Fishtank Everywhere: Improving Viewing Experience over 3D Content

Lucas S. Figueiredo[1], Edvar Vilar Neto[1], Ermano Arruda[1], João Marcelo Teixeira[1,2], and Veronica Teichrieb[1]

[1] Federal University of Pernambuco, Recife PE 50740-560, Brazil
{lsf,excvn,eaa3,jmxnt,vt}@cin.ufpe.br
cin.ufpe.br/voxarlabs

[2] Federal Rural University of Pernambuco, Recife PE 52171-900, Brazil
www.deinfo.ufrpe.br

Abstract. The goal of this work is to analyze the user experience of the motion parallax effect on common use displays, such as monitors, tvs and mobile devices. The analysis has been done individually for each device and comparing each other to understand the impact on the immersion of such media. Moreover, we focused on understanding the user impression on the change of an usual passive visualization paradigm to the interactive visualization possibility allied to the motion parallax effect.

Keywords: depth perception, cross platform experience, fishtank effect, interactive visualization.

1 Introduction

Since the early stages of Virtual Reality research the perception of depth is a highlighted issue. Among the explored depth cues there is the motion parallax as an option for interactive visualization. Motion parallax is the effect that allows the user to distinguish depth on the scene by moving his viewpoint (i.e. user eyes). The perception of the third dimension comes by the intrinsic comparison of which object in the scene is moving faster or slower in relation to the viewpoint displacement.

This type of depth perception is simulated on planar displays by a technique called Fishtank Virtual Reality, which uses the information of the user viewpoint (usually gathered by an additional sensor) to change the scene according to this movement. Considering the current availability of depth cameras and new face detection and tracking algorithms, the Fishtank technique gains space in the common use scenario. Moreover, as we show later, nowadays it is possible to enable the effect without requiring the use of glasses or any other additional attached sensors or displays. That said, it is possible to apply the Fishtank technique on nearly any 3D content on present devices.

The user experience of the Fishtank Virtual Reality technique working on different displays has been evaluated in this work. In our experiment, people used applications (e.g. 3D rendered games) coupled with the Fishtank Virtual

A. Marcus (Ed.): DUXU 2014, Part I, LNCS 8517, pp. 560–571, 2014.

Reality technique and without it. The effect experience was analyzed regarding its relevance and the user satisfaction and acceptance of the technique. For this, after the participants completed the proposed tasks, they participated on a semi-structured interview and answered a Likert-scale questionnaire, aimed to collect subjective impressions.

The paper is structured as follows. Section two exposes the chronology and the background concepts about the motion parallax illusion. Section three explains the proposed methodology and experiment. Section four shows the experiment results and analysis. Finally, section five presents the conclusion and future work.

2 The Fishtank Virtual Reality Technique

2.1 Chronology

This section provides information regarding the history and evolution of the Fishtank Virtual Reality technology, from the 1960's when it was only an utopic idea, to nowadays when it is possible to apply it also on telepresence scenarios with real content displayed. This list is not intended to cover all scientific works related to the technology; instead it points important marks along its history.

In the year of 1965, Ivan Sutherland, known by the research community as the "father of virtual reality", discussed about what the "ultimate display" should look like [14]. He stated that such utopic device should pose as a window to the virtual world, capable of simulating a complete immersive environment, seamlessly from reality itself. Beyond displaying 2D images, it should also provide tridimensional perception, different smells and tactile experiences. By conveying such information, it would be possible to exploit most human senses.

In 1992, Steuer contributed to the definition of the term Virtual Reality [13]. Up to this date, the concept was directly related to the hardware being used, instead of being based on sensations the users experienced. It then defines Virtual Reality in terms of telepresence, in a way that the physical person is transported to and feels like being in a different world (the virtual one). From this moment on, all the hardware used were simply considered instruments or means to implement telepresence. Later in 1992, the first system capable of changing the viewing perspective as the user changed his position was created [4]. Since the displays at that time were CRT and curved, a mathematical model was required to cope with such display shape. This work was named "High Resolution Virtual Reality", since it claimed to have a higher resolution than the head-mounted displays at that time, but it was later called "Fishtank Virtual Reality". In 1993, Ware et al. [15] compare the Fish Tank technique in two distinct scenarios: alone and combined with stereoscopy. This was the first work to relate the technology with other similar systems. Back in 1999, Brooks Jr. et al. discussed several issues regarding virtual reality, from its definition to its history, the technology it requires and devices typically used [2]. Finally, it lists the open problems and remembers us that we are still far from Sutherland's "ultimate display".

