

uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Tamboli, Roopak R., Kara, Peter A., Cserkaszky, Aron, Barsi, Attila, Martini, Maria G., Appina, Balasubramanyam, Channappayya, Sumohana S. and Jana, Soumya (2018) 3D objective quality assessment of light field video frames. In 2018 - 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON) . Piscataway, U.S. : Institute of Electrical and Electronics Engineers, Inc. ISSN (online) 2161-203X ISBN 9781538661253

<https://doi.org/10.1109/3DTV.2018.8478557>

3D OBJECTIVE QUALITY ASSESSMENT OF LIGHT FIELD VIDEO FRAMES

Roopak R. Tamboli^{†,‡}, Peter A. Kara[§], Aron Cserkaszkzy[†], Attila Barsi[†],
Maria G. Martini[§], Balasubramanyam Appina[†], Sumohana S. Channappayya[‡], Soumya Jana[‡]

[†]*Holografika, Budapest, Hungary*

Email: {r.tamboli, a.cserkaszkzy, a.barsi}@holografika.com

[‡]*Indian Institute of Technology Hyderabad, Kandi, Sangareddy, India*

Email: {ee13p0008, ee13m14p100001, sumohana, jana}@iith.ac.in

[§]*WMN Research Group, Kingston University, Kingston upon Thames, UK*

Email: {p.kara, m.martini}@kingston.ac.uk

ABSTRACT

With the rapid advances in light field displays and cameras, research in light field content creation, visualization, coding and quality assessment is now beyond a state of emergence; it has already emerged and started attracting a significant part of the scientific community. The capability of light field displays to offer glasses-free 3D experience simultaneously for multiple users has opened new avenues in subjective and objective quality assessment of light field image content, and video is also becoming research target of such quality evaluation methods. Yet it needs to be stated that while static light field contents have evidently received relatively more attention, the research on light field video content still remains largely unexplored. In this paper, we present results of the objective quality assessment of key frames extracted from light field video content. To this end, we use our own full-reference 3D objective quality metric.

Index Terms — Light field, 3D visualization, 3D display, quality assessment, objective metric

1. INTRODUCTION

The recent advent of glasses-free light field displays has enabled life-like 3D visualization [1]. Such displays eliminate most of the unpleasant disturbances in the viewing experience associated with stereoscopic and auto-stereoscopic (lenticular and parallax-barrier) displays [2]. While providing an ultra-realistic experience due to the continuous motion parallax effect, light field displays provide new challenges in content acquisition, reconstruction, coding, communication and rendering [3] [4].

In general, each of the aforementioned steps dictate quality parameters and therefore the end user's Quality of Experience (QoE), in all kinds of visualization technologies, namely 2D, stereoscopic and auto-stereoscopic. Naturally, one attempts to assess the quality of the visualized content, which, in turn may be used as a feedback to the various processing stages. The research on the subjective and objective quality assessment of 2D and stereoscopic images and video content has progressed significantly in recent years. Well-established objective assessment methods have

nearly eliminated the need for subjective evaluation in common use cases of 2D and stereoscopic media [5]. In comparison, the light field visualization technology itself is in its so-called "infancy". A variety of implementations are now available for light field content generation and its visualization [6]. The new frontiers of light field visualization pose several research questions. Leveraging on the experience and expertise available from research on 2D and stereoscopic media, it may be beneficial to seek for hitherto unexplored aspects of light field visualization right in the beginning and accordingly devise efficient methods in the aforementioned light field media pipeline. Such an approach is immediately limited since the light field visualization technology is yet to be standardized and to become a commonplace. To this end, Cserkaszkzy *et al.* recently proposed a display-independent light field format [7].

One of the central aspects of any light field content visualization method which has been emphasized recently is its spatio-angular nature [8]. In the related research, subjective quality evaluations were conducted with several combinations of spatial and angular resolutions, a variety of rating methods, seated as well as freely moving subjects (within the valid field of view of the display), synthetic and real-life contents. However, nearly each and every research disseminated until now focuses on the assessment of still content.

