

An Exploratory Comparison of the Visual Quality of Virtual Reality Systems Based on Device-Independent Testsets

Robert Manthey¹(✉), Marc Ritter², Manuel Heinzig¹, and Danny Kowerko¹

¹ Junior Professorship Media Computing,

Technische Universität Chemnitz, 09107 Chemnitz, Germany

{robert.manthey,manuel.heinzig,danny.kowerko}@informatik.tu-chemnitz.de

² Professur Medieninformatik, Hochschule Mittweida, University of Applied Sciences,

Technikumplatz 17, 09648 Mittweida, Germany

marc.ritter@hs-mittweida.de

Abstract. Nowadays, several different devices exist to offer virtual, augmented and mixed reality to show artificial objects. Measurements of the quality or the correctness of their resulting visual structures are not developed as sophisticated as in the classical areas of 2D image and video processing. Common testsets for image and video processing frequently contain sequences from the real world to reproduce their intrinsic characteristics and properties as well as artificial structures to provoke potential visual errors (see Fig. 1a). These common but traditional testsets are nowadays faced with rapid technical developments and changes like HD, UHD etc. improved surround sound or multiple data streams. That results in a limitation of the testsets usability and their ability to evoke visual errors. To overcome those limitations, we developed a system to create device-independent testsets to be used in the area of virtual reality devices and 3D environments. We conduct an empirical evaluation of most recent virtual reality devices like HTC Vive and Zeiss Cinemizer OLED, aiming to explore whether the technical hardware properties of the devices or the provided software interfaces may introduce errors in the visual representation. The devices are going to be evaluated by a group with technical skills and mostly advanced knowledge in computer graphics. All perceived visual and technical saliences are recorded in order to evaluate the correctness and the quality of the devices and the constraints.

1 Introduction

Many of today's electronic systems produce and process a massive amount of multimedia data like video, audio, position information etc. Production systems observe the workflows with cameras and scanning barcodes to optimize, to verify and to track the delivery services of the products. Surveillance systems monitor the traffic of cars to detect potential problems like traffic jam and to provide informations to advanced driver assistance systems. Most of these systems receive, produce and send many different kind of data and often combine

them to one file or a group of files to create streams of multimedia data. At the same time the number of systems as well as their complexity rapidly grow. The standard video resolution further increases from Full HD to 4k and 8k UHD and beyond while the acoustical standards also make use of additional channels ranging from the well-known 5.1 surround sound to the Hamasaki 22.2 surround sound system. Additional capabilities like 3D, 360-degree, Virtual Reality as well as Augmented and Mixed Reality are also included and need to be addressed. Furthermore, the application domains expands from normal TV and computer screens to smartwatches, smartphones, huge projectors and systems showing artificial objects in 3D with the capability to extend real world image scenes.

In contrast to these trends, common studies of accessibility, correctness, performance, and especially quality are frequently performed by using media samples of small sizes which not seldom originate from the last century, produced in standard-television formats like NTSC, PAL or SECAM being accompanied by stereo sound (cf. to Fig. 1). They often contain recorded sequences from the real world to reproduce intrinsic characteristics and properties as well as artificial structures to invoke potential visual errors. However and due to their nature of a comparatively low-resolution of a past technology, results cannot be easily transferred to the new challenges described beforehand.

Regardless of the system, any stream of multimedia data can be described as a chain of various steps as shown in Fig. 2. Not all the steps need to be included into every system and some steps may repeatedly occur depending on the task of the system. However, each step has its own characteristics and a



Fig. 1. Samples of commonly used visual test data (a) RCA Indian-head test image (http://www.forensicgenealogy.info/images/bulova.indian_head_test_patt.jpg). (b) Lena test image (<http://sipi.usc.edu/database/download.php?vol=misc&img=4.2.04>). (c) Frame of the Flower test video (http://media.xiph.org/video/derf/y4m/flower_cif.y4m). (Color figure online).