In 2006, Demiralp et al. compare the Fish Tank approach to a CAVE environment [5]. The important conclusion of this work was the suggestion that

Fish Tank systems are more effective than CAVEs. In 2008, Kooima generalized perspective transforms so that they could mathematically describe distorted frustums [10]. Such knowledge is necessary in order to correctly distort the viewing frustum to create a Fish Tank Virtual Reality system. In 2009, Maksakov developed an extension for the Fish Tank Virtual Reality technique that comprises much larger screens and cooperative work by using separate viewports [12]. This work solves a previous limitation of the technology because it was not designed for team cooperation. By using a device attached to each user's head, it was possible to modify the scene view accordingly. In 2010, Andersen et al. propose the combination of oblique perspective changes with stereoscopy [1]. They also propose a modified graphics pipeline in order to achieve applicable results, specifically targeting games.

In 2011, Francone et al. explore the application of head-coupled perspective in mobile devices by means of their camera and a face tracking algorithm [8]. It manages to track the user's face, estimate its position, and use it to change the perspective of the scene. This creates the impression of 3D perception in mobile devices and can be used to improve user experience. In 2012, Halamkar et al. redefine Virtual Reality as a computer-simulated environment [6]. They categorize different levels of Virtual Reality, some factors that one should consider when designing such environments, its origins, probable future and challenges that have still to be overcome in order to create ideal virtual environments. Later in 2012, instead of simulating the 3D environment on the TV screen, Heirichs et al. make use of a real scene captured by a robotic camera [7]. The camera position is controlled by tracking the user's head position in order to simulate the Fish Tank Virtual Reality effect, which is associated with the parallax effect to create a 3D perception illusion.

2.2 How It Works

The purpose of the Fishtank technique is to simulate a display behaving like a glass window to the 3D content. This simulation requires the knowledge of the user viewpoint in real time. The new pose of the scenes virtual camera is then calculated in order to render the 3D content considering the new viewpoint but without changing the original viewport (which represents the boundaries of the glass window metaphor). Figure 1 illustrates the concept.

Motion parallax can be experienced in a speeding car, as it can be perceived that trees located far away move slower than the ones closer to the car. The human brain uses this depth cue to define which objects are closer based on their angular relative motion to the observer. Its reasonable to say that during motion, our brain analyses the sequence of different images of an object, which were acquired from different points of view, and combines them together in order to estimate its depth. In contrast, the binocular depth cue of stereoscopy helps humans and other animals to estimate distances from objects without necessarily being in relative motion to them. The slightly two different images formed in each eye retina is sufficient to infer depth, and judge distances accurately. If theres no motion involved, motion parallax simply doesnt work, thus, making

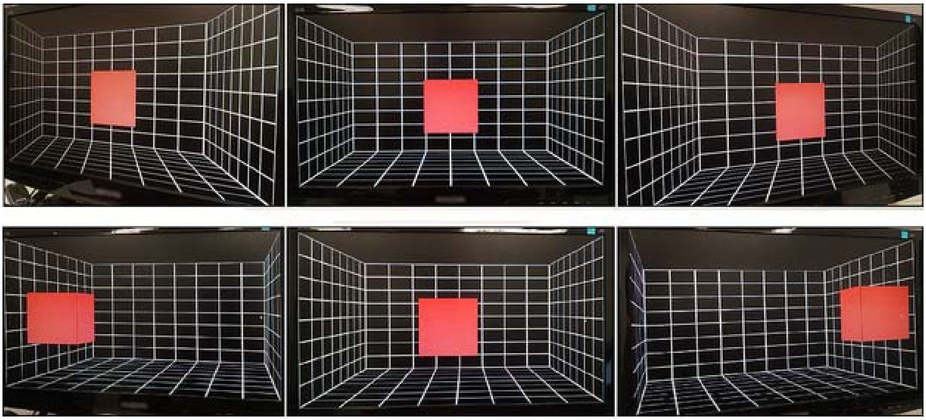

Fig. 1. Top: three different viewpoints of 3D content displayed on a monitor with regular visualization. Bottom: three similar viewpoints of the same 3D content displayed using the Fishtank approach, improving depth perception and immersion.