This motivated us to investigate the following research questions: (a) Can a metric designed for still content be an efficient way to assess video content? (b) How do common visual degradations used for still image quality measurements apply to video frames with high temporal information and affect their objective assessments?

In this paper, we introduce our research on the visual quality assessment of light field video frames. As at the time of this paper no objective video quality metric exists for light field content, we applied our own objective metric for light field image quality to the selected video frames. The frames were chosen based on their values of temporal information, due to the evident dependency on the content.

The remainder of this paper is structured as follows: Section 2 overviews the related work in the area of light field visualization and details the state-of-the-art 2D and 3D objective methodologies of quality assessment. Section 3 introduces the methodology used in our research, with emphasis on the objective quality metric and the applied visual degradations. The investigated video content is described with its temporal information in Section 4, based on

The work in this paper was funded from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreements No 643072, Network QoE-Net and No 676401, European Training Network on Full Parallax Imaging.

which the video frames were selected. The results of the research are analyzed and discussed in Section 5. The paper is concluded in Section 6.

2. RELATED WORK

Due to the requirement for a large number of viewpoints, real-life light fields have primarily been captured for static scenes, using either lenticular arrays for a single sensor, a single camera densely scanning a static scene or multiple densely spaced cameras [9]. Existing research on light field video generally made use of publicly available multi-view video plus depth (MVD) data-sets [2] [10] [11]. Intermediate views, if required, were generated by view synthesis. Considering the associated synthesis artifacts, Battisti *et al.* proposed objective image quality assessment of 3D synthesized views [10]. Their suggested metric, called 3D Synthesized view Image Quality Metric (3DSwIM), applies to quality assessment of static views synthesized using depth image base rendering (DIBR). Ekmekcioglu *et al.* evaluated the quality of the synthesized video frames against the reference video frames using the existing 2D IQA metrics [12]. However, the subjective assessment was carried out only on a regular 2D display in their works. Tikanmaki *et al.* studied the quality assessment of 3D video in rate-allocation experiments on a lenticular auto-stereoscopic display [13]. Using the existing 2D metrics, they compared the views synthesized from uncompressed image and depth data with those synthesized from the corresponding compressed image and depth data. Liu *et al.* reported subjective and objective video quality assessment of 3D synthesized views under texture and depth compression [11].

Recently, Viola *et al.* reported objective and subjective evaluation of compression algorithms targeted for lenslet-based plenoptic camera images [14]. In a review of existing light field data-sets and associated challenges, Adhikarla *et al.* presented an evaluation, analysis, and extensions of existing quality metrics in the context of light fields [15]. Concerning the quality assessment of 3D content on a light field display, Dricot *et al.* reported the subjective evaluation of compressed (SMV) content on a horizontal-parallax light field display [2]. In particular, the effect of various compression schemes on perceived quality was investigated using classical image quality metrics.

3. METHODOLOGY

We now briefly explain the full-reference 3D objective quality metric used in this paper [16]. The metric, considering the spatio-angular nature of the light field content, evaluates the spatial and angular quality scores of a 3D perspective visualized on a light field display, and then pools them into a 3D quality score using a pooling parameter.

The spatial quality score, Q_{2D} involves steerable pyramid decomposition of each of the constituent image of a 3D view followed by fitting an univariate generalized Gaussian distribution (UGGD) on the coefficients. A feature vector corresponding to a 3D view is formed by stacking the parameters of UGGD for all the constituent images. Then, the spatial quality index Q_{2D} is the distance between a feature vector of a pristine 3D view and feature vector of a distorted 3D view, where each constituent image is distorted.