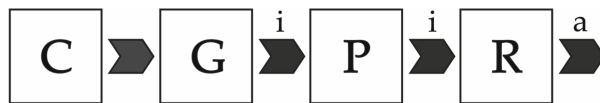


Fig. 2. Chain of multimedia data processing with “C, input from camera; G, grab image (digitize and store); P, preprocess; R, recognize (i, image data; a, abstract data)” [1].

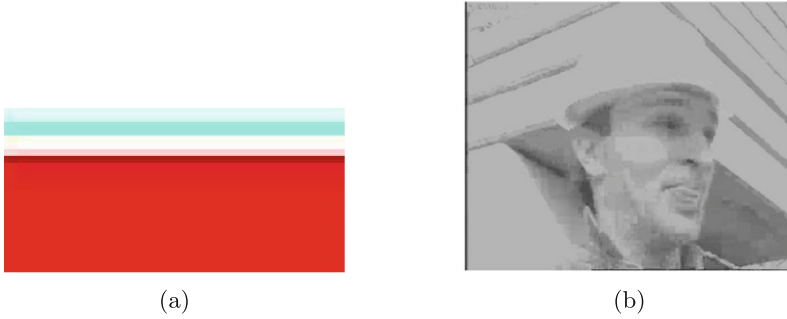


Fig. 3. Samples of common visual artefacts (a) Ringing artefacts at the transition from the red to the white part of the image. (b) Blocking artefacts as result of image compression [8]. (Color figure online).

potential to inflict errors. Such can be noticed for instance as visual artefacts shown in Fig. 3a and b or as clicks or disturbances in acoustical data. On the one hand the characteristics of each artefact and their rate of occurrence are strongly correlated to parameters like resolution, framerate, and color space, on the other hand they also depend on the implementation of the underlying transcoding system, their settings, and the data itself.

In order to reduce the artefacts and to optimize the quality of the multimedia data, various test patterns already exist. Each pattern can detect at least one specific kind of an artefact even though the total number of artefacts is innumerable. In addition, some artefacts will not appear in a single test pattern. Thus, combined and more complex patterns are needed. They commonly appear, for instance, in rapid changes of the image content, movements or image transformations like rotations or translations. Generally, testsets often need to be hand-crafted to cause the anticipated error and make them abundantly clear visible. For example, a minor color error in one of the flowers of Fig. 1c may occur but almost appear nearly invisible since the contrast to the surrounding is too small or too big, in contrast to a image with larger unicolored planes. In some fields like image retrieval, digital archiving [4] or image understanding [5] additional constraints like size or resolution are important to minimize the overall time of the test. On the other hand, many changes of fundamental properties like the aspect ratio, affects the effectiveness of the test and therefore a new test must be created. *Manthey et al.* [3] develops a highly flexible system to create synthetic testsets as independent as possible to overcome that problem and show its use with a short evaluation of visual data with the commonly used video encoding systems FFmpeg¹, Adobe Media Encoder CC 2015.0.1(7.2)², and Telestream Episode 6.4.6³. Some results present artefacts as in Figs. 4 and 5.

¹ <https://ffmpeg.org/>.

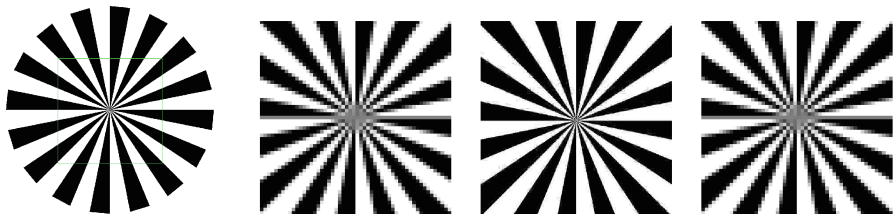
² <http://www.adobe.com/de/products/media-encoder.html>.

³ <http://www.telestream.net/episode/overview.htm>.



(a) Original pattern of the test case rotating around the center of the image. (b) Result frame showing grain like random artefacts. (c) Result frame showing randomly grouped artefacts.