stereoscopy the main depth cue in such scenarios. A Fishtank Virtual Reality system uses motion parallax as its main depth cue to convey depth perception. Depending on the distance between the object and the observer, motion parallax can be even more important depth cue than stereoscopy itself [3].

To generate the motion parallax depth perception, the system must deform the viewing frustum and move the camera pose according to the users viewpoint position. Thus, a way of tracking the users head position is needed. Formerly, it was done by attaching extra tracking devices to the users head [15], [12], but this approach has inconveniences regarding user physical freedom. Hence, a good solution should not require extra devices attached to the users body. Nowadays, its possible to solve this problem by tracking the users head 3D position on images retrieved from a webcam or similar sensors. Even trackers which can only retrieve the 2D position of the users head, with the addition of an extra step can be used to estimate the 3D viewpoint of the user head based on the image bounding box.

By tracking the 3D position of the users head (x,y,z) a dynamic asymmetric frustum is defined as a function of the users head position. The frustum deforms according to the users current viewpoint, being responsible for half of the motion parallax effect. The other half of the effect is performed by displacing the virtual camera position (without rotating it) according to the users head new position. This displacement brings the virtual screen viewport (four points defining the virtual window) to the same position it was before the distortion of the frustum. As desired by the effect, only the virtual camera point of view changes according to the user head movement, maintaining the virtual screen on the same place. By using this mechanics, it is possible to apply the motion parallax nearly on any 3D rendered content and on any device with head or face tracking capabilities.

3 Cross-Platform Experiment

The main goal of the proposed experiment is to understand the impact of the motion parallax effect over different platform experiences. With that in mind, the setup simulated three different scenarios, namely a TV-like scenario, a Desktop scenario and a Mobile scenario, using the same application over each one. The application used was an open source game called GLTron, which is available for download and can be compiled for PC and Android devices.

3.1 Case Study Application

The choice of GLTron as the case study application aimed to promote a more engaging activity, exploring game mechanics as an enrichment of the experience over 3D contents. Moreover, unlike for example Starcraft and Street Fighter 4, GLTron is a game that needs to be rendered in 3D, because the interaction depends on that. This factor adds relevance to the depth perception and 3D environment exploration by the user. The simplicity of the GLTron interaction is also in favor of a more abrangent experiment, which requires minor user training and still can be challenging and engaging.

The game consists on controlling a futuristic motorcycle on a large and limited square area by turning it to the left or the right. During the movement each player motorcycle creates a wall following its path and this wall blocks the way of every player on the scene. The game was conducted with a single human player and other three artificial intelligence players. Figure 2 illustrates the GLTron game.

The game was compiled to run on each scenario, and then it was properly coupled with head/face tracking capabilities and the needed calculation of both frustum and virtual camera displacements. All three scenarios explored existent tracking technologies, which are widely available. The scenarios are shown in Figure 3 and detailed as follows.

3.2 Scenarios

The first scenario was intended to be a living room experience, in which the user has a 50 inch screen space (displayed by a projector), which was set about 3 meters away from him. In this case the user point of view was tracked using the Kinect device and the body tracking algorithm provided by the Microsoft Kinect Toolkit. Usually the algorithm tracks twenty body points but it was set to the seat-mode in order to be robust to both cases of the user standing and seated. The captured head 3D point was then used as input for the Fishtank effect. In this scenario the user was able to explore the effect seated or if wanted, he could walk in the room (within the Kinect field of view). To interact with the game a Xbox joystick was provided, so the experience aimed to be a game-like activity common in Xbox equipped living rooms.