The angular quality score Q_{θ} finds structural similarity between optical flow images obtained for a pristine 3D view and a distorted 3D view. Optical flow values are calculated between successive images of a 3D view. The key idea is that differences be-

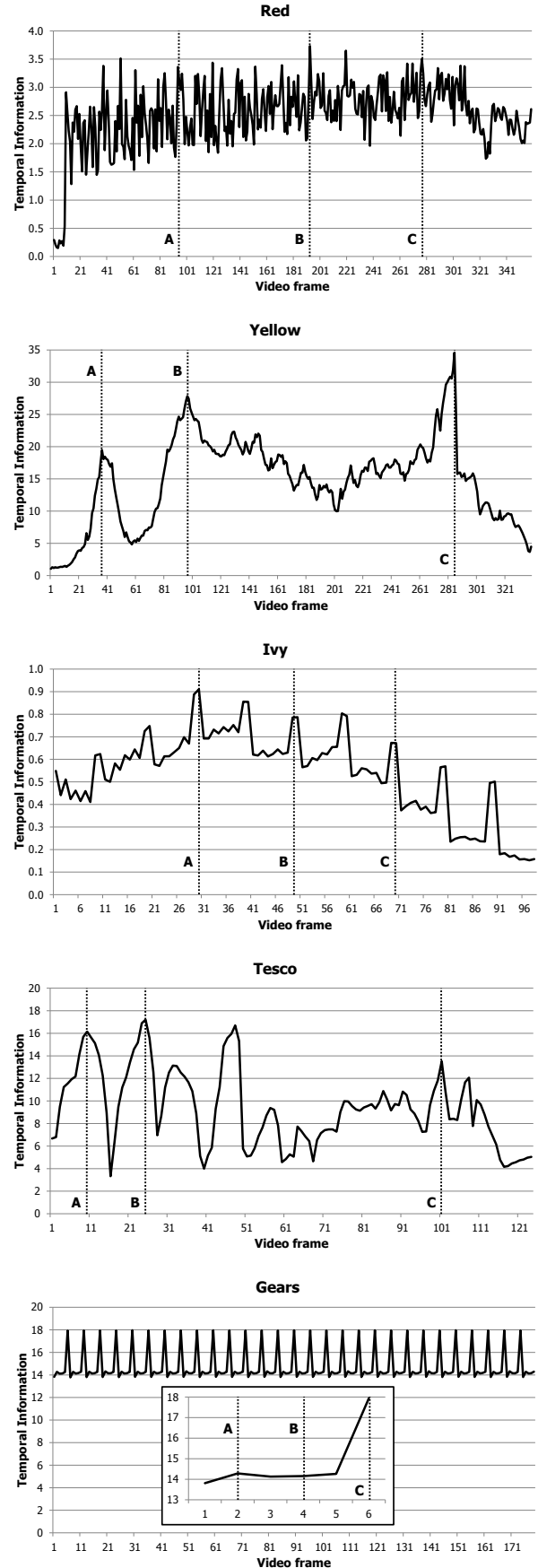


Figure 1: Temporal Information of the video contents. The video frames selected for objective assessment are labeled in chronological order as A, B and C.