Fig. 4. The test video with rotating stripes in Fig. 4a is compressed with FFmpeg showing heavily disintegrated content in Fig. 4b and c (Color figure online).



(a) Original unrotating Siemens star pattern with green cutout marking. (b) Cutout from result frame 1. (c) Cutout from result frame 7. (d) Cutout from result frame 15.

Fig. 5. The test video without any change in Fig. 5a is compressed with FFmpeg and results in a sequence of frames with strongly changing visual quality.

In the field of virtual reality systems the quality and the properties can be different for each of both eyes as shown in Fig. 6. Consequently, the amount of examinations increases at least by a factor of two and represents an additional constraint to the testset. The studies of *Kreylos* [2] and *Tate* [6] which use traditional testsets like checkerboards and grids to measure the distortion of the lenses and the chromatic aberration as in Fig. 7a as well as the field-of-view in Fig. 7b. Also the perspective, motion and occlusion have to be taken into consideration.

The remainder of this paper is organized as follows: Sect. 2 gives an overview about the structure and the workflow of the creation of our device-independent testset. Section 3 describes the exploratory comparison of the virtual reality devices with our testset and Sect. 4 present the results. A brief summary and an outlook into future work is given in Sect. 5.

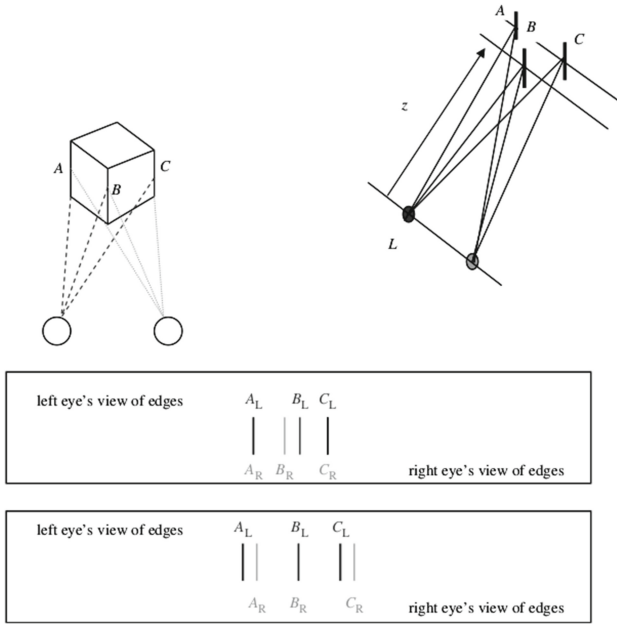


Fig. 6. Scheme of a cube viewed by two eyes. Showing the different position of the edges A , B and C in the visual field of the left and the right eye allowing the calculation of depth [7].

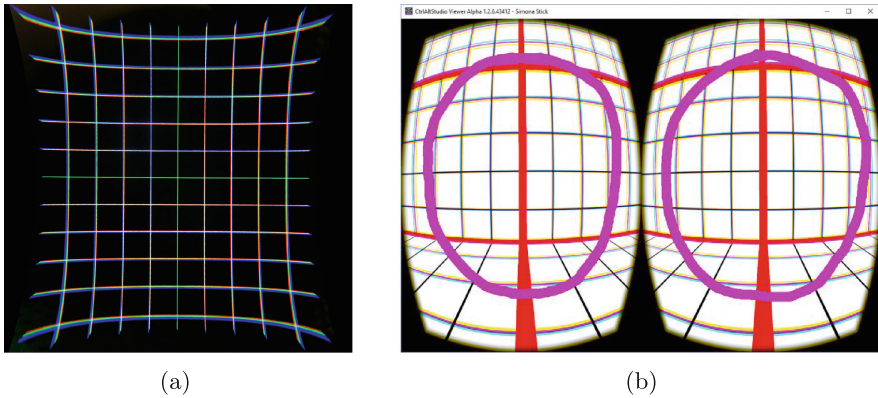


Fig. 7. Examples of traditional testsets used in the field of virtual reality systems (a) Example of the distortion of a lens and chromatic aberration near the periphery [2]. (b) Example of the field-of-view (oval) of the left and right eye in a virtual reality system with two displays [6].

