A Modular Scheme for Artifact Detection in Stereoscopic Omni-Directional Images

Sebastian Knorr, Simone Croci and Aljosa Smolic

School of Computer Science and Statistics
Trinity College Dublin, the University of Dublin, Ireland.

Abstract

With the release of new head-mounted displays (HMDs) and new omni-directional capture systems, 360-degree video is one of the latest and most powerful trends in immersive media, with an increasing potential for the next decades. However, especially creating 360-degree content in 3D is still an error-prone task with many limitations to overcome. This paper describes the critical aspects of 3D content creation for 360-degree video. In particular, conflicts of depth cues and binocular rivalry are reviewed in detail, as these cause eye fatigue, headache, and even nausea. Both the reasons for the appearance of the conflicts and how to detect some of these conflicts by objective image analysis methods are detailed in this paper. The latter is the main contribution of this paper and part of long-term research roadmap of the authors in order to provide a comprehensive framework for artifact detection and correction in 360-degree videos. Then, experimental results are demonstrating the performance of the proposed approaches in terms of objective measures and visual feedback. Finally, the paper concludes with a discussion and future work.

Keywords: 360-degree video, omni-directional images, 3D quality assessment, binocular rivalry, conflicts of depth cues

1 Introduction

360-degree video, also called live-action virtual reality (VR), is one of the latest and most powerful trends in immersive media, with an increasing potential for the next decades. In particular, head-mounted display (HMD) technology like e.g. HTC Vive, Oculus Rift and Samsung Gear VR is maturing and entering professional and consumer markets. On the other side, capture devices like e.g. Facebook's Surround 360 camera, Nokia Ozo and Google Odyssee are some of the latest technologies to capture 360-degree video in stereoscopic 3D (S3D).

Figure 1: Example of overlapping errors

However, capturing 360-degree videos is not an easy task as there are many physical limitations which need to be overcome, especially for capturing and post-processing in S3D. In general, such limitations result in artifacts which cause visual discomfort when watching the content with a HMD. The artifacts or issues can be divided into three categories: binocular rivalry issues, conflicts of depth cues and artifacts which occur in both monocular and stereoscopic 360-degree content production (see Section 2 for further details). Issues of the first two categories have been investigated for standard S3D content e.g. for cinema screens and 3D-TV [Knorr et al., 2012], [Vatolin et al., 2016], [Lambooij et al., 2009]. The third category consists of typical artifacts which only occur in multi-camera systems used for panorama capturing. As native S3D 360-degree video production is still very error-prone, especially with respect to binocular rivalry issues, many high-end S3D productions are shot in 2D 360-degree and post-converted to S3D.

This paper is dealing with automatic artifact detection in omni-directional images (ODIs) for quality control within the post-production workflow. To our knowledge, there is no scientific publication in this area and thus an open research field of high importance. Currently, the post-production workflow basically consists of six steps: 1) data ingest, 2) rough stitching of camera views (automatically), 3) fine stitching (manually), 4) color-grading, 5) editing and 6) finishing (rendering). Especially, the fine stitching process, which includes removal of stitching and blending artifacts as well as wire-, rig-, shadow- and contamination removal, is a labor intensive process with many intermediate rendering steps in order to check the quality of the results on HMDs. It is our goal to provide algorithms and tools for automatic detection and, if possible, correction of artifacts in order to give automatic feedback to artists and reduce time and efforts in post.
The paper is structured as follows. In Section 2, the state of the art is reviewed, in particular the technical challenges of 360-degree capturing and the resulting artifacts and issues. Then, in Section 3, we describe the proposed modular system and approaches for color mismatch and geometrical misalignment detection, which is the main contribution of this paper and part of our long-term research roadmap for quality control in 360-degree videos. In Section 4, experimental results for a test dataset of 13 ODIs captured with 7 different 360 capture devices (Google Odyssee, Jaunt, OmniCam-3D, Ozo, Panocam, Surround-360 and a self-constructed system with Mobius cameras) are demonstrating the performance of the proposed approaches in terms of objective measures and visual feedback. Finally, the paper concludes with a discussion and future work in Section 5.