The second scenario setup consisted of the user seated on a table with a laptop. The face tracking was performed using the laptop webcam images and a tracking algorithm called Face TLD (also known as Predator) [9]. Face TLD is able to

Fig. 2. GLTron game screenshots

detect the user face and track it over time, improving the tracking results as it learns the appearance of the user face. The algorithm, after some execution time, is able to track the user face even considering the head inclination over shoulders and the rotation to left and right. The result of the tracking is a bounding box containing the user face, which is further transformed in a 3D point using the camera horizontal field of view information and providing an average real size width (in meters) for the bounding box. In this case the user was free to decide if he wanted to interact with a joystick or using the laptop keyboard.

The last scenario aimed to reproduce the game experience on mobile devices. The used device was a tablet, and the face tracking was performed by a native Android function for face capturing. The tracking result is a 2D bounding box of the face, which is later converted to a 3D point in order to be used as input for the needed distortions.

3.3 Methodology

The experiment was conducted starting with a few questions about users profile and related to previous experience with games and gesture interaction. A minor test was also conducted to help them discover their dominant eye, this way this information could later be used to adjust the best 3D position as result of the tracking phase.

In sequence, the motion parallax effect was explained and a simple cubic room interactive example was shown to illustrate the functioning of the effect and help

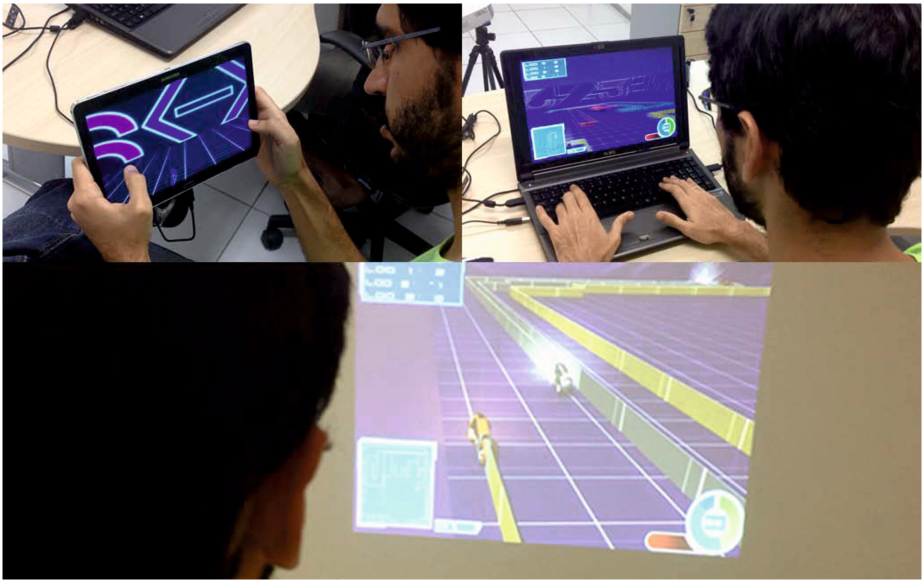

Fig. 3. The three different scenarios tested. Top-left: Mobile. Top-right: Desktop. Bottom: TV-like.

users understand what happens when they move the head and why it happens. This step was necessary since some users may initially be confuse about the effect as a way to control the virtual camera rotation, besides, it is important to make users understand that the interaction of coming closer to the screen does not make the 3D objects to appear bigger. Essentially the opposite occurs, since the idea is that the virtual objects are linked to the real world, and their occupied screen space should decrease once the user comes closer, precisely to compensate that motion. The window metaphor was used to aid the explanation, helping users to understand that when they are closer to a window it is possible to see more of the other side in the same screen space.

After that, users experimented each described scenario with and without the motion parallax illusion. The goal was to play the game for a small amount of time (around five minutes) and try to win at least once on each condition (with and without the effect). This goal was set mainly to help the users engagement with the application, for example encouraging them to visually explore the game 3D world. The experience was observed and documented, and at the end of each scenario some questions were answered. The questions were the following: 1) The use of the effect provides a better visual exploration of the scene; 2) The effect extended immersion in the game scenario; 3) It was more fun playing with the effect. Each question was answered using a 1-5 Likert scale in which 1 means disagree completely, 3 means no impact and 5 means agree completely.

