Tamboli, Roopak R., Cserkaszky, Aron, Kara, Peter A., Barsi, Attila and Martini, Maria G. (2018) Objective quality evaluation of an angularly-continuous light-field format. In 2018 International Conference on 3D Immersion (IC3D) : proceedings. Piscataway : IEEE. ISSN (2379-1780), ISBN: 9781538675908.

https://doi.org/10.1109/IC3D.2018.8657876

OBJECTIVE QUALITY EVALUATION OF AN ANGULARLY-CONTINUOUS LIGHT-FIELD FORMAT

Roopak R. Tamboli[*,†], *Aron Cserkaszky*[*], *Peter A. Kara*[*,‡,§], *Attila Barsi*[*], *Maria G. Martini*[‡]

[*]Holografika, Budapest, Hungary
Email: {r.tamboli, a.cserkaszky, a.barsi}@holografika.com
[†]Indian Institute of Technology Hyderabad, Kandi, Sangareddy, India
Email: ee13p0008@iith.ac.in
[‡]WMN Research Group, Kingston University, UK
Email: {p.kara, m.martini}@kingston.ac.uk
[§]Budapest University of Technology and Economics, Budapest, Hungary
Email: kara@hit.bme.hu

ABSTRACT

With the rapid advances in light-field display and camera technology, research in light-field content creation, visualization, coding and quality assessment has gained momentum. While light-field cameras are already available to the consumer, light-field displays need to overcome several obstacles in order to become a commonplace. One of these challenges is the unavailability of a light-field visualization format which can be used across various light-field displays. Existing light-field representations are optimized for specific displays and converting them for visualization on a different display is a computationally expensive operation, often resulting in the degradation of perceptual quality. To this end, an intermediate, display-independent and angularly-continuous light-field representation format has been proposed recently, targeted towards large-field-of-view light-field displays. In this paper, we evaluate the said data format in terms of degradation in objective quality under three compression methods. We found that, while offering display-independence, the intermediate light-field format maintains the same objective quality in general and achieves higher objective quality in some cases compared to the conventional linear camera representation.

***Index Terms*—** Light-field, visualization format, objective quality assessment

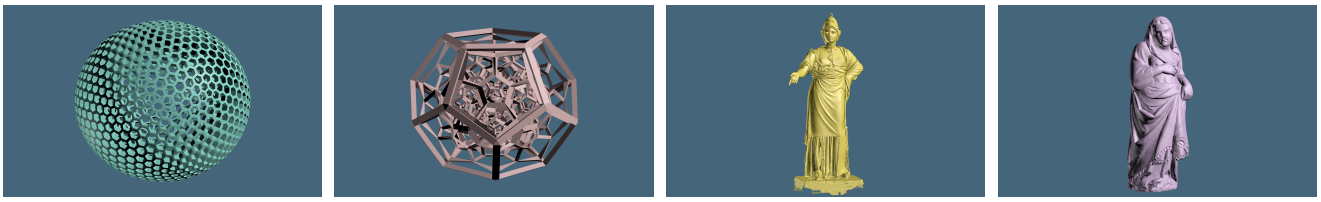
1. INTRODUCTION

Light-field (LF) capture and visualization has progressed significantly in recent years [1]. The current generation of LF displays are used mostly in the industry and research institutions. On the other hand, LF cameras are already available on the consumer market. These cameras capture narrow-baseline LF images with their microlense-based optics [1]. Such formats are not suitable for wide-baseline LF displays.

Wide-baseline LF displays offer life-like immersive experience via continuous motion parallax. In order to do so, such displays need a large amount of input data, often captured using a multi-camera array [2]. The physical size of the cameras limits how densely they can be arranged. Furthermore, deploying a dense camera rig is economically and technologically prohibitive. In practice, multi-view content is acquired using sparse camera arrays, followed by view interpolation. A computationally expensive conversion of the captured LF to the display-specific LF is then performed. The quality of such converted LF depends on the closeness between the sampling of the ray space of the two LFs. Thus, LF data optimized for a certain display cannot be used on another display without several inefficiencies in conversion. In future applications of LF — such as dynamic adaptive video streaming [3] and teleconferencing [4] — low latency requirements necessitate display-specific LF mapping to be performed at the acquisition side as opposed to the display side [5]. It is clearly not possible for the acquisition side to convert and transmit display-specific LF for all available display types due to huge amounts of data. In such situations, the availability of LF representation formats that are oblivious to the acquisition side — as well as the display side — becomes a necessity. To this end, Damghanian *et al.* recently proposed a camera-agnostic format and processing for LF data [6], and Cserkaszky *et al.* focused on wide-baseline displays and proposed an angularly-continuous intermediate LF representation [5].

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

[1] Stanford light field file format, `http://graphics.stanford.edu/software/lightpack/lightpack.html`



(a) Source stimulus A (b) Source stimulus B (c) Source stimulus C (d) Source stimulus D

Fig. 1: 2D views of the 3D models used to generate stimuli in 's-t-phi' format and perspective camera format.