2 State of the Art

Omni-directional image stitching is a process of synthesizing multiple views together on a common virtual surface. The overlapping regions between the cameras are first matched using different planar transformation models (e.g. affine, perspective or cubic transformation models), and the transition between the overlapping parts is estimated via blending parameters. Then, the views are blended and warped onto the omni-directional surface using the estimated geometric relation between the omni-directional surface and the image coordinates.
ODI stitching is challenging as the capturing devices have some inevitable drawbacks, e.g. the optical centers of the individual cameras do not share the same center of projection. However, applying planar transformation models in order to synthesize multiple views together on a common virtual surface is only valid if the captured scene is a planar surface itself or the cameras share the same center of projection [Hartley and Zisserman, 2003].

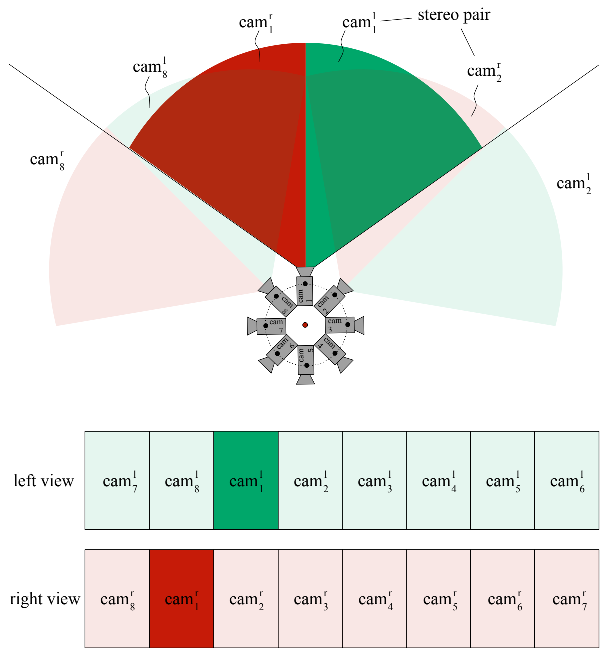

Figure 2: Principles of an off-centered slit camera model for capturing 360-degree in S3D

For off-centered cameras, transformation errors occur which increase with the off-center distance and the amount of depth within the captured scene. Figure 1 shows exemplary the viewport of an ODI with a stitching and blending error caused by a cross-fading in the overlapping area of two adjacent cameras.
In order to reduce stitching errors, the baseline between the cameras should be minimized. On the other side, the baseline between the cameras of different views needs to be increased for S3D content creation as parallax is required for generating a 3D effect, i.e. stitching and blending errors actually increase in S3D 360-degree content.
Figure 2 shows the principles of an off-centered slit camera model for capturing 360-degree in S3D. Half of the field of view of each camera is dedicated to either the left view of a stereoscopic ODI and the other half

is dedicated to the right view of a stereoscopic ODI. At this point, it needs to be mentioned that two adjacent cameras must share at least 50% of their fields of view as the right half of one camera view (e.g. cam_1^l) and the left half of the adjacent view (e.g. cam_2^r) form a stereoscopic image pair as illustrated in Figure 2.

2.1 Common artifacts

Whether native S3D or conversion from 2D to stereoscopic 3D, they both must display a high technical quality, while also taking all aspects of the human binocular visual system into account in order to reduce visual discomfort [Knorr et al., 2012]. This is even more essential, the higher the degree of immersion is, which is the case for VR by using HMDs. Artifacts arising from improper camera alignments, physical limitations of the chosen capture system, errors in post-production or compositing, etc. are still inevitable and can be categorized into three categories: binocular rivalry issues, conflicts of depth cues and artifacts which occur in both monocular and stereoscopic 360-degree content production.