At last, a semi-structured interview was conducted to gather insight about the user overall experience. The guiding questions were: 1) Does the fishtank effect tend to improve user experience? Why? 2) Would you want to see this effect working on other 3D applications? Other games? 3) Were there any differences in experiencing the effect over the different scenarios/displays? What was the difference?

4 Results

A total of six users participated on the experiments, five male and one female, with their age varying from 19 to 29 years old, four having the right eye as dominant and two the left one, and all of them had previous experiences with 3D games and gesture interaction applications. Although the tests were performed with a small group of users, the provided insight was enough to understand main aspects of the interaction coupled with the motion parallax effect across the experimented platforms. Moreover, the users profile showed to be relevant to the experiment considering they already had experience or experienced all three proposed scenarios on common day interactions. This way they were able to correlate the test experience to real life and better answer questions like the one which asks about the use of Fishtank systems on existing applications (question 2 of the semi-structured interview).

The functioning of the effect was easily understood by the users as well. Additionally, on each platform, before playing the game using the effect the user experienced the effect on a static scene, to be familiar with the viewpoint changing inside the game and also to understand the impact his movements would have on the gameplay. Another important aspect of the used game (GLTron) is that during the experience the user is compelled to laterally observe the virtual scene in order to preview further collisions or other approaching players.

4.1 TV-Like Results

This setup drew amusement from most users at the first moment. They started to walk in the room and experience the effect. They had the option to stay seated, even though all of them (after a few seconds of being seated) decided to play the game standing and moving around to explore the virtual scene according to their gameplay needs. The influence on gameplay can be further analyzed and maybe used on the game design. For example, in this case, when using the effect a common move was reproduced by most users. They went in a straight line near and alongside one of the stage boundaries (from where no surprises could arise) and then moved to the other side of the screen to better visualize the other part of the scene from where other players could come and block the path.

As side-effect was perceived as the users intended to move the viewpoint by crouching in order to better see the horizon. Normally this viewpoint is not available, but since the users had the possibility to explore the visualization sometimes they sacrificed their comfort in order to better play the game. This lead to another thought of experience design in which the application should consider that sometimes the user may exploit the provided visualization freedom in a way that may not be wanted in first place (e.g.: the game is made to be played for hours, in a comfortably seated position).

When the users played the game without the effect they immediately felt the difference. A common behavior was to reproduce the interactive visualization movements but after a few tries they were disappointed but convinced that they would have to play without this possibility. One of the users at the first moment of experience without the effect said hey, its a lot different.

4.2 Desktop Results

In general, the playing experience using the effect on the laptop was not well accepted. The first perceived problem was the face tracking output jitter and drift. These are common tracking problems and have always been a challenge for all sort of algorithms. The jitter is result of the tracking imprecision and produces a shaking effect even if the tracked object remains still. The drift problem is a behaviour in which the tracking result is always delayed in relation to the real tracked object position. The Face TLD algorithm is a state of the art solution for tracking faces, still, its results present an amount of jitter and drift which turns the interactive visualization in a not reliable experience.

Another perceived point was that since the user is seated in front of the laptop in a more restricted body position the head movements are not effective as wanted. Even when the virtual camera accompanied the head correctly (regardless the jitter and drift problems) the viewpoint displacement was not enough to show the intended part of the scene, and the user could not move further due to the camera field of view limitation and also due to body position limitations (since the user was seated). Once this was perceived additional experiments were conducted using an increased movement effectiveness, i.e. the head movement was set to move more than the normal ratio, and this way the virtual camera moved the double of the normal movement to facilitate scene exploration. In this variation we perceived the effect could be effectively used to explore the virtual scene, which means a scale factor between the head movement and the virtual camera displacement may be useful. This additional test was conducted out of record and did not influence the responses showed in Figure 4.