In this paper, we evaluate the intermediate LF format proposed by Cserkaszky *et al.* [5]. Specifically, the said format was compared with perspective camera representation under various compression schemes and a few objective quality metrics. It was observed that the objective quality values for the intermediate LF format were the same as those for the perspective camera format. In some cases, the intermediate format offered higher objective quality values. Thus, display-independence using the intermediate format is achieved without any compromise in objective quality values when compared to the conventional format.

The rest of this paper is organized as follows: Section 2 presents the related work on LF representation formats. The dataset, compression methods and objective quality metrics used in this paper are described in Section 3. Results of this study are presented in Section 4. A discussion is provided in Section 5. Section 6 concludes the paper.

2. RELATED WORK

Existing LF formats are either designed to represent dense and uniformly sampled LFs or they contain a set of images captured from different positions by perspective cameras along with calibration parameters [5][7]. Narrow-baseline LF data can be captured using plenoptic cameras. On the other hand, wide-baseline LF content needs to be captured using either a single moving camera (restricted to static scenes) or an array of cameras. A wide variety of datasets have been used as LF content. In a recent call for evaluations of super multi-view (SMV) and free-viewpoint content, several camera configurations to capture multiview data were proposed, such as 1D linear arrangement, 2D or full-parallax arrangement, convergent arc setups, etc. [8].

Both the aforementioned regimes of LF acquisition generate a huge amount of data and therefore necessitate compression. Existing LF compression schemes generally adapt intra- and inter-frame coding [9]. For example, Ahmed *et al.* treated views from multi-camera system as frames of a multiview sequence, and compressed them using the Multiview extension of High Efficiency Video Coding [10]. Similarly, Guo *et al.* proposed a two-pass encoding system for pseudo-temporal sequence of LF data captured using plenoptic cameras [11]. While LF compression has gained significamt attention, research towards an LF representation that does not adhere to a specific LF camera or LF display is in its infancy.

Recently, Damghanian *et al.* proposed a camera-agnostic format and processing for LF data [6]. Their LF processing pipeline decouples the LF capture system from LF storage and processing. The feasibility of their two camera-agnostic data formats — which store geometry and color content of the LF data, respectively — was demonstrated using a depth extraction algorithm, applied to LF data captured by four types of camera setups.

Pertaining to wide-baseline LF displays, Cserkaszky *et al.* recently proposed a novel parameterization of LF data [5]. This representation lies between the source content (e.g., perspective camera images) and the final LF slices (optical module images). The format only assumes that the screen of the display is approximately flat, and the rays it can emit have a symmetry in the angular domain. The format describes the 4D LF with two spatial coordinates (that indicate the start positions of the rays), and two angular coordinates (that give the directions of the rays). The header of the format contains the properties of the LF: the number of pixels and their physical size in each spatial dimension, the field of view (FOV) and the number of angular views in each angular dimension. The said dimensions are denoted with s, t, ϕ and θ [12]. From the intermediate representation, it is clearly possible to make a geometry-specific conversion for a specific LF display. Due to the horizontal-parallax-only (HPO) nature of the existing, commercially available LF displays, Cserkaszky *et al.* refer to their novel format as the 's-t-phi format', because the second angle θ is not required in this case. Since the representation provides one image per angle, it is also referred to as an 'angularly-continuous light-field format'.

3. EXPERIMENTS

We begin with the description of the investigated LF content, rendered in both perspective camera format and s-t-phi format. Next, we provide details of the three compression schemes used in this study. Finally, we describe the 2D image quality metrics, as well as the 3D objective quality metric used in this paper.