Table 1 gives an overview of the first category. Most of the issues only occur in native S3D productions while issues of the second category (see Table 2) mainly appear in 2D to S3D conversion. Finally, Table 3 details issues which can occur in both 2D and S3D production as well as in native S3D or 2D to S3D conversion.

Artifact/ Issue	**Characteristics**	**Caused by**
Geometrical misalignment	Improper (vertical) alignment of left and right images	Cameras or lenses not properly aligned Tilting head or changing yaw while looking at the pole caps with a HMD
Luminance/ colorimetry	Difference in hue, saturation and/or intensity between left and right image	Cameras not properly matched (e.g. different aperture) Varying lighting conditions at different camera locations
Visual mismatch	Reflections, lens flares, polarization Contamination Missing or different objects in one of the views	Varying lighting conditions at different camera locations Contamination due to environmental conditions (e.g. rain, dust, etc.) Compositing errors in post
Depth of field/ sharpness mismatch	Difference in sharpness or depth of field	Different aperture settings of cameras Focal length of cameras not properly matched
Synchronization	Left and right image sequences are not synchronized	Cameras are not synchronized/ gen-locked Editing errors in post
Hyperconvergence/ hyperdivergence	Objects are too close to or too far from the viewer's eyes	Too much negative or positive parallax between left and right image
Pseudo-3D	Left and right images are swapped	Swapped images in HMDs Editing error in post
Ghosting	Double edges of objects	Stitching and blending artifacts in post

Table 1: Binocular rivalry issues in stereoscopic 360-degree videos

2.2 Quality assessment

Over the last years, many publications focused on the assessment of 3D quality in terms of subjective and objective quality metrics. In [Khaustova et al., 2015], the authors investigated how viewer annoyance depends on various technical parameters such as vertical disparity, rotation and field-of-view mismatches as well as color and luminance mismatches between the views. [Chen et al., 2014] proposed several objective metrics for luminance mismatch and evaluated their correlation with the results of subjective experiments. In [Goldmann et al., 2010], an artifact specific to S3D video is analyzed in depth by evaluating visual discomfort caused by temporal asynchrony. Finally, in [Battisti et al., 2015], a full-reference metric is presented by evaluation of a large variety of measures by taking 2D picture quality, binocular rivalry and depth map degradation into account. The authors maximized the correlation with the mean opinion score (MOS) by using linear regression. However, none of the 3D quality assessment approaches in the literature deal with ODIs.

In this paper, however, the focus lies on artifact detection in order to support the artist by giving direct quality feedback during post-production, in particular for geometrical misalignment and color mismatch. Thus, full-reference objective quality metrics can not be applied in this application. The authors of [Dong et al., 2013] propose a stereo camera distortion detecting method based on statistical models in order to detect vertical misalignment, camera rotation, unsynchronized zooming, and color misalignment in native S3D content.

Depth conflict	Characteristics	Caused by
Vergence vs. accommodation	Eyes accommodate on screen plane but converge or diverge on objects in front or behind the screen plane	Parallax between objects in the left and right view
Stereopsis vs. interposition	Foreground objects are occluded by background objects	3D compositing errors in post
Accommodation vs. depth of field	Eyes accommodate on screen plane but scene or part of scene is out of focus	Wide aperture of cameras
Stereopsis vs. (aerial) perspective	Monocular depth cue "perspective" or "aerial perspective" does not match with binocular depth cue "stereopsis"	3D compositing errors in post
Stereopsis vs. motion parallax	Motion of objects does not match with their distance	3D compositing errors in post
Stereopsis vs. size	Relative or familiar size of objects does not match with their distance	3D compositing errors in post
Stereopsis vs. light and shading	Distance or shape of objects does not match with their shadings	3D compositing errors in post
Stereopsis vs. texture gradient	Texture gradients are not in line with the descending of depth in the scene	3D compositing errors in post

