

© 2017 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other 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.

Cserkaszky, Aron, Barsi, Attila, Kara, Peter A. and Martini, Maria G. (2017) To interpolate or not to interpolate: subjective assessment of interpolation performance on a light field display. IProceedings of the 2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW). pp. 55-60. Piscataway, U.S. : Institute of Electrical and Electronics Engineers, Inc. ISBN (electronic) 9781538605608.

<https://doi.org/10.1109/ICMEW.2017.8026223>

TO INTERPOLATE OR NOT TO INTERPOLATE: SUBJECTIVE ASSESSMENT OF INTERPOLATION PERFORMANCE ON A LIGHT FIELD DISPLAY

Aron Cserkaszkzy, Attila Barsi

Holografika
{a.cserkaszkzy, a.barsi}@holografika.com

Peter A. Kara, Maria G. Martini

WMN Research Group, Kingston University
{p.kara, m.martini}@kingston.ac.uk

ABSTRACT

Interpolating virtual views from sparse visual content increases the angular resolution of the visualization. However, such techniques may degrade the image quality through inaccurate view estimation. The smooth motion parallax through increased angular resolution, and image quality are essential components of the overall user experience. The choice of interpolation initiates an unclear trade-off between them. In this paper, we introduce our research on interpolation techniques for 3D light field visualization. We examined the performance and perceived quality of selected algorithms, in a series of subjective quality assessment experiments. Our results show an obvious interpolation technique preference on visual contents with low angular resolutions.

Index Terms— Light field display, interpolation, Quality of Experience, subjective quality assessment

1. INTRODUCTION

Interpolation is an operation that generates virtual camera images from real camera images. Interpolation can increase the angular resolution of a visualized content, and thus improve the smoothness of the motion parallax. However, interpolation techniques by definition are estimations, and the inaccuracies of the interpolation can lead to degradations of visual quality, e.g., artefacts. This creates a trade-off for visual content interpolation on light field displays: on one hand, an original content with low angular resolution can perceptually benefit from the higher number of views, but on the other hand, the possible degradations in visual quality can reduce user satisfaction. Both of these factors are vital components of the Quality of Experience (QoE) achievable on light field displays, and the preference may vary between observers; while for some, the smoothness of the 3D experience may be more important, some others may value the visual quality of the content more.

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 676401, European Training Network on Full Parallax Imaging and No 643072, Network QoE-Net.

In this paper, we present our results obtained from a series of subjective tests that investigated the performance of selected interpolation techniques on a light field display. The main scope of this paper is to answer the following primary research questions: (1) Is there a perceivable difference between visual contents with genuine high angular resolutions and interpolated contents with the same angular resolution? If there is, how significant is it? (2) How much can the perceived quality of a given content with low angular resolution benefit from interpolation? How different are the inputs and the outputs of the investigated techniques regarding visual quality? (3) Which interpolation technique performs better when they are subjectively assessed by human observers on a real light field display?

The remainder of this paper is structured as follows. Section 2 provides a brief overview of the relevant related work, followed by the detailed introduction of the interpolation techniques used in our experiment in Section 3. The experiment itself is described in Section 4, results of which are presented in Section 5. The paper is concluded in Section 6, which also points out potential continuations of the research topic.

2. RELATED WORK

In the area of imaging and cameras, interpolation traditionally referred to an operation between the pixels of a given image, facilitating digital zooming, anti-aliasing and creating various forms of image distortions. One of the earliest methods to use interpolation in the sense of computing an image captured by a virtual camera from existing camera images was published by Chen and Williams in '93 [1]. Since then, the technique of image interpolation has been used for several purposes and it is a basic tool in image-based rendering.

One obvious use of image interpolation is making motions in animations and video streams smoother. This technique is called motion-compensated frame interpolation [2]. It is used to make motion more continuous in low-frame-rate videos and also for increasing the efficiency of video compression algorithms [3].

In the work of Dricot *et al.* [4], specific views were skipped

and then replaced with synthesized ones, and the introduced visual degradations and artefacts were assessed during subjective assessments. The work of Kara *et al.* [5] compares the effects of low angular resolution and reduced visual quality on user experience. In the research, the view interpolation was performed on rectified camera images using inpainting in the shearlet domain [6]. The results indicate subjective observer sensitivity towards both aspects of quality, and shows that on the selected visual stimuli, the degradations introduced by the shearlet transform were more difficult to tolerate than the angular resolution of 1.5-2 views per degree. However, on lower angular resolutions, light field reconstruction performed similarly. In case depth information is additionally available, depth image-based rendering (DIBR) view synthesis can be used, as in the work of Bosc *et al.* [7].