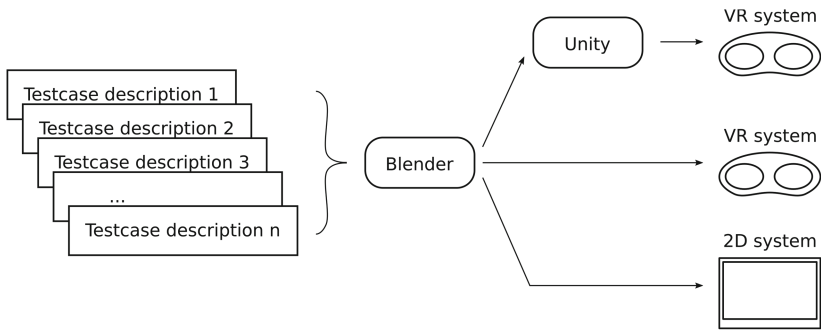


Fig. 8. Schematic view of the process of the generation and the application of the testsets. Descriptions of the testcases be combined into a testset to be applied by Blender to provide them to the designated virtual reality device or monitoring 2D devices. In some cases the transfer to Unity is necessary to provide the testset to the virtual reality device.

2 System Architecture and Workflow

To generate testsets that are able to cover the given constraints in a flexible and adaptable way, we decided to describe them in an abstract, vectorized and device-independent form following the experience from *Manthey et al.* [3]. Each element of a testcase is defined by the shape of the structure, the color, the position and properties of the movement as shown in Fig. 9, aside of affine transformations like translation, rotation, scaling, shearing and reflection of the base elements. In that way a 3D scene is constructed with one or multiple grouped elements in order to build complex test cases.

We use the build-in Blender/Python API⁴ to realize the description as the first step shown in Fig. 8. A second step comprises the selection of a subset of all the testcases to create a testset which is afterwards applied to the designated device. If another tool like the cross-platform game-engine Unity⁵ is needed to use devices like HTC Vive⁶, Oculus Rift⁷ or Android-based smartphones, the testset is exported and executed locally. Other devices like the *Zeiss Cinemizer OLED*⁸ or simple 2D displays can be directly operated and rendered by Blender. In each case the settings which happens to be more device-specific like the size of the test object, resolution, framerate etc. are set by the current tool, for instance Blender or Unity if necessary. The result is send to the designated device and the test is realized. The comparison of the given data from the generator and the presented visual data allows an inference of the performance, the quality as well as the constraints of the tested devices.

⁴ https://docs.blender.org/api/blender_python_api_2.78a_release/.

⁵ <https://unity3d.com/de>.

⁶ <https://www.vive.com/de>.

⁷ <https://www.oculus.com>.

⁸ <http://cinemizeroled.com>.

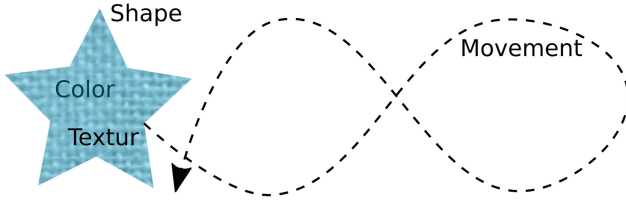
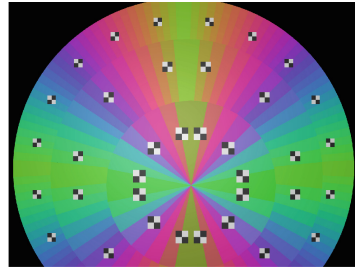
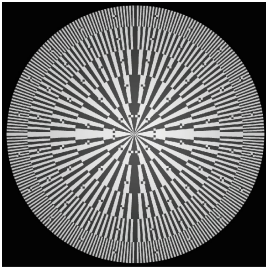
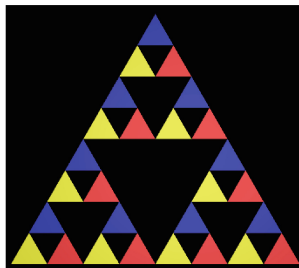


Fig. 9. Schematic view of the definition of a element being part of a testcase consists of the shape, color, texture and movement.