Table 2: Depth conflicts in stereoscopic 360-degree videos

Artifact/ Issue	Characteristics	Caused by
Stitching artifacts	Visible seams and misaligned/ broken edges	Improper camera arrangement Registration and alignment errors in post
Blending artifacts	Visible color- and luminance mismatches of regions within an ODI	Varying lighting conditions at different camera locations Compositing errors in post
Warping artifacts	Visible deformations of objects	Improper camera arrangement Registration and alignment errors in post
Wobbling artifacts	Unsteady scene appearance over time	Temporal inconsistent stitching of camera views (non-stabilized image sequences)

Table 3: Artifacts in both monocular and stereoscopic 360-degree videos

[Voronov et al., 2013] introduce a large variety of artifact detection methods, including color mismatch and vertical disparity. With respect to the color mismatch approach, the RGB color space is used which, in our application, is inappropriate as the HSV color space is usually preferred in post-production.

3 Proposed Approach

Figure 3 shows an overview of the proposed modular system to detect color mismatch and vertical misalignment as part of an overall framework for detecting and correcting artifacts outlined in Subsection 2.1. All of the underlying algorithms are implemented in OFX and thus useable as plugins in professional post-production applications like Nuke, Fusion, Mamba FX or Natron (see Figure 4). The following subsections describe the underlying methods in more detail.

3.1 Geometrical misalignment

Geometrical misalignment, in particular vertical parallax, is present when objects in the scene are not vertically aligned between the left and right stereo images. In order to detect vertical parallax, we first compute sparse

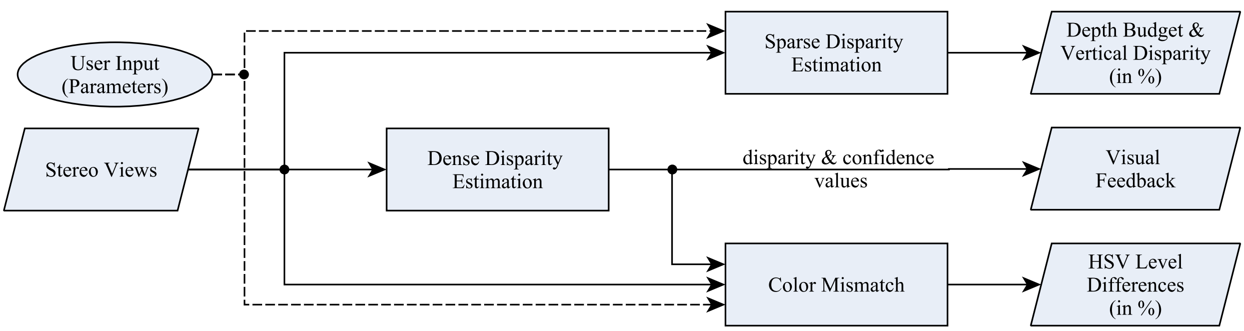

Figure 3: Overview of proposed system to detect color mismatch and vertical misalignment