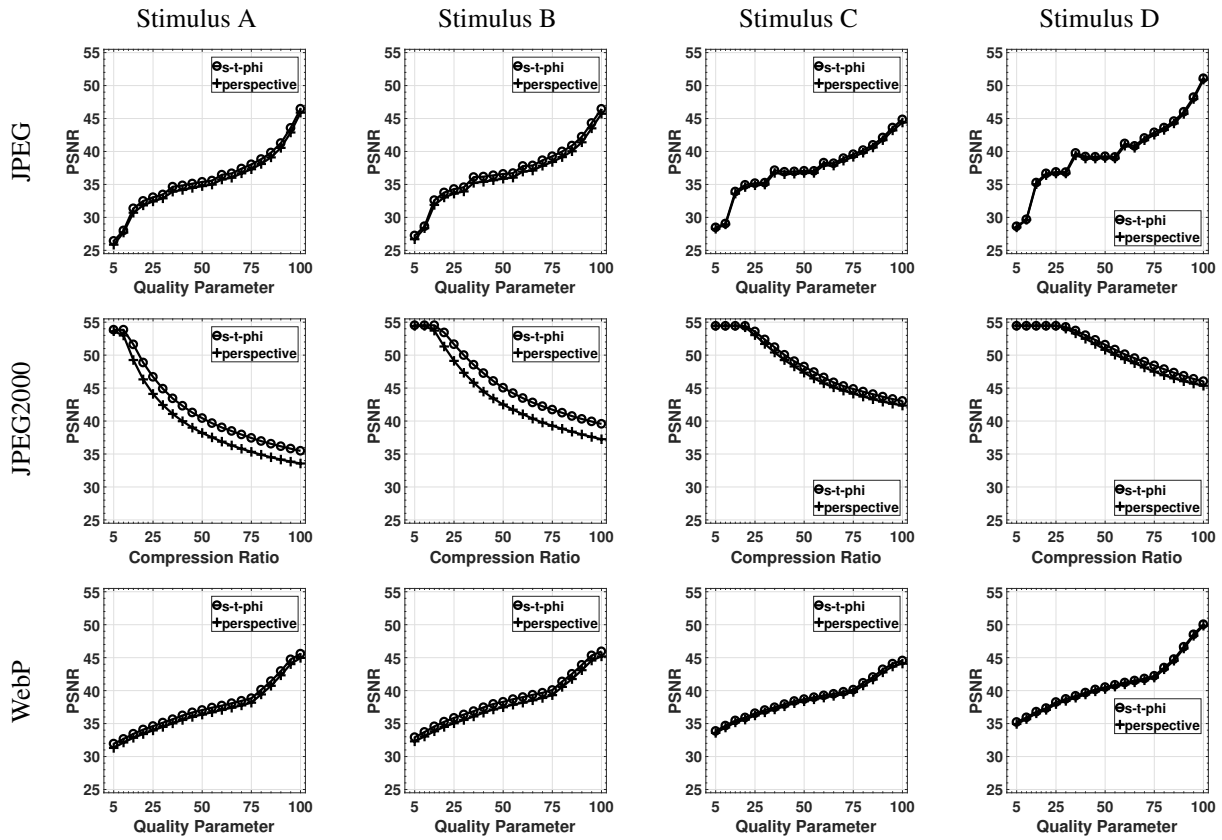


Fig. 2: 2D quality assessment using PSNR metric.

3.1. Investigated light-field content

In this section, we describe the LF content chosen for compression experiments [13]. A subset of this content was also used to derive stimuli for the expert evaluation of the s-t-phi format by Cserkaszky *et al.*[12]. The 2D views of the 3D models used to generate the stimuli are shown in Figure 1. In this work, for both the perspective camera format and the s-t-phi format, 101 images were rendered, corresponding to the 101 linearly arranged cameras and the 101 angles, respectively. The number of images/angles were chosen in order to provide an angular resolution of 2 views per degree (also denoted as an angular resolution of 0.5 degree) for an HPO display calibrated for 50° FOV.

Stimulus A and B were complex mathematical bodies with large depths[2], stimulus C and D were laser-scanned statues with smaller depths[3]. The difference between A and B was that while A (polyhedron with 972 faces) had a detailed, uniform grid on the front (closest to the observer), stimulus B (structure of 120 regular dodecahedra) had a simple, smooth surface segment on the side of the object. In general, both A and B suffer significant penalties in the perceived visual quality, even at the slightest compression level.

3.2. Investigated image compression methods

We now describe the three compression methods used in this work. Each source stimulus — rendered in both s-t-phi format and perspective camera format — was compressed using JPEG[4], JPEG2000[5] and WebP[6] compression methods at 20 quality/compression levels. The stimuli were compressed before the display-specific LF conversion. Thus, the 4 source stimuli compressed using the 3 compression methods at 20 quality levels resulted in 240 sets of 101 images. Including the 4 uncompressed sets, the total number of sets amounted to 244. The value of 'Quality Parameter' (or Compression Ratio in case of JPEG2000) was varied from 5 to 100 in steps of 5. The JPEG and JPEG2000 compressions were achieved in MATLAB. For WebP compression, we built the '`cwebp`' and '`dwebp`' tools from the WebP codec v0.6.1.

Each compressed image was compared with the corresponding reference image using the objective metrics described