The so-called "holy grail" of image interpolation is the streamable free-viewpoint video. This means that a scene is captured with cameras from different angles, then streamed to the viewer, who can choose an arbitrary viewpoint to view the scene from. An excellent example for this is a sports event, in case of which cameras are located all around the stadium, but the sports fans would like to watch the events live from a specific point of the field from close distance and in high detail. A lot of effort has been done towards this goal [8][9][10].

3. INVESTIGATED INTERPOLATION TECHNIQUES

3.1. Disparity based interpolation

Disparity is the row-wise difference between adjacent 2D images [11]. Disparity calculation between two images transforms one rectified image row-wise into the other, in order to allow such algorithms to run in real time. Rectification refers to the method of transforming the camera coordinate system of two adjacent cameras into a new camera system, so that the camera directions are equal and they enclose a right angle with the line connecting the camera positions. The new coordinate systems are chosen so that the epipolar lines of the original camera images are falling on the same horizontal pixel rows in the new camera images. A virtual image can be interpolated between the two rectified cameras images, by finding the linear interpolation value based on the distance from the left and the right camera and interpolating the left and right disparity values to transform the pixels of both source images. In the final step these two images are blended together based on the interpolation value. Due to the simple per pixel shifts of color, the disparity based interpolation can be implemented efficiently on massively parallel architectures (e.g., GPU). In this paper, test materials that are outputs of the disparity based interpolation are denoted with D .

3.2. Sweeping planes based interpolation

A more sophisticated method of image interpolation is the plane sweeping approach [12][13]. The essence of this tech-

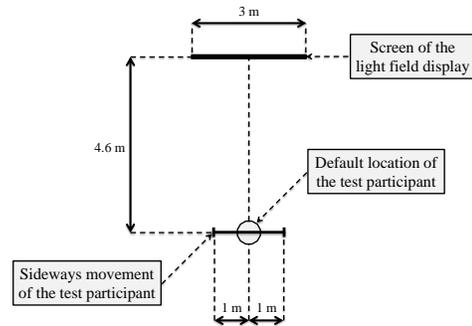


Fig. 1: Viewing conditions of the experiment.

nique is that we put planes perpendicular to the average of camera directions to different depth levels in front of the cameras and we project the colors seen by the cameras on these. By evaluating the color match consistency of this projection, for every pixel in the cameras we find the most probable depth level where the light seen by these pixels was emitted from. Then the interpolation is the reprojection of these depth planes onto the new virtual camera. This operation is computationally expensive as the projections have to be evaluated for all adjacent camera pairs on many planes. In this paper, test materials that are outputs of the sweeping planes based interpolation are denoted with S .

4. EXPERIMENTAL SETUP

4.1. Display and test environment

The subjective assessment of quality was carried out on Holografika's HoloVizio C80 light field cinema system¹, which is a 3-meter wide projection-based light field display with a brightness of 1500 cd/m². The tests were performed in an isolated laboratory environment with lighting conditions of 20 lx. The display itself was calibrated in a way to enable a 45-degree field of view.

The test participants observed the screen of the display from a distance of 4.6 meters (see Figure 1), which corresponded to 2.5H, as the height of the screen was 1.85 m. From the middle position on the line of observation, the test participants had to move a meter to the left and to the right during the visualization of a test stimulus, in order to fully experience the notion of horizontal motion parallax.

4.2. Visualized source content

The experiment involved 4 source stimuli, all of which were still image models in front of a plain background (see Figure

¹HoloVizio C80 light field cinema system, www.holografika.com/Documents/HoloVizio_C80.pdf

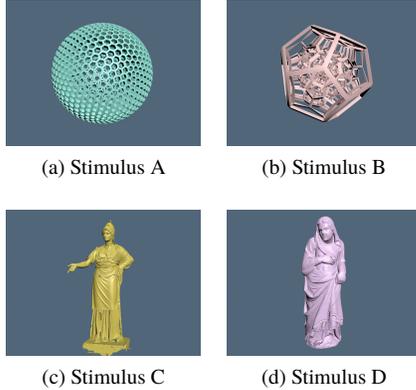


Fig. 2: The source stimuli.

2). Two of the models were mathematical bodies² and the other two were laser-scanned statues³. We used a linear setup of virtual cameras to render the stimuli in given angular resolutions, which were then used as inputs of the interpolation techniques. All the visualized stimuli in the experiment had the same spatial resolution of 1440×1080 .