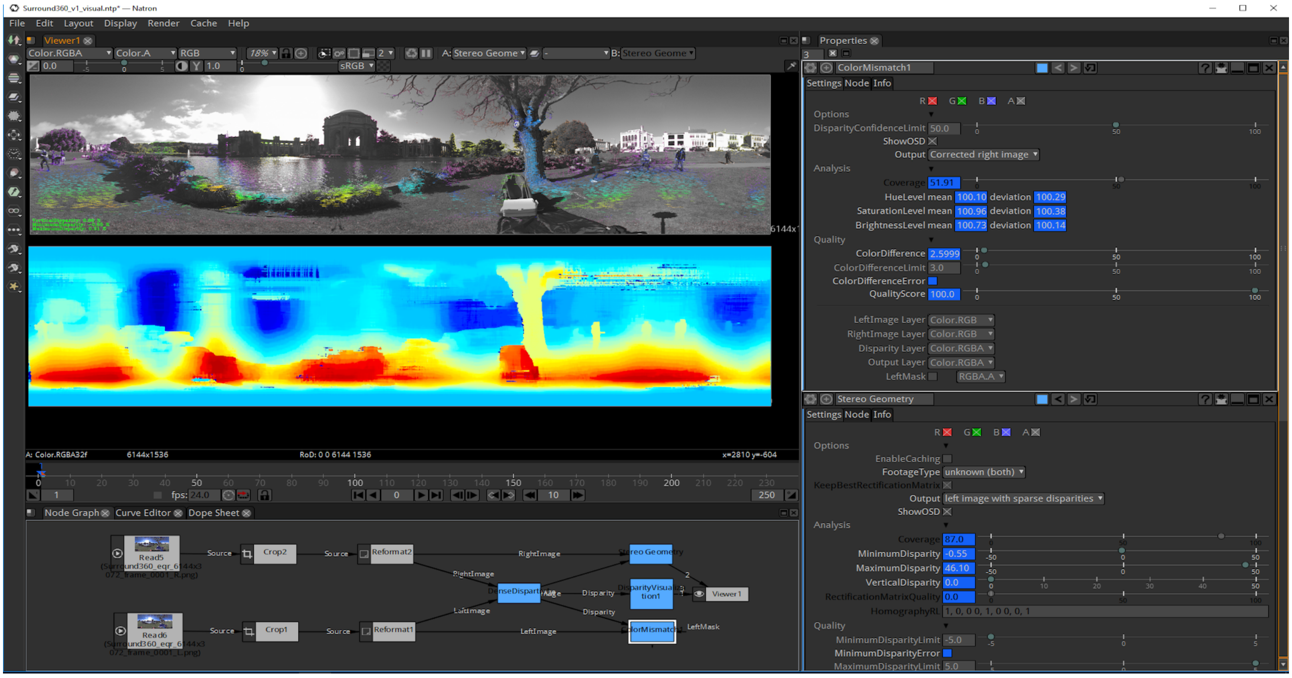

Figure 4: Screenshot of Natron including the node graph with the modules, visualization of sparse and dense disparities, and user input and output parameters for the *lake* ODI captured with the Surround-360 rig

disparities between left and right view. First, distinctive features are extracted in both stereo images using the SURF feature point detector and descriptor SURF [Bay et al., 2006]. Then, for each extracted feature point a descriptor is computed that represents the local properties of the image around the feature point.

The feature points are then matched between the two stereo images as follows. Each feature point in the left image is compared to the feature points in the reference image by calculating the Euclidean distance between their descriptor vectors. A matching pair is detected if its distance is closer than 0.7 times the distance of the second nearest neighbour [Bay et al., 2006]. Unreliable matches are eliminated with RANSAC [Fischler and Bolles, 1981] using the epipolar geometry as validation model. Finally, the vertical component of the disparities of all remaining matches is computed and displayed as percentage of the image width.

3.2 Color mismatch

The color mismatch module compares the color properties between the two stereo images. In the first step, pixels which are present in both the left and right images are detected using the semi-global matching approach for dense disparity estimation as described in [Hirschmuller, 2008]. The confidence maps of the resulting left-to-right and right-to-left disparity maps can be thresholded in order to take only higher reliable disparities into account. In the experimental results in Section 4, we applied a threshold of 50%. For all corresponding pixels above the threshold, the color properties, i.e. mean and standard deviation of the color channels, are computed for the left and right images, respectively, as introduced by [Reinhard et al., 2001]. Instead of using the $l\alpha\beta$ color space, as proposed by the authors, we extract the same statistics but in the HSV color space for reasons mentioned above.