[2]George W. Harts Rapid Prototyping Web Page, `www.georgehart.com/rp/rp.html`

[3]Jotero.com 3D-Scan and 3D Measurement, `forum.jotero.com/viewtopic.php?t=3`

[4]`https://jpeg.org/jpeg/index.html`

[5]`https://jpeg.org/jpeg2000/index.html`

[6]`https://developers.google.com/speed/webp/`

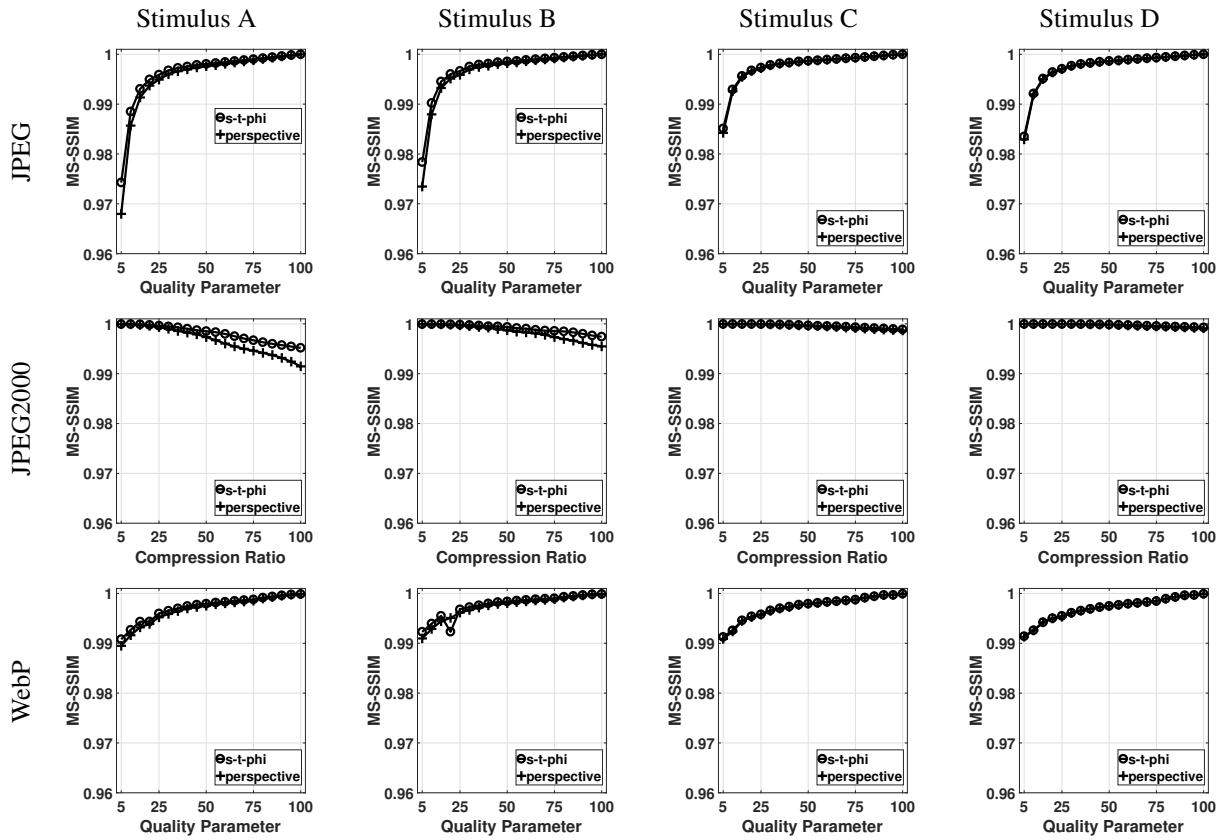


Fig. 3: 2D quality assessment using MS-SSIM metric.

in Section 3.3. A static 3D view on a LF display is composed of a set of images in our experiment. Therefore, an average of the per-image quality values was considered as the quality of the 3D view, computed in a full-reference (FR) setting.

3.3. 2D objective quality metrics

The FR 2D image quality metrics used in this work can be classified into the following categories: pixel-based, structure-based and scene-statistics-based. The metrics selected from these categories were Peak Signal-to-Noise Ratio (PSNR), Multi-Scale Structural SIMilarity (MS-SSIM) [14], Feature Similarity Index Measure (FSIM) [15] and Information Fidelity Criterion (IFC) [16], respectively.

MS-SSIM relies on the ability of the human visual system (HVS) to extract structural information from a scene and assesses image quality based on the degradation of structural information. FSIM is based on two components. The first component, termed as phase congruency, is a dimensionless measure of the significance of local structure. The second component, called the image gradient magnitude, accounts for the contrast information. IFC relies on natural scene statistics and assesses the perceptual quality by quantifying the mutual information between the reference and the distorted images.

3.4. 3D objective quality metric

We now briefly explain the FR 3D objective quality metric used in this paper [17]. The metric — considering the spatio-angular nature of the LF content — evaluates the spatial and angular quality scores of a 3D perspective visualized on a LF 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 score Q_{2D} is the distance between a feature vector of a reference 3D view and a feature vector of a distorted 3D view, where each constituent image is distorted (compressed).

The angular quality score Q_θ finds structural similarity between optical flow arrays computed for a pristine and a distorted 3D view. Optical flow is computed between successive constituent images of a 3D view. The difference between an optical flow array for a distorted 3D view and the corresponding reference optical flow array indicates disturbances in angular continuity, which can be measured by any objective metric from the class of structural similarity measures.