Stimulus A was a polyhedron with 972 faces. The reason why we chose this complex mathematical body is that even though the far side of the model was partially visible through the front grid, its interpolation was challenging due to the occluded areas. Stimulus B was also a complex mathematical body, namely a structure of 120 regular dodecahedra. Here the fine-grained details of the central structure of the model and also the occlusions posed the greatest challenge, while the frontal part in the given orientation – closest to the observer – had a simple surface. Stimulus C was a laser scan of an actual statue. Beyond the surface of the thin, low-depth body, the out-reaching arm was also a notable component, as it reacted differently to reduced angular resolution than the torso due to depth distances. Stimulus D was also a laser-scanned statue, with different visualization parameters. Although there was no specific part of the model standing out from the whole – like the arm of the previous statue – depth values were greater and there were deeper wrinkles on the surface.

In this paper, test materials that are directly rendered in given angular resolutions and used as interpolation inputs are denoted with R . The interpolated test materials are denoted differently – depending on the selected interpolation techniques – as specified earlier.

4.3. Test conditions

As defined in previous subsections of the paper, we used three main types of stimuli in our experiment: the directly rendered stimuli (R), and the outputs of disparity based interpolation

²George W. Hart's Rapid Prototyping Web Page www.georgehart.com/rp/rp.html

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

Type	Values	Respective views
R(rendered)	10, 20, 45	10, 20, 45
D(isparity)	10, 20, 45	181, 191, 221
S(sweeping)	10, 20, 45	181, 191, 221
G(round truth)	181, 191, 221	181, 191, 221

Table 1: Investigated test conditions.

(D) and sweeping planes based interpolation (S). As interpolation was performed on 3 different angular resolutions (10, 20 and 45 views) which resulted in 3 increased angular resolution (181, 191 and 221 views, respectively), we also rendered stimuli with the corresponding high angular resolutions as ground truth – denoted with G in the paper. This defines a total of 12 different visual stimuli (see Table 1). We chose angular resolutions at 1 view per degree (45 views in this case) and below, as such low values can provide poor perceived quality [14] and can benefit from interpolation.

In this paper, the identification numbers for the interpolated stimuli (D and S) refer to the number of views in the input, and the numbers for the rendered stimuli (R) and the ground truth (G) determine the number of views in which the content was rendered. As an example, $D45$ means that the input consisting of 45 views ($R45$) was interpolated based on disparity, that resulted in 221 views, which has the same angular resolution as $G221$, but may have a different, degraded visual quality.

4.4. Interpolation outputs

In general, we can state that the investigated interpolation algorithms performed well; Figure 3 is a good example, as the difference between the ground truth views and the interpolated views are minimal. Artefacts were only introduced in two cases: in case of occluded visuals, where the interpolation algorithms had no information to work with (see Figure 4) and in case of the disparity based interpolation, when the available views were insufficiently low and the color difference between adjacent views were too high, thus the algorithm could not always find the appropriate color correspondences.

4.5. Subjective quality assessment task and structure

The task of the test participants was to compare the test conditions for each source content during a single-stimulus paired comparison. A seven-point scale from ITU-R Rec. BT.500-13 was used [15], with points *Much worse*, *Worse*, *Slightly worse*, *Same*, *Slightly better*, *Better* and *Much better*. However, comparison was performed on two identical instances of this scale, comparing different aspects of quality. One focused particularly on the visual quality of the visualized object (i.e., considering degradations through artefacts) and the other separately rated the 3D performance and the smoothness of

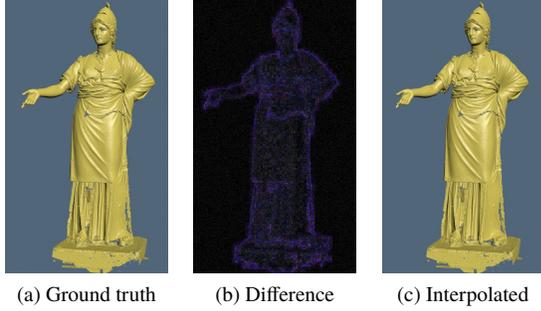


Fig. 3: Highlighted difference (1600%) at the center view of D45 on stimulus C.

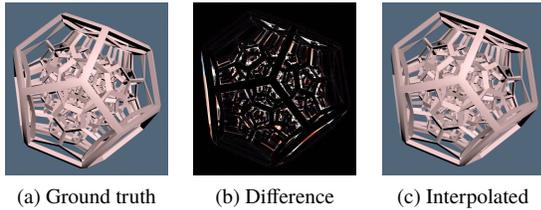


Fig. 4: Highlighted difference (400%) at the center view of S20 on stimulus B.

the horizontal motion parallax (i.e., considering degradation through reduced angular resolution).

Measuring all factors on a single scale would have provided a better insight to effect on the whole; how the combination of the two shape the overall user experience. However, by separating them, we can directly analyze the possible negative impacts of interpolation artefacts and the benefits of the higher angular resolutions.