With Ω_l and Ω_r as the sets of corresponding pixels in the left image I_l and right image I_r, and with $I_l(p)$ and $I_r(p)$ as the colors at the pixel p defined in the HSV color space, the means for each channel are defined as

$$\mu_x = \frac{1}{|\Omega_x|} \sum_{p \in \Omega_x} I_x(p), \quad x \in \{l, r\} \tag{1}$$

and the standard deviations for each color channel as

$$\sigma_x = \sqrt{\frac{1}{|\Omega_x|} \sum_{p \in \Omega_x} (I_x(p) - \mu_x)^2}, \quad x \in \{l, r\}. \tag{2}$$

For comparison, we take the left image as reference and compute the mean difference $(\mu_r - \mu_l)/\mu_l$ and the standard deviation difference $(\sigma_r - \sigma_l)/\sigma_l$. Finally, the determined statistics can be used for color transfer and thus applied in order to correct the right or left image, respectively.

4 Experimental Results

In our test scenario, we chose exemplary a dataset consisting of 13 equirectangular stereo ODIs captured with 7 different 360 capture devices (Google Odyssee, Jaunt, OmniCam-3D, Nokia Ozo, Panocam, Facebook's Surround-360 and a self-constructed system with Mobius cameras by Jim Waters[1]) and analyzed the images with respect to geometrical misalignment and color mismatch as proposed in Section 3. As the results will also heavily depend on the captured content and the amount of post-production efforts, they can only be seen as an indiction for the characteristics of the camera systems under test.
In order to have similar conditions for all images, we only selected a vertical field of view of 60 degree instead of 180 degree for the equirectangular ODIs for two reasons: 1) Some of the images do either have no pole caps (like Google Odyssee and OmniCam-3D) or the rig at the nadir has not been removed (Panocam images) and 2) The zenith and nadir have a large degree of distortion within the equirectangular ODI (e.g. the first row represents a single pixel in the sphere).

4.1 Geometrical misalignment

Figure 5 illustrates the vertical parallax for each of the input images in percentage of the image width. It should be noted at this point, that here the image width was chosen to one fourth of the original image width as most of the HMDs only have a horizontal field of view of about 90 degrees.
The results show that the ODIs captured with the Panocam have by far the largest geometrical misalignments with up to 1.2% vertical parallax for the *corridor* sequence. In relation to S3D cinema movies, where the vertical parallax ranges between 0 and 0,14% according to the study in [Vatolin et al., 2016], most of the geometrical misalignments are extremely high, and thus would cause a high degree of visual discomfort.

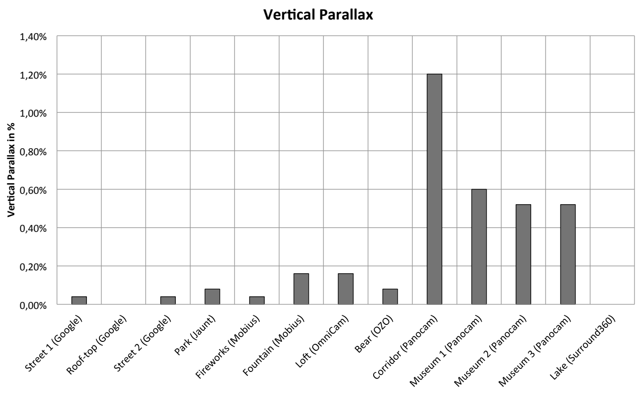

Figure 5: Vertical Parallax

4.2 Color mismatch

Figure 6 illustrates the color mismatches of the right images relative to the left images in terms of mean differences and standard deviation differences for each channel in HSV color space in percentage. The results show that the ODIs captured with the OmniCam have by far the larges color differences with a mean difference of brightness of 19.77%, followed by the *corridor* sequence (Panocam) and the self-constructed rigs with Mobius cameras. Most of the ODIs do only have a minor color mismatch between left and right view. However, we noticed visible color mismatches when applying a toggle view between the left and right images.
The reason for this discrepancy is quite obvious. While the OmniCam, Panocam and Mobius rig capture left and right ODIs independently, color differences between the views are inevitable as described in Section 2. Furthermore, the OmniCam is actually a mirror rig and thus, it is more error-prone to lighting conditions at different viewpoints. All the other ODIs were captured with an off-centered slit camera system as illustrated in Figure 2, i.e. left and right ODIs are captured with the same cameras. Although the overall color differences