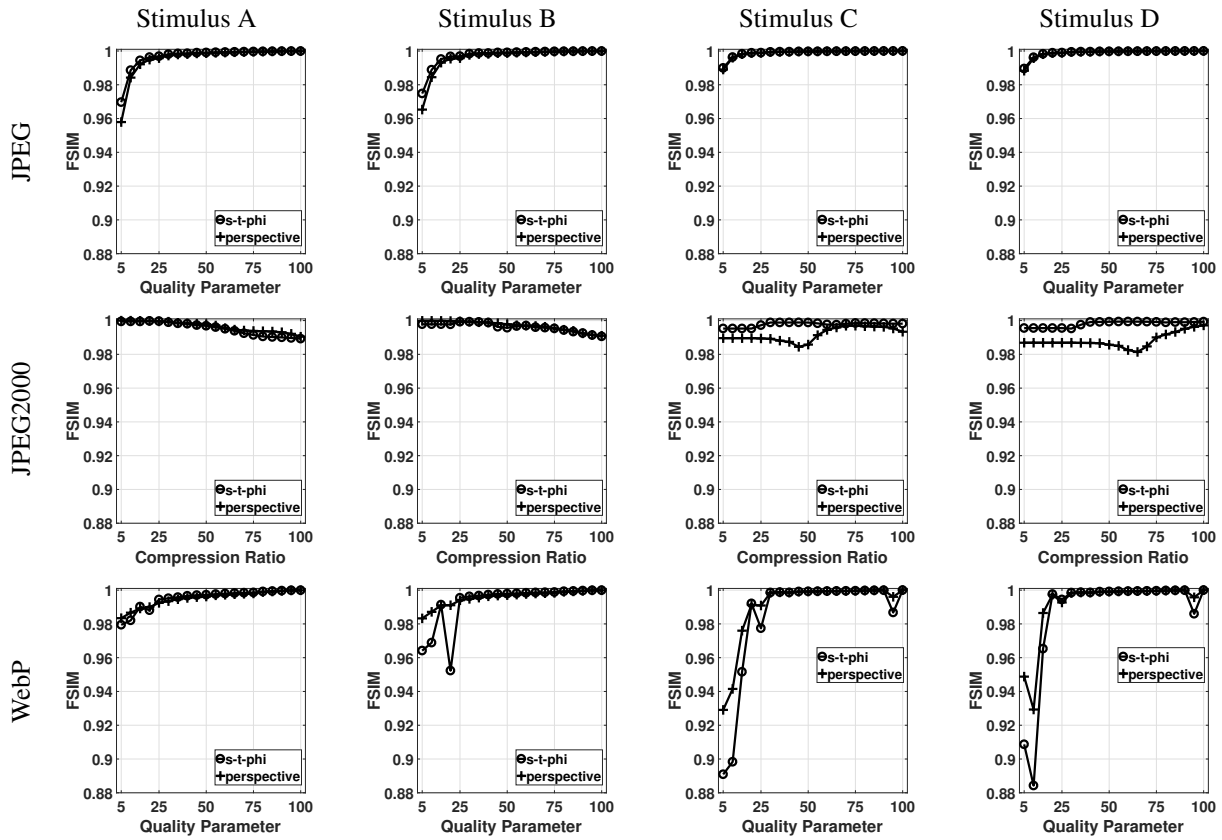


Fig. 4: 2D quality assessment using FSIM metric.

The spatial and angular quality scores are pooled into 3D quality score as $Q_{3D} = Q_{2D}^{1-\alpha} \times Q_{\theta}^{\alpha}$. In this work, the distance metric used in computing Q_{2D} was 'WaveHedges' [18]. 'MS-SSIM' was used to find similarity between the two optical flow arrays. The value of the pooling parameter α was '0.89'. These parameters were found to be optimal in terms of correlation between subjective and objective scores in a study conducted by Tamboli *et al.* [17]. In their study, spatial distortions were added to the content while angular resolution was fixed. Similarly, in this paper, spatial distortions (compression) were added to the content and angular resolution was fixed. Therefore, we used the aforementioned quality metric in its original settings, without any optimization specific to the LF content or distortions.

4. RESULTS

We now present the results of the source coding experiments. Figures 2, 3, 4, 5 and 6 depict the results calculated using the PSNR, MS-SSIM, FSIM, IFC and Q3D metrics, respectively. In these figures, columns correspond to the different stimuli and rows correspond to the different compression methods. For JPEG and WebP methods, the X-axis represents 'Quality Parameter' whereas for JPEG2000 method, the X-axis represents 'Compression Ratio'. Therefore, the nature of plots for JPEG2000 is opposite to that of the other two compression schemes. Furthermore, the 2D objective metrics used in this work are similarity metrics whereas the 3D objective metric is a combination of a distance metric and a similarity metric. The key objective of this work — to compare the two formats in terms of difference in objective quality — can be served by the use of any kind of metric.

As seen in Figure 2, the PSNR values for the s-t-phi format on average were higher than those for the perspective camera format, for all three compression schemes. Especially, for stimuli A and B, which were complex mathematical objects, the s-t-phi format maintained higher PSNR values compared to the perspective camera format when compressed using JPEG2000 method. Similar behavior was observed for the MS-SSIM metric (Figure 3).

In case of the FSIM metric (Figure 4), the s-t-phi format exhibited higher objective quality than the perspective format for stimuli C and D under JPEG2000 compression scheme. The perspective camera format appears to perform better at lower compression levels of WebP method, albeit with a narrow margin. For JPEG compression, no significant difference in the objective quality values was observed for all the four stimuli.