(a) The Siemens-Star contain a group of black and white equal sized stripes filling a quadrant of the whole circle and the rest by mirroring. In this example, we form a 3D object by stacking together five different cylinders with a radius of one to five units and a height of one unit similar to a pyramid. The texture of stripes of size-adjusted quantity is put on the surface as well as small squares with checkerboard structure.

(b) The Rainbow-Star object in this sample is constructed in a similar way as the Siemens-Star, but with a single color applied to each stripe of the generator quadrant and with a stack of only three cylinders.



(c) This plane pattern is based on a Sierpinski-triangle with red, green and blue triangles of equal size.

Fig. 10. 2D view of some examples of the testcases being part of the testset used in the comparison (Color figure online).

3 Exploratory Comparison

In order to realize the comparison we create a group of testcases. They contain circles and cylinders with black and white and with colored stripes like the samples shown in Fig. 10a and b. Further test cases consist of similarly colored, parallel tubular frames of equal length and diameter. As illustrated in Fig. 10c, some are constructed as Sierpinski-triangle and Sierpinski-carpet with fixed red, green, blue and yellow colored elements respectively. Each version is implemented without movement and with one of the following movements represented in Fig. 11. Move along a sinus-shaped curve, along a circle with one and five units radius through the zero point of the scale as well as orbiting that point, and along a rectangle of one and five units length. One additional movement realizes the circling of the scene camera representing the position of the virtual reality device in the virtual reality world like lunar orbit around the earth.

The test cases are created and deployed to each of the virtual reality devices with a resolution of HD or the closest possible depending on the device and with 24 bit color depth and 25 frames per second. Afterwards the set is presented to our exploratory group using a HTC Vive and a Zeiss Cinemizer OLED respectively. The group consists of five persons in the age between 20 and 40 years with technical skills with advanced knowledge in computer graphics.

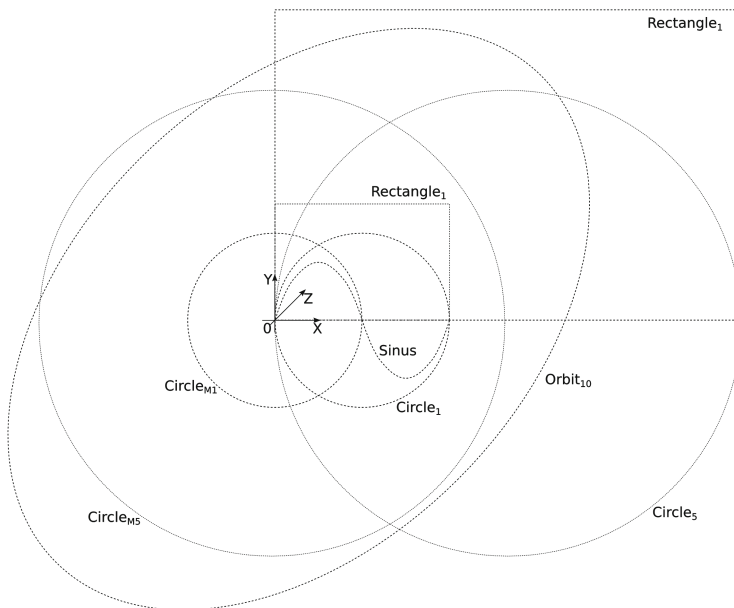


Fig. 11. Scheme of the realized sequences of movements following a sinus-shaped curve, circles of different diameters and centers of rotation, rectangles of different size and an orbit surrounding the center of the virtual reality scene.