[1] `https://photocreations.ca/3D/mobius_camera_rig.html`

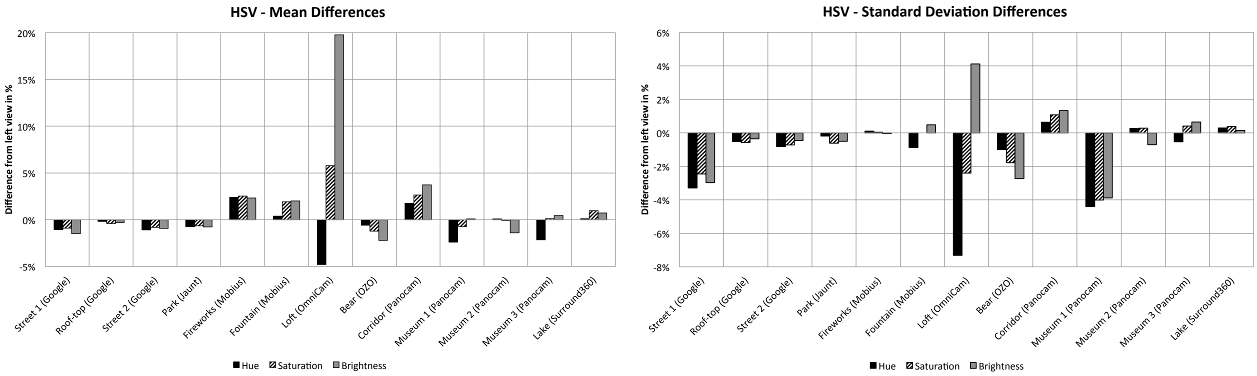

Figure 6: Color difference (mean and standard deviation) of the right image compared to the left image

for these cameras are small, the color differences between left and right images within a viewport of an HMD are still existing as each stereo pair of the input images belong to different cameras.

5 Conclusion and future work

The paper described a modular system for artifact detection in stereoscopic omni-directional images within the post-production, in particular the binocular rivalry issues: geometrical misalignment and color mismatch. The algorithms were exemplary applied to a dataset of 13 stereoscopic ODIs captured with 7 different 360-degree camera rigs. From the results, we can derive a couple of interesting properties and further challenges which need to be addressed. First, vertical parallax seems to be a serious issue in stereoscopic ODIs as it is much larger than in standard 3D cinema or 3DTV footage. Furthermore, capture devices which use independent stereo camera pairs tend to have an even larger vertical parallax. We have to note again, however, that we do not have any knowledge about the amount of post-production efforts spent for each of the ODIs under evaluation.
Secondly, global color mismatches between left and right ODI seem to be relatively small, except for the capture devices which use independent stereo camera pairs like OmniCam-3D, Panocam and the Mobius rig. However, visual inspection of the left and right ODIs often show significant local color mismatches. Thus, more investigation of local color mismatches, i.e. within the viewports is necessary as this is the area which the user actually sees and where binocular rivalry needs to be measured. The main challenges, however, are heavy distortions within each ODI of a stereo pair, in particular stitching, blending and warping artifacts which heavily degrade the dense disparity estimation results, and which affect the subsequent detection modules. One could argue, however, that the modules should support an artist within the fine-stitching process in post. If the distortions are too extreme for confident dense disparity estimation, the artist would probably notice this even without the support of the detection modules and would try to fix it.
The natural evolution of the current work is the extension of the artifact detection tools to the view adapted temporal dimension, i.e. the analysis of the viewports in stereoscopic 360-degree images and videos by also considering saliency as introduced in [Ana De Abreu, Cagri Ozcinar, 2017]. Furthermore, we will extend the system for the detection and possibly correction of other artifacts like e.g. sharpness mismatch. The modules for visual mismatch and pseudo-3D detection are already implemented and under evaluation. Finally, we want to motivate researchers from other research institutes to also focus on this research area in order to improve the quality of 360-degree videos.

