B., M.B, M.V.; visualization, A.P., A.S and T.V.. All authors have read and agreed to the published version of the manuscript.

**Funding:**

This research was funded by Kaunas University of Technology.

**Institutional Review Board Statement**

Ethical review and approval were waived for this study, as this study involves no more than minimal risk to subjects.

## **Informed Consent Statement**

Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement**

The data presented in this study are available on request from the corresponding author. The data are not publicly available due to the data restriction policy by the grant provider.

**Conflicts of Interest:** The authors declare no conflict of interest.

**References**

1. Boletsis, C., & Cedergren, J. E. (2019). VR locomotion in the new era of virtual reality: an empirical comparison of prevalent techniques. Advances in Human-Computer Interaction, 2019.
2. Takatori, H., Yano, H., & Iwata, H. (2015). Panoramic movie-rendering method with superimposed computer graphics for immersive walk-through system. In Haptic Interaction (pp. 277-284). Springer, Tokyo.
3. Kim, Y. M., Lee, Y., Rhiu, I., & Yun, M. H. (2021). Evaluation of locomotion methods in virtual reality navigation environments: An involuntary position shift and task performance. International Journal of Human-Computer Studies, 155, 102691.
4. Soler-Domínguez, J. L., de Juan, C., Contero, M., & Alcañiz, M. (2020). I walk, therefore I am: a multidimensional study on the influence of the locomotion method upon presence in virtual reality. Journal of Computational Design and Engineering, 7(5), 577-590.
5. Cherni, H., Métayer, N., & Souliman, N. (2020). Literature review of locomotion techniques in virtual reality. International Journal of Virtual Reality.
6. Mayor, J., Raya, L., Bayona, S., & Sanchez, A. (2021). Multi-technique Redirected Walking Method. IEEE Transactions on Emerging Topics in Computing.
7. Cherni, H., Nicolas, S., & Metayer, N. (2021). Using virtual reality treadmill as a locomotion technique in a navigation task: Impact on user experience–case of the KatWalk. International Journal of Virtual Reality, 21(1), 1-14.
8. Rantala, J., Kangas, J., Koskinen, O., Nukarinen, T., & Raisamo, R. (2021). Comparison of controller-based locomotion techniques for visual observation in virtual reality. Multimodal Technologies and Interaction, 5(7), 31.
9. Paris, R. A., Buck, L. E., McNamara, T. P., & Bodenheimer, B. (2022, March). Evaluating the Impact of Limited Physical Space on the Navigation Performance of Two Locomotion Methods in Immersive Virtual Environments. In 2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR) (pp. 821-831). IEEE.
10. Pallavicini, F., Pepe, A., & Minissi, M. E. (2019). Gaming in virtual reality: What changes in terms of usability, emotional response and sense of presence compared to non-immersive video games?. Simulation & Gaming, 50(2), 136-159.
11. Arya, S. (2020). New Technologies and Applications of Motion Tracking in Virtual Reality Systems. International Journal of Advanced Research in Computer Science, 11(5).
12. Noghabaei, M., & Han, K. (2021). Object manipulation in immersive virtual environments: Hand Motion tracking technology and snap-to-fit function. Automation in Construction, 124, 103594.
13. Qian, K., Arichi, T., Price, A., Dall’Orso, S., Eden, J., Noh, Y., ... & Hajnal, J. V. (2021). An eye tracking based virtual reality system for use inside magnetic resonance imaging systems. Scientific reports, 11(1), 1-17.
14. Li, G. (2021). Optimization and simulation of virtual experiment system of human sports science based on VR. Complexity, 2021.
15. Miller, M. R., Herrera, F., Jun, H., Landay, J. A., & Bailenson, J. N. (2020). Personal identifiability of user tracking data during observation of 360-degree VR video. Scientific Reports, 10(1), 1-10.
16. Lee, Y., Do, W., Yoon, H., Heo, J., Lee, W., & Lee, D. (2021). Visual-inertial hand motion tracking with robustness against occlusion, interference, and contact. Science Robotics, 6(58), eabe1315.
17. Baldominos, A., Saez, Y., & Del Pozo, C. G. (2015). An approach to physical rehabilitation using state-of-the-art virtual reality and motion tracking technologies. Procedia Computer Science, 64, 10-16.
18. Fang, W., Zheng, L., Deng, H., & Zhang, H. (2017). Real-time motion tracking for mobile augmented/virtual reality using adaptive visual-inertial fusion. Sensors, 17(5), 1037.
19. Oh, A. S. (2019). A Study on MTL Device Design and Motion Tracking in Virtual Reality Environments. Journal of information and communication convergence engineering, 17(3), 205-212.
20. Korisky, U., & Mudrik, L. (2021). Dimensions of perception: 3D real-life objects are more readily detected than their 2D images. Psychological Science, 32(10), 1636-1648.
21. Weller, P. D., Rabovsky, M., Rahman, R. A. (2019). Semantic knowledge enhances conscious awareness of visual objects. Journal of Cognitive Neuroscience, 31(8), 1216–1226.
22. Marini, F., Breeding, K. A., Snow, J. C. (2019). Distinct visuo-motor brain dynamics for real-world objects versus planar images. NeuroImage, 195, 232–242.
23. Borghi, A. M., Riggio, L. (2015). Stable and variable affordances are both automatic and flexible. Frontiers in Human Neuroscience, 9, Article 351.
24. Gledhill, D., Tian, G. Y., Taylor, D., & Clarke, D. (2003). Panoramic imaging—a review. Computers & Graphics, 27(3), 435-445.
25. Neumann, U., Pintaric, T., & Rizzo, A. (2000, October). Immersive panoramic video. In Proceedings of the eighth ACM international conference on Multimedia (pp. 493-494).
26. Schäfer, A., Reis, G., & Stricker, D. (2021). Investigating the sense of presence between handcrafted and panorama based virtual environments. In Mensch und Computer 2021 (pp. 402-405).
27. Lai, P. K., Xie, S., Lang, J., & Laganière, R. (2019, March). Real-time panoramic depth maps from omni-directional stereo images for 6 dof videos in virtual reality. In 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR) (pp. 405-412). IEEE.
28. Gil, A., Khurshid, A., Postal, J., & Figueira, T. (2019, October). Visual assessment of equirectangular images for virtual reality applications in unity. In Anais Estendidos da XXXII Conference on Graphics, Patterns and Images (pp. 237-242). SBC.