Any visual artefact perceived by the participants is registered and a picture is taken by a Canon IXUS 980 IS digital camera at the position of the eye of the perceiving participant. This is compared with the deployed testset and the presentation at the 2D device to get a better isolation of the reason. For each artefact, a subjective estimation is given by each participant representing its relevance with rating from 1 (insignificant, i.e. less important) to 5 (severe, i.e. heavily affects the quality of perception). Finally, all ratings from the group members of each artefact are averaged to get an overall rating.

4 Results

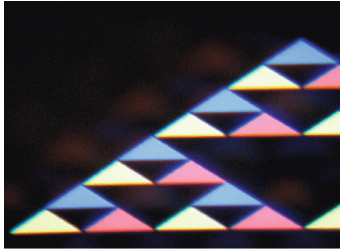
After the deployment and the presentation of the testset to our exploratory group the visually perceived artefacts and their ratings are taken into account to select the most salient as well as the strongest artefacts from the total amount. In a similar way the rating of the testcases are processed.

As a result, our comparison shows that the biggest influence of the visual quality of all the testcases is represented by the hardware of the virtual reality devices, mostly as expectable mostly depending on resolution and the quality of the incorporated hardware. Furthermore, a lowering of the quality of the rendering system can result in visual artefacts but also in the introduction of new abnormalities, especially during movements. However, a selection of the best depictable artefacts are shown in Fig. 12.

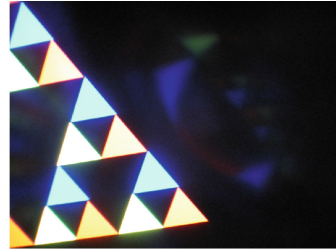
We found that especially the testcases with high contrast between their elements and surrounding objects appear as reasonable indications for artefacts mostly caused by the lenses. Combined with the movements of the test objects, they become salient and easier recognizable. In general, the Siemens-Star and the Sierpinski-triangle tend to create shadow-alike structures and reflections as shown in Fig. 12a, b, and e presumably caused by the structure of the lenses and their position in relation to the main light source of the scene. Some features create regular spatial recurring errors as shown in Fig. 12f which are induced by the Fresnel lenses of the devices and moire patterns (cmp. to Fig. 12d). Testcases with lower contrast like in Fig. 12c amplify a blurring of the transitions of the borders of colored areas presumably caused by the low resolutions of the virtual reality devices. Errors like in Fig. 12g are independent from the content but appear in our instance of the devices as a component fault.

With the implementation of the different movements a reproducible observation of objects containing the testcases is enabled. This facilitates the detection of some artefacts since they are emphasized by the dynamic changes.

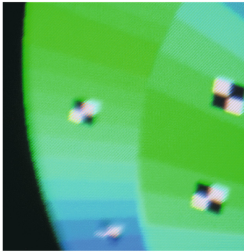
The results of the comparison as well as the subjective impressions given by the participants show that the HTC Vive creates a good and elaborated immersion into virtual reality at the price of lower resolution and more visual artefacts with stronger manifestation which are covered by intrinsic actions of movement in the 3D environment, especially in fast-paced games. The Zeiss Cinemizer OLED performs a better visual realization with mostly higher quality but a lower clarity of the virtual reality.



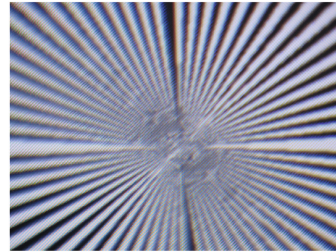
(a) A Sierpinski-triangle produces artefacts of left-shifted colored shadows of the elements.



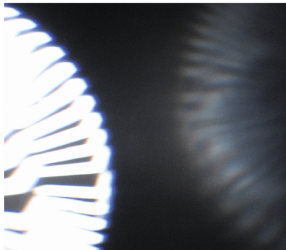
(b) Moving (a) along a predefined path enhances the perceptibility of artefacts yielding to a curved reflection on right.