Acknowledgments

The authors would like to thank Lutz Goldmann, Sebastian Schmiedecke and Ronald Kluth for their work on 3D quality control plugins.

References

[Ana De Abreu, Cagri Ozcinar, 2017] Ana De Abreu, Cagri Ozcinar, A. S. (2017). Look around you: saliency maps for omnidirectional images in VR applictions. In *9th International Conference on Quality of Multimedia Experience (QoMEX)*.

[Battisti et al., 2015] Battisti, F., Carli, M., Stramacci, A., Boev, A., and Gotchev, A. (2015). A perceptual quality metric for high-definition stereoscopic 3D video. In *Image Processing: Algorithms and Systems XIII, 939916*.

[Bay et al., 2006] Bay, H., Tuytelaars, T., and Van Gool, L. (2006). SURF: Speeded up robust features. In *Lecture Notes in Computer Science*, volume 3951, pages 404–417.

[Chen et al., 2014] Chen, J., Zhou, J., Sun, J., and Bovik, A. C. (2014). Binocular mismatch induced by luminance discrepancies on stereoscopic images. In *IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6.

[Dong et al., 2013] Dong, Q., Zhou, T., Guo, Z., and Xiao, J. (2013). A stereo camera distortion detecting method for 3DTV video quality assessment. In *2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, pages 1–4.

[Fischler and Bolles, 1981] Fischler, M. and Bolles, R. (1981). RANdom SAmpling Consensus: a paradigm for model fitting with application to image analysis and automated cartography. *Commun. Assoc. Comp. Mach.*, 24:381–395.

[Goldmann et al., 2010] Goldmann, L., Lee, J. S., and Ebrahimi, T. (2010). Temporal synchronization in stereoscopic video: Influence on quality of experience and automatic asynchrony detection. *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, pages 3241–3244.

[Hartley and Zisserman, 2003] Hartley, R. and Zisserman, A. (2003). *Multiple View Geometry in Computer Vision*. Cambridge University Press.

[Hirschmuller, 2008] Hirschmuller, H. (2008). Stereo Processing by Semiglobal Matching and Mutual Information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2):328–341.

[Khaustova et al., 2015] Khaustova, D., Fournier, J., Wyckens, E., and Le Meur, O. (2015). An objective method for 3D quality prediction using visual annoyance and acceptability level.

[Knorr et al., 2012] Knorr, S., Ide, K., Kunter, M., and Sikora, T. (2012). The Avoidance of Visual Discomfort and Basic Rules for Producing "Good 3D" Pictures. *SMPTE Motion Imaging Journal*, 121(7):72–79.

[Lambooij et al., 2009] Lambooij, M., IJsselsteijn, W., Fortuin, M., and Heynderickx, I. (2009). Visual Discomfort and Visual Fatigue of Stereoscopic Displays: A Review. *Journal of Imaging Science and Technology*, 53(3):030201.

[Reinhard et al., 2001] Reinhard, E., Ashikhmin, M., Gooch, B., and Shirley, P. (2001). Color transfer between images. *IEEE Computer Graphics and Applications*, 21(5):34–41.

[Vatolin et al., 2016] Vatolin, D., Bokov, A., Erofeev, M., and Napadovsky, V. (2016). Trends in S3D-Movie Quality Evaluated on 105 Films Using 10 Metrics. *Proceedings of the SPIE, Stereoscopic Displays and Applications XXVII*, 2016(5):1–10.

[Voronov et al., 2013] Voronov, A., Vatolin, D., Sumin, D., Napadovsky, V., and Borisov, A. (2013). Methodology for stereoscopic motion-picture quality assessment. In *Proceedings of the SPIE, Stereoscopic Displays and Applications XXIV*, volume 8648.