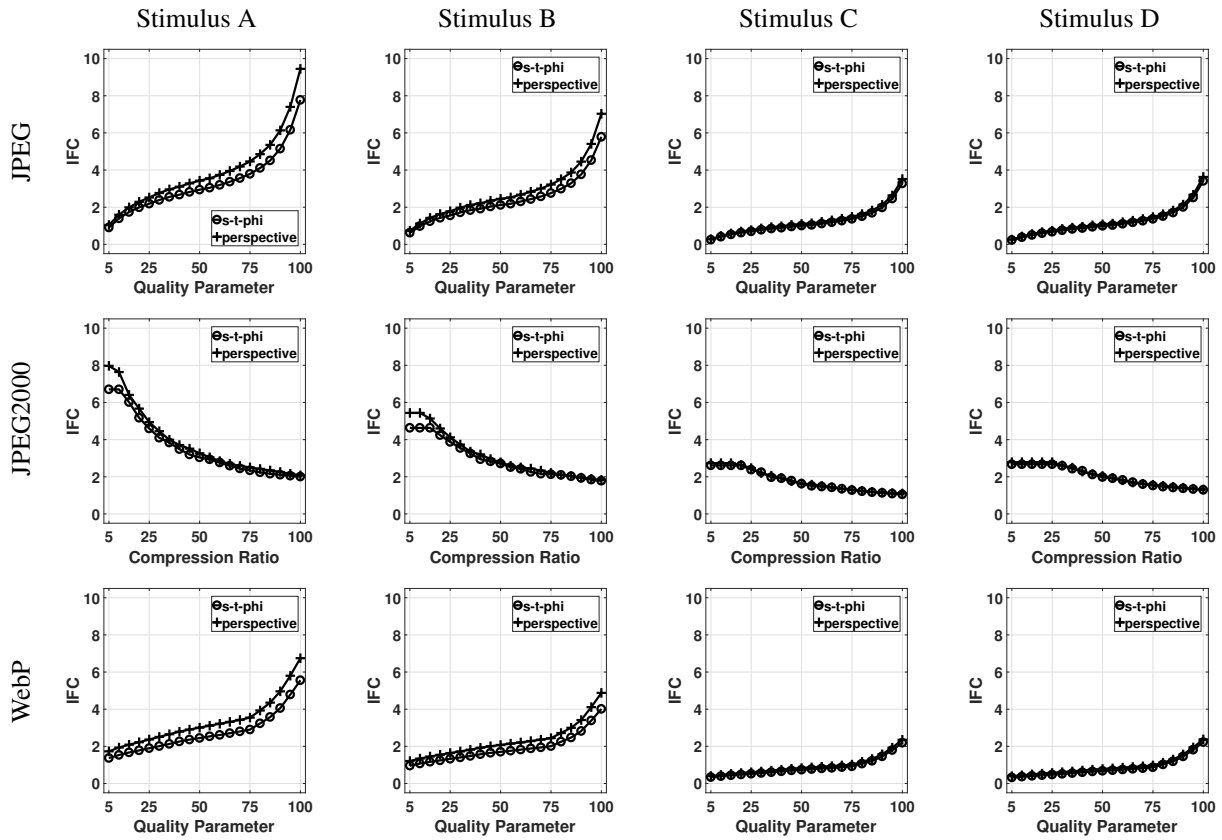


Fig. 5: 2D quality assessment using IFC metric.

The results for the IFC metric are shown in Figure 5. The perspective camera format was found to be better for stimuli A and B under JPEG and WebP compression methods. Nevertheless, the objective quality values were very close at higher compression levels for both the formats. In the remaining cases, the differences in objective quality values were not significant.

It should be noted that in the results discussed above, the quality of 3D content was computed as the average of quality values for constituent 2D images. This assumption ignores the fact that the viewers' 3D experience is affected by spatial as well as angular properties of the content presented. To this end, we used a FR 3D objective quality metric, described earlier in Section 3.4. Results for the final 3D quality score Q_{3D} are shown in Figure 6. No significant difference was observed between two formats in terms of objective quality value. For stimuli A, B and D, variation in Q_{3D} scores was very small, whereas for stimulus C, Q_{3D} scores varied significantly. This can be attributed to the fact that stimulus C has large depth variations compared to other three stimuli [19]. Also, the absence of explicit angular distortions may have resulted in low variations in Q_{3D}, as the angular quality score has higher weight during the pooling operation. The spike in Q_{3D} value at quality parameter value of 95 for stimulus C under WebP compression arose due to the artifacts introduced by the '`dwebp`' tool. The said tool was used to convert '`webp`' images to '`png`' format images for computations in MATLAB. This anomaly was observed in all images generated for stimuli C and D, in both representation formats, even with the newer versions of the '`dwebp`' tool.

The range of values taken by Q_{2D}, Q_θ and Q_{3D} were found to be [16.01, 1668.31], [1.36, 1.41] and [1.85, 3.06], respectively. It was found that the minor differences that exist in some cases were due to large differences in corresponding spatial quality scores Q_{2D}. Indeed, computing the element-wise absolute differences in the Q_{2D} values, followed by computing variances across the quality parameter values/ compression ratios revealed that the variances were of the order 10^5. Among the three compression methods, variances were high for the JPEG2000 method. Across the stimuli, although no clear pattern was observed, variances of Q_{2D} values for stimuli C and D were high in general.