4.3 Mobile Results

Regarding the face tracking, a similar issue to the laptop emerged in this case. The used algorithm was the face capturing available on the Android 4.3 version and returned the bounding box of the face, which later was converted to the face 3D point. The jitter and drift were also present and undermined the interaction. On the other hand, the horizontal field of view of the tablet front camera was wider than the laptop one and the movement restriction was not an issue since it was easier to move the head in relation to the device. So regardless the tracking precision and response time issues, the virtual scenario exploration was successful (without the need of any scale factor).

Another interesting behaviour was observed in which the users instead of moving the head to see from a different point of view, rotated the tablet to one side or to the other. Since the effect uses the relationship between the user face and the screen, the rotation interaction produced a similar result, showing the same desired virtual scene part on the screen. This type of interaction was more comfortable, requiring minor physical effort, thus it was reproduced along the experience time.

On the other hand, this rotation movement presents two issues. The first is that the relationship between the real world and the virtual world is changed once the user moves the mobile device. This does not occur in the other two scenarios since the display is fixed in the real world. By moving the tablet the virtual world moves accordingly, this way the impression is that the tablet is a window to the virtual world however this virtual world is coupled on the tablet and does not have a fixed relationship with the real environment.

The second issue related to the rotation interaction was present when the user choses to rotate and at the same time had to press the left or right in order to turn the motorcycle. Considering these two simultaneous movements, it was perceived that sometimes the user reaction was side-inverted in response to his needs.

Figure 4 presents the overall results of the objective questions about visual exploration, immersion and fun. As discussed before, the issues presented by the tracking algorithm impacted user experience on the Mobile (tablet) and Desktop (laptop) cases. On the other hand the living room (TV-like) experience was very well accepted.

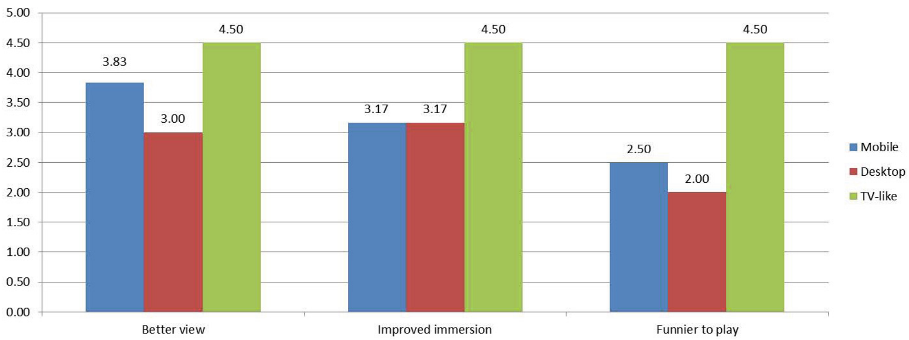

Fig. 4. Result comparison for the three tested scenarios. The values represent the mean scores given by all six tested subjects.

The semi-structured interview returned interesting additional feedback as well. Here follow some of the user responses on question 1 asking if the fish-tank effect had improved the experience: "Yes, because you feel more immersed in the game. This game is improved with the effect" and "In an ideal scenario, I think it has the potential to improve the experience".

Question 2 asked the users if they would like to use the effect on other games or 3D applications: “It would be cool. Because it lets us look at the best objects”, “Yes, I think it would be really cool if there was a more stable version” and “Yes, because with this effect it was possible to play in a fun way. Without it I would have found the game boring”.

Finally, question 3 asked about the experience on the different scenarios: “Yes, the screen size influences. Using the tablet is worse because you keep your hands attached to the device and by moving the device I often moved the camera unintentionally”, “Yes, the Kinect provided the best experience, because the sensitivity was higher. When comparing them, the Desktop version presented almost none viewpoint change” and “The tablet was the worst experience. It was the worst way of controlling the camera and holding the device, it was annoying and tiring”.