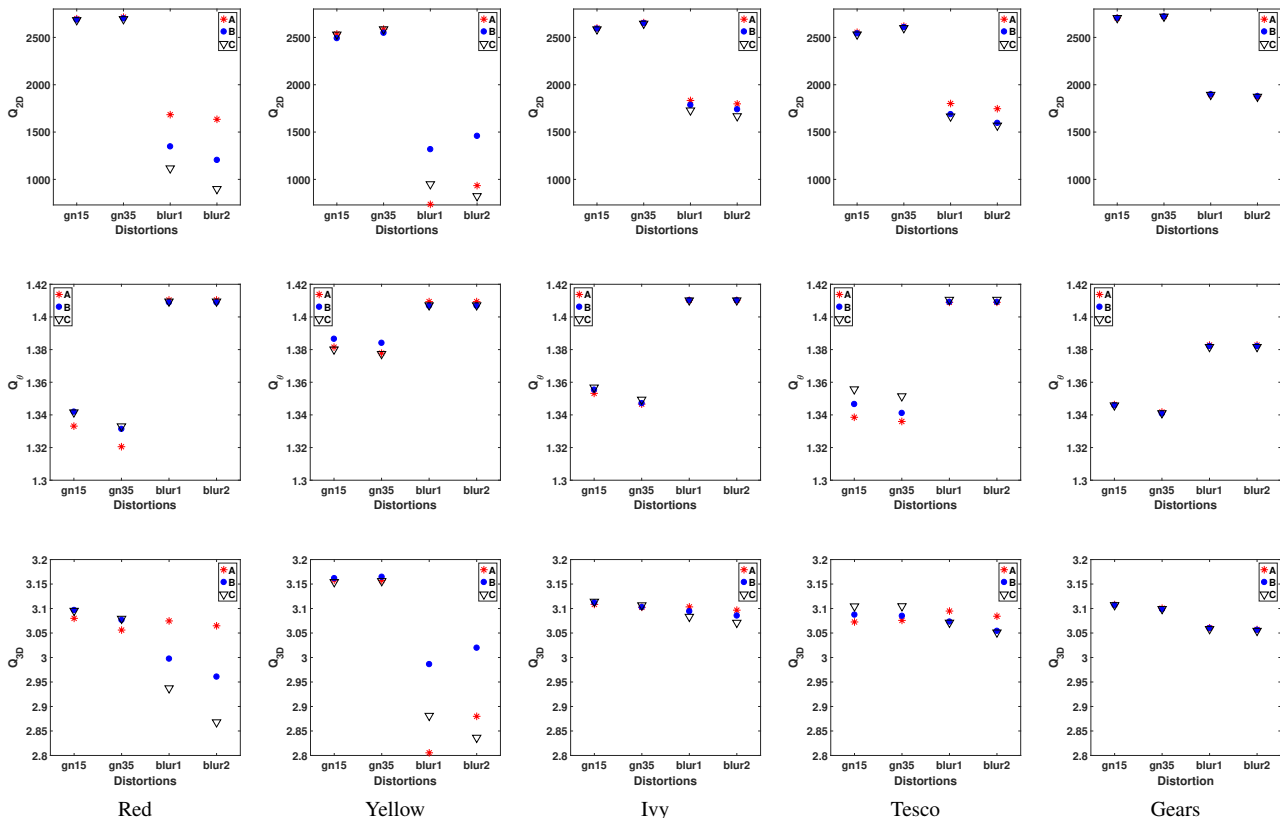


Figure 2: Results of 3D objective quality evaluation of key frames. Top row: Q_{2D} , middle row: Q_{θ} and bottom row: Q_{3D} .

tween an optical flow image noisy case and the corresponding reference optical flow image indicates disturbances in angular continuity and can be measured by any objective metric from the class of structural similarity measures.

In this work, we used the aforementioned quality metric as it is. The distance metric used in computing Q_{2D} was “Wave-Hedges”. Multi-scale structural similarity measure (MS-SSIM) was used to find similarity between two optical flow arrays. The value of pooling parameter was 0.89. For the sake of comparison with the previously reported results, the 2D distortions introduced to the source images, before light field conversion, were as follows: (i) Gaussian noise, $\mu = 0$, $\sigma = 0.15$, (ii) Gaussian noise, $\mu = 0$, $\sigma = 0.15$, (iii) Gaussian blur, $\sigma = 1$, and (iv) Gaussian blur, $\sigma = 2$. In this paper, these distortions are denoted as “gn15”, “gn35”, “blur1” and “blur2”, respectively.

The aforementioned distortions were introduced to three key frames of certain video contents. Thus, we evaluated quality of static 3D views composed of key frames from multiple cameras, or in other words, 3D key frames of a light field video. The video content and the process of extracting key frames are described next.