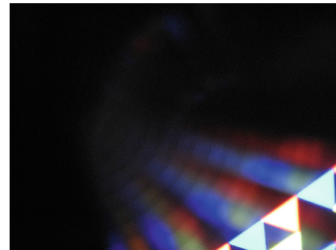
(c) A Rainbow-Star demonstrates the blurring artefact at the borders of the areas.



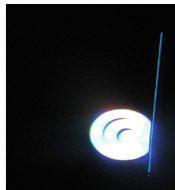
(d) With a Siemens-Star, moiré effects and the pixel grid becomes visible, especially in the upper left area.



(e) During movement the reflection of a Siemens-Star becomes indistinct but visible because of the high contrast to the background, whereas at some positions distinct perceptions are perceivable.



(f) At an particular position during the movement of a Sierpinski-triangle and their shadows the rings of the Fresnel lens become clearly visible.



(g) A defect in the construction of our Zeiss Cinemizer shows a content independent bright bar on right of the left display.

Fig. 12. Pictures of best depictable visual artefacts caused by the testset (Color figure online).

5 Summary and Future Work

In conclusion, we demonstrated the use of testcases based on abstract device-independent descriptions of objects and movements in virtual reality scenes. They are generated and deployed to different virtual reality devices being observed by our exploratory group in order to compare the two virtual reality devices and to estimate the usefulness of generated testsets as well as the generation process. The observed visual artefacts demonstrate the properness of the approach and its potential. Especially future development and integration of automatic image capturing devices strikingly increases the capabilities of quality measuring and assurance of the devices and their components like lenses and displays.

Acknowledgments. This work was partially accomplished within the project *localizeIT* (funding code 03IPT608X) funded by the *Federal Ministry of Education and Research* (BMBF, Germany) in the program of *Entrepreneurial Regions InnoProfile-Transfer*.

References

1. Davies, E.: Machine Vision. Morgan Kaufmann, Amsterdam (2005)
2. Kreylos, O.: Optical Properties of Current VR HMDs. Technical report. <http://doc-ok.org/?p=1414>
3. Manthey, R., Conrad, S., Ritter, M.: A framework for generation of testsets for recent multimedia workflows. In: Antona, M., Stephanidis, C. (eds.) UAHCI 2016. LNCS, vol. 9739, pp. 460–467. Springer, Cham (2016). doi:[10.1007/978-3-319-40238-3_44](https://doi.org/10.1007/978-3-319-40238-3_44)
4. Manthey, R., Herms, R., Ritter, M., Storz, M., Eibl, M.: A support framework for automated video and multimedia workflows for production and archive. In: Yamamoto, S. (ed.) HIMI 2013. LNCS, vol. 8018, pp. 336–341. Springer, Heidelberg (2013). doi:[10.1007/978-3-642-39226-9_37](https://doi.org/10.1007/978-3-642-39226-9_37)
5. Ritter, M.: Optimization of algorithms for video analysis: a framework to fit the demands of local television stations. In: Wissenschaftliche Schriftenreihe Dissertationen der Medieninformatik, vol. 3, pp. 1–336. Universitätsverlag der Technischen Universität Chemnitz, Germany (2014). <http://nbn-resolving.de/urn:nbn:de:bsz:ch1-qucosa-133517>
6. Tate, A.: VRL and a community and test region for virtual reality in virtual worlds. Technical report, Artificial Intelligence Applications Institute, University of Edinburgh (July 2016), <http://blog.inf.ed.ac.uk/atate/2016/07/20/vrland-a-community-and-test-region-for-virtual-reality-in-virtual-worlds/>
7. Westheimer, G.: Three-dimensional displays and stereo vision. Proc. Roy. Soc. **278**, 2241–2248 (2011)
8. Wiegand, T., Sullivan, G.J., Bjntegaard, G., Luthra, A.: Overview of the H.264/AVC video coding standard. IEEE Trans. Circuits Syst. Video Technol. **13**(7), 560–576 (2003)