5 Conclusion

We have applied and analyzed how the Fishtank Virtual Reality technique can be significant and useful for immersive interactions in different common displays, such as monitors, TVs and tablets. In our study, we could notice that users had no difficulties in comprehending the technique. However, while using it in the performed experiments, some issues arose mainly related to current face tracking solutions and scenario restrictions. The case study application was the GLTron game, and the effect showed to be more than a cosmetic improvement, being useful for the gameplay, providing additional visualization capabilities that improved user performance on the game and enjoyment of the game. This pointed a possibility of changing different visualization paradigms for users, emerging from a passive to an interactive viewing experience of 3D contents, but also pointed the need of more stable, precise and faster face tracking methods for the use on monitors and tablets. The use of the Kinect for tracking was well accepted, which suggests that similar depth-enabled devices coupled with laptops and mobile devices should provide a good enough tracking result.

As future works, the first intent is to improve the used face tracking method by using new algorithms and devices to get rid of the jitter and drift problems observed during the experiment. New depth sensors are emerging nowadays and it may represent a significant improvement on user experience. Moreover, an additional work is planned to use the Fishtank technique on mobile devices also including the devices additional sensors to correctly place the virtual world registered to the real environment. Moreover, this technique can be applied to Augmented Reality scenarios, making the tablet or smartphone look like a transparent glass (rather than a window metaphor presented on the Virtual Reality case). This way, the user should experience the augmentation more naturally.

References

1. Holst, J., Andersen, A.S., Vestergaard, S.E.: The implementation of fish tank virtual reality in games: Exploring the concepts of motion parallax simulation and stereoscopy (January 2014)
2. Brooks, F.P.: What's real about virtual reality? IEEE Comput. Graph. Appl. 19(6), 16–27 (1999)
3. Cutting, J.E., Vishton, P.M.: Perceiving layout and knowing distances: the integration, relative potency and contextual use of different information about depth. In: Epstein, W., Rogers, S. (eds.) Handbook of Perception and Cognition. Perception of Space and Motion, vol. 5, pp. 69–117 (1995)
4. Deering, M.: High resolution virtual reality. In: Proceedings of the 19th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1992, pp. 195–202. ACM, New York (1992)
5. Demiralp, A., Jackson, C.D., Karelitz, D.B., Zhang, S., Laidlaw, D.H.: Cave and fishtank virtual-reality displays: A qualitative and quantitative comparison. IEEE Trans. Vis. Comput. Graph. 12(3), 323–330 (2006)
6. Halarnkar, P., Shah, S., Shah, H., Shah, H., Shah, A.: A review on virtual reality. IJCSI International Journal of Computer Science Issues 9(6), 323–330 (2012)
7. Heinrichs, C., McPherson, A.: Recreating the parallax effect associated with fishtank vr in a real-time telepresence system using head-tracking and a robotic camera. In: ISMAR, pp. 283–284. IEEE Computer Society (2012)
8. Laurence, N., Francone, J.: Using the user's point of view for interaction on mobile devices. In: Conference Proceedings of IHM (October 2011)
9. Kalal, Z., Matas, J., Mikolajczyk, K.: P-n learning: Bootstrapping binary classifiers by structural constraints. In: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 49–56 (June 2010)
10. Kooima, R.: Generalized perspective projection (January 2014)
11. Lee, J.C.: Hacking the nintendo wii remote. IEEE Pervasive Computing 7(3), 39–45 (2008)
12. Maksakov, E., Booth, K.S., Hawkey, K.: Whale tank virtual reality. In: Proceedings of Graphics Interface, GI 2010, pp. 185–192. Canadian Information Processing Society, Toronto (2010)
13. Steuer, J.: Defining Virtual Reality: Dimensions Determining Telepresence. In: Communication in the Age of Virtual Reality, pp. 33–56. L. Erlbaum Associates Inc., Hillsdale (1995)
14. Sutherland, I.E.: The ultimate display. In: Proceedings of the IFIP Congress, pp. 506–508 (1965)
15. Ware, C., Arthur, K., Booth, K.S.: Fish tank virtual reality. In: Proceedings of the INTERACT 1993 and CHI 1993 Conference on Human Factors in Computing Systems, pp. 37–42. ACM, New York (1993)