4. VISUAL STIMULI

In our work, we used 5 video sequences, and from each video, we selected 3 frames, based on Temporal Information. The angular resolution of each and every source sequence was 0.5 degree (2 source views per degree inside the field of view) and spatial resolution varied. *Red* and *Yellow* (both 1024×768) were sets of columns performing spatial movements and rotations along

the different axes (and also changing color in case of *Yellow*), *Ivy* (960×540) was a slow-paced subtle animation, *Tesco* was a highly mobile scene¹, and *Gears* was a looping animation (6 frames repeating) of three rotating gears. The Temporal Information of these video sequences and the selected frames are reported in Figure 1.

We decided to select frames with high values of Temporal Information, as those tend to have the highest level of motion (or generally variation) in the content, and therefore pose the greatest challenges (e.g., motion blur or defocus). A high-magnitude spike in such data can be caused by a sudden scene change. Here we need to state that all 5 sequences were without any scene change, and furthermore, the camera position, angle and level of zoom was constant for all frames.

5. RESULTS

We now present the results of the objective quality evaluation of the key frames. Figure 2 depicts the Q_{2D} , Q_{θ} , and Q_{3D} values in top, middle and bottom rows, respectively. As expected, the Q_{2D} values vary considerably across (i) frames of a data-set, (ii) data-sets, and (iii) distortion categories. The variation in Q_{2D} values across frames of a data-set within a distortion category can be attributed to the content itself.

Interestingly, Q_{θ} values also exhibit some variation, although no explicit angular distortion was introduced to the content, revealing interplay between spatial and angular components of the

¹The source videos of *Ivy* and *Tesco* were created by Post Edison (<http://www.postedison.hu/>) and were provided to Holografika for research purposes.

quality [8]. The Q_θ attains its maximum value of $\sqrt{2}$ when angular disturbances are absent. In case of “blur1” and “blur2” distortions, Q_θ values are closer to $\sqrt{2}$ than those for “gn15” and “gn35” distortions. This indicates that Gaussian noise added to individual images degraded the quality of 3D view composed of those images.

The Q_{3D} scores were obtained by pooling Q_{2D} and Q_θ as $Q_{3D} = Q_{2D}^{1-\alpha} \times Q_\theta^\alpha$. The value of ‘ α ’ was chosen to be 0.89, which results in significantly more weight to the angular quality than the spatial quality. Nevertheless, in the metric space, different distortion categories are well-separated for some contents (“Red”, “Yellow” and “Tesco”) and are indistinguishable for the remaining content. It would be interesting to see if such objective assessment is in agreement with perceptual separation among the contents and distortion categories.

6. CONCLUSION

3D displays in general and light field displays in particular have potential to provide substantial benefits to several applications such as edutainment, telepresence, telemedicine etc. Even further, Cserkaszkzy *et al.* envisage individual consumer-centric use cases such as light field video-on-demand [7]. Therefore, all the stage of light field processing pipeline from content acquisition to visualization should be optimized to enhance end user’s QoE. We take a step to this end by evaluating quality of 3D key frames of a light field video.

In this paper, we presented the objective quality assessment of key frames extracted from light field video contents. Since, to the best knowledge of the authors, no objective metric for assessing quality of light field video was available at the time of this paper, we used a full-reference 3D objective metric which was originally proposed for static light field content, optimized for a few distortions [16]. The efficacy of this metric was reported by Tamboli *et al.* in terms of correlation among subjective scores of their study and the objective metric values. In this paper, we (i) added same distortions to the source images and (ii) retained the default parameter settings of the objective metric to evaluate the quality of 3D views composed of key frames, i.e., 3D key frames of a 3D video. The purpose of this study was to examine if the quality of 3D key frames is an indicator of perceptual quality of the corresponding 3D video. To this end, we observed that the distortions were sufficiently separated in metric space, even with the parameters setting unoptimized for the content at hand. In future work, we would plan to test if objective scores correlate with the subjective opinions, using default and optimized parameter settings for the aforementioned metric.