On the other hand, the angular quality scores Q_θ were not very different. The variances of the differences in angular quality scores Q_θ of the two LF formats — calculated separately across the stimuli, across the compression methods, as well as across the quality parameter values/ compression ratios — were of the order 10^{-3}. Since the metric was used

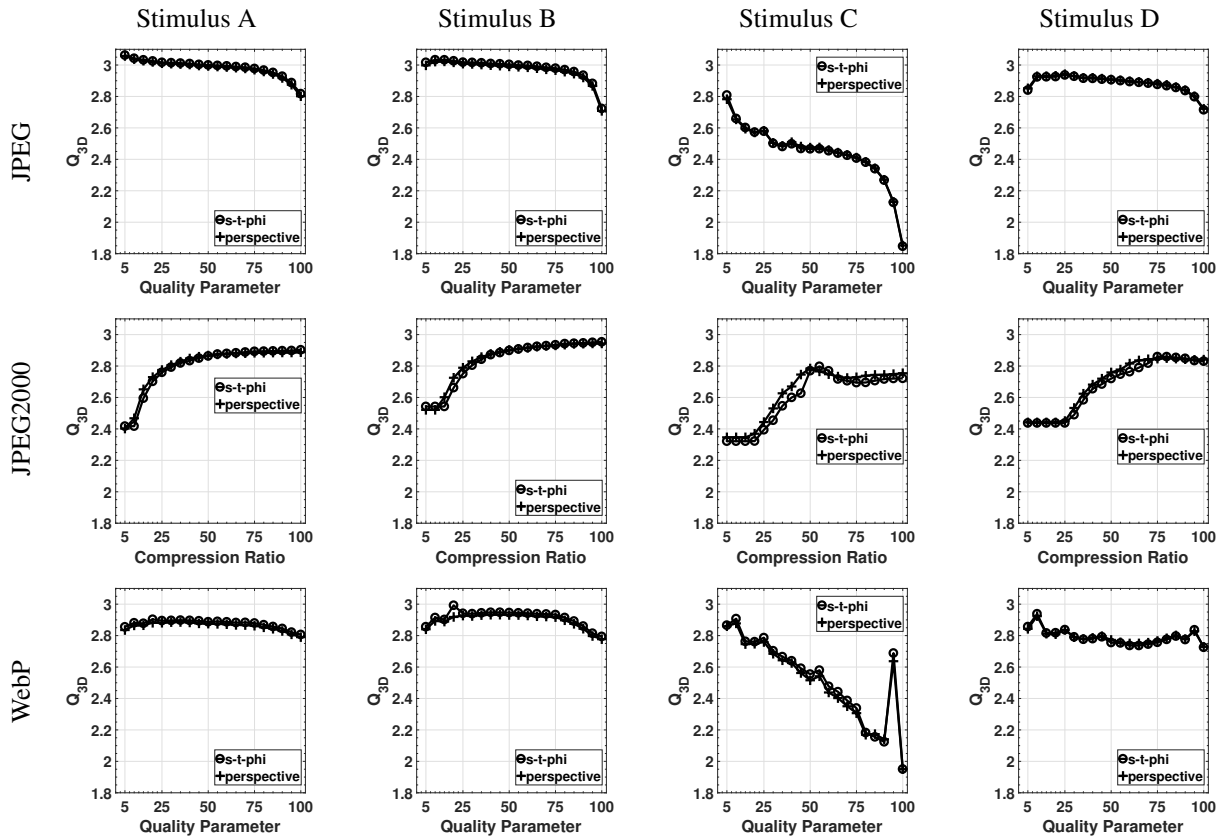


Fig. 6: 3D quality assessment using Q_{3D} metric.

with its original settings, the angular quality scores received higher weight, which resulted in a minuscule difference in the overall 3D quality score.

5. DISCUSSION

The results presented in Section 4 corroborate with the observations made by Cserkaszky *et al.* in their expert evaluation study of the proposed s-t-phi format [12]. Specifically, the stimuli rendered in s-t-phi format were found to be better than those rendered in perspective camera format, for low angular resolutions. On the other hand, for sufficient angular resolutions, both formats were found to provide similar perceptual experience. Their subjective tests were conducted on the HoloVizio C80 light-field cinema system [7]. In this paper, the content was rendered at a recommended angular resolution of 2 views per degree [13]. Therefore, although no explicit angular distortions were studied in this paper, the objective results can be qualitatively compared to the perceptual quality of stimuli rendered at sufficient or high angular resolutions used in the aforementioned expert evaluation study.

The FR 3D objective quality metric used in this work was found to be a good indicator of perceived quality on a large LF display in an earlier work of Tamboli *et al.* [17]. Specifically, certain spatial distortions were added to multi-camera datasets before the display-specific LF conversion and the objective quality assessment was performed. The objective scores were found to correlate well with subjective score obtained through a test conducted on Holografika's HV721RC display [8]. In this paper, a similar study was conducted where only spatial distortions (compression artifacts) were introduced to the content without any display-specific LF conversion. Therefore, we believe that, even in this case, the objective quality score Q_{3D} provides a good estimate of perceptual experience if the contents were to be visualized on a LF display.

6. CONCLUSION

In this paper, we evaluated objective quality for light-field datasets rendered in a novel intermediate format, as well as in the conventional perspective camera format. The datasets were compressed using three distinct compression methods

[7] HoloVizio C80 light-field cinema system, http://holografika.com/c80-glasses-free-3d-cinema/

[8] HoloVizio 721RC, http://www.archive.holografika.com/Products/HoloVizio-721RC.html

and full-reference quality assessment was performed using 2D and 3D quality metrics. It was observed that the proposed s-t-phi format retains objective quality levels at par with the perspective camera format. In some cases, the s-t-phi format was found to be better. Thus, the intermediate light-field representation offers several advantages over the conventional format [5], without any compromise in objective quality as observed in the experiments carried out in this paper. In the future, we plan to evaluate this format extensively via subjective and objective tests, with combinations of spatial and angular distortions. We also plan to explore use cases involving light-field video and its coding with the s-t-phi format.

7. REFERENCES

[1] C. Guillemot and R. Farrugia, "Light field image processing: overview and research issues," *MMTC Communications-Frontiers*, vol. 12, no. 4, 2017.

[2] M. Domański, T. Grajek, C. Conti, C. J. Debono, S. M. de Faria, P. Kovacs, L. F. Lucas, P. Nunes, C. Perra, N. M. Rodrigues *et al.*, "Emerging imaging technologies: trends and challenges," in *3D Visual Content Creation, Coding and Delivery*. Springer, 2019, pp. 5–39.

[3] P. A. Kara, A. Cserkaszky, M. G. Martini, A. Barsi, L. Bokor, and T. Balogh, "Evaluation of the concept of dynamic adaptive streaming of light field video," *IEEE Transactions on Broadcasting*, vol. 64, no. 2, pp. 407–421, 2018.

[4] A. Cserkaszky, A. Barsi, Z. Nagy, G. Puhr, T. Balogh, and P. A. Kara, "Real-time light-field 3D telepresence," in *7th European Workshop on Visual Information Processing (EUVIP)*, 2018.

[5] A. Cserkaszky, A. Barsi, P. A. Kara, and M. G. Martini, "Towards display-independent light-field formats," in *International Conference on 3D Immersion (IC3D)*, 2017.

[6] M. Damghanian, P. Kerbiriou, V. Drazic, D. Doyen, and L. Blond, "Camera-agnostic format and processing for light-field data," in *IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, 2017, pp. 7–12.

[7] T. Ebrahimi, S. Foessel, F. Pereira, and P. Schelkens, "JPEG pleno: toward an efficient representation of visual reality," *IEEE MultiMedia*, vol. 23, no. 4, pp. 14–20, 2016.

[8] "MPEG-I visual test material summary, ISO/IEC JTC1/SC29/WG11 MPEG2016/N16731," https://mpeg.chiariglione.org/sites/default/files/files/standards/parts/docs/W16731%20MPEG-I%20phase%202%20test%20material.docx, Accessed: Sept. 2018.

[9] I. Viola, M. Rerabek, T. Bruylants, P. Schelkens, F. Pereira, and T. Ebrahimi, "Objective and subjective evaluation of light field image compression algorithms," in *Picture Coding Symposium (PCS)*, 2016.

[10] W. Ahmad, M. Sjöström, and R. Olsson, "Compression scheme for sparsely sampled light field data based on pseudo multi-view sequences," in *Optics, Photonics, and Digital Technologies for Imaging Applications V*, 2018.

[11] B. Guo, J. Wen, and Y. Han, "Two-pass light field image compression for spatial quality and angular consistency," *arXiv:1808.00630*, 2018.

[12] A. Cserkaszky, P. A. Kara, A. Barsi, and M. G. Martini, "Expert evaluation of a novel light-field visualization format," in *3DTV Conference*, 2018.

[13] P. A. Kara, A. Cserkaszky, S. Darukumalli, A. Barsi, and M. G. Martini, "On the edge of the seat: reduced angular resolution of a light field cinema with fixed observer positions," in *Ninth International Conference on Quality of Multimedia Experience (QoMEX)*, 2017.

[14] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *The Thrity-Seventh Asilomar Conference on Signals, Systems Computers*, vol. 2, 2003, pp. 1398–1402.

[15] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: a feature similarity index for image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386.

[16] H. R. Sheikh, A. C. Bovik, and G. de Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Transactions on Image Processing*, vol. 14, no. 12, pp. 2117–2128, 2005.

[17] 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.

[18] M. M. Deza and E. Deza, *Encyclopedia of distances*. Springer, 2009.

[19] A. Cserkaszky, A. Barsi, P. A. Kara, and M. G. Martini, "To interpolate or not to interpolate: subjective assessment of interpolation performance on a light field display," in *IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, 2017, pp. 55–60.