7. REFERENCES

- [1] J. Y. Son, H. Lee, B. R. Lee, and K. H. Lee, “Holographic and Light-Field Imaging as Future 3-D Displays,” *Proceedings of the IEEE*, vol. 105, no. 5, pp. 789–804, May 2017.
- [2] A. Dricot, J. Jung, M. Cagnazzo, B. Pesquet, F. Dufaux, P. T. Kovács, and V. K. Adhikarla, “Subjective evaluation of Super Multi-View compressed contents on high-end light-field 3D displays,” *Signal Processing: Image Communication*, vol. 39, pp. 369–385, 2015.
- [3] F. Dufaux, B. Pesquet-Popescu, and M. Cagnazzo, *Emerging Technologies for 3D Video: Creation, Coding, Transmission and Rendering*, 1st ed. Wiley Publishing, 2013.
- [4] A. Kondozi and T. Dagiuklas, *Novel 3D Media Technologies*. Springer Publishing Company, Incorporated, 2014.
- [5] S. Chikkerur, V. Sundaram, M. Reisslein, and L. J. Karam, “Objective Video Quality Assessment Methods: A Classification, Review, and Performance Comparison,” *IEEE Transactions on Broadcasting*, vol. 57, no. 2, pp. 165–182, June 2011.
- [6] N. S. Holliman, N. A. Dodgson, G. E. Favalora, and L. Pockett, “Three-dimensional displays: a review and applications analysis,” *IEEE transactions on Broadcasting*, vol. 57, no. 2, pp. 362–371, 2011.
- [7] A. Cserkaszkzy, A. Barsi, P. A. Kara, and M. G. Martini, “Towards display-independent light-field formats,” in *2017 International Conference on 3D Immersion (IC3D)*, Dec 2017, pp. 1–7.
- [8] P. A. Kara, A. Cserkaszkzy, A. Barsi, T. Papp, M. G. Martini, and L. Bokor, “The interdependence of spatial and angular resolution in the quality of experience of light field visualization,” in *2017 International Conference on 3D Immersion (IC3D)*, Dec 2017, pp. 1–8.
- [9] A. Mustafa, M. Volino, J.-y. Guillemaut, and A. Hilton, “4D Temporally Coherent Light-field Video,” *arXiv preprint arXiv:1804.11276*, 2018.
- [10] F. Battisti, E. Bosc, M. Carli, P. Le Callet, and S. Perugia, “Objective image quality assessment of 3D synthesized views,” *Signal Processing: Image Communication*, vol. 30, pp. 78–88, 2015.
- [11] X. Liu, Y. Zhang, S. Hu, S. Kwong, C.-C. Kuo, and Q. Peng, “Subjective and Objective Video Quality Assessment of 3D Synthesized Views With Texture/Depth Compression Distortion,” *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 4847–4861, Dec 2015.
- [12] E. Ekmekcioglu, S. Worrall, D. De Silva, A. Fernando, and A. M. Kondozi, “Depth Based Perceptual Quality Assessment for Synthesised Camera Viewpoints,” in *User Centric Media*. Springer, 2012, pp. 76–83.
- [13] A. Tikanmaki, A. Gotchev, A. Smolic, and K. Miller, “Quality assessment of 3D video in rate allocation experiments,” in *IEEE International Symposium on Consumer Electronics*, 2008, April 2008, pp. 1–4.
- [14] I. Viola, M. Rerabek, and T. Ebrahimi, “Comparison and Evaluation of Light Field Image Coding Approaches,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 7, pp. 1092–1106, Oct 2017.
- [15] V. K. Adhikarla, M. Vinkler, D. Sumin, R. K. Mantiuk, K. Myszkowski, H.-P. Seidel, and P. Didyk, “Towards a quality metric for dense light fields,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2017, pp. 3720–3729.
- [16] R. R. Tamboli, B. Appina, S. Channappayya, and S. Jana, “Super-multiview content with high angular resolution: 3D quality assessment on horizontal-parallax lightfield display,” *Signal Processing: Image Communication*, vol. 47, pp. 42–55, 2016.