Each stimulus was visualized for 10 seconds. The stimuli in the investigated pairs were separated with a 5-second long black screen, and there was a 10-second blank screen after the second stimulus in the pair, during which the rating on the two scales was performed. As each pair had a duration of approximately 35 seconds (with rating included), and there were 60 pairs in the experiment (15 pairs for 4 source stimuli), thus the total duration for a subjective test was roughly 35 minutes. The tests were conducted without breaks.

4.6. Investigated test condition pairs

The previous subsection stated that there were 60 comparisons in total, with 15 different condition pairs. In the experiment, we compared each interpolated stimulus (IS) with the other type of interpolation (C-IS), the corresponding rendered stimulus with same high angular resolution (G) and the input of the interpolation technique with the original low angular resolution (R). This would define 18 pairs (see Table 2); however, 3 of them comparing the performance of the two interpolation techniques are identical, which leaves us 15 comparisons.

IS	C-IS	G	R
D45	S45	G221	R45
S45	D45	G221	R45
D20	S20	G191	R20
S20	D20	G191	R20
D10	S10	G181	R10
S10	D10	G181	R10

Table 2: Investigated test condition pairs.

4.7. Test participants

A total of 21 test participants completed the quality assessment task, with an average age of 31. From these observers 16 were male and 5 were female. None of the test participants were experts; only naïve human observers participated.

5. RESULTS

In this section of the paper, we first introduce real camera captures of the test stimuli visualized on the light field display, and then present the results of the subjective quality assessment experiment. The aim of this dissemination of our results is to address our research questions via the mean subjective comparison scores, but in-depth statistical analysis, content-dependent result separation, correlation calculation and scoring pattern analysis are not included in the scope of this paper.

5.1. Visualized interpolations

We took pictures of the light field display with a DSLR camera from the default location of the test participants (4.6 meters from the screen). Figure 5 illustrates the artefacts introduced by both algorithms in the occluded areas, as it was seen by the participants. Figures 6 and 7 show the cases, where the disparity based interpolation generated artefacts due to the low angular resolution of the input image set. These examples also show that the sweeping planes interpolation was able to maintain image quality and increased the angular resolution.

5.2. Subjective assessment results

The subjective quality assessment experiment gathered scores on two comparative scales for the smoothness of the horizontal motion parallax and for the visual quality, as specified earlier. In this analysis, we refer to the smoothness of the horizontal motion parallax as "motion" – especially since test participants were particularly instructed to move to the left and to the right during assessment – and to the visual quality of the visualized object simply as "quality". The 95% confidence intervals of the data points are not displayed on the graphs. Their values are all within the [0.23, 0.63] range with an average value of 0.40.

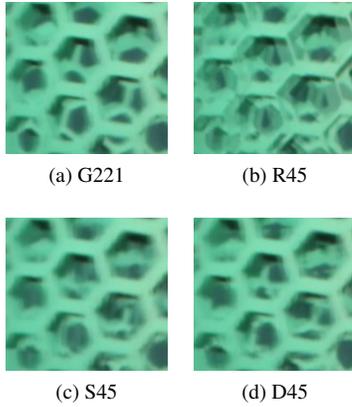


Fig. 5: Camera capture of the grid of stimulus A.

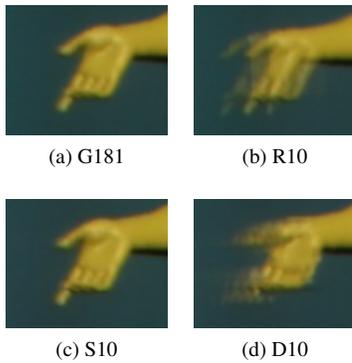


Fig. 6: Camera capture of the statue's out-reaching hand of stimulus C.



Fig. 7: Camera capture of the statue's head of stimulus D.

Both algorithms performed at an acceptable level compared to the ground truth when the input image set 45 views (Figure 8). At lower angular resolutions, the artefacts introduced by the disparity based interpolation decreased the mean comparison scores rapidly.

Compared to the input image sets, we can state that the sweeping planes interpolation always improved the visual ex-

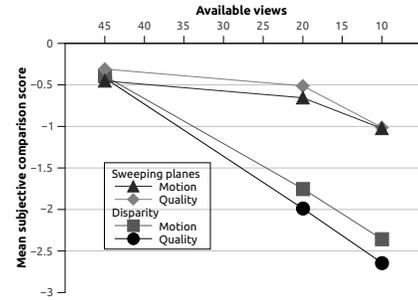


Fig. 8: Interpolated stimuli (IS) compared to the corresponding ground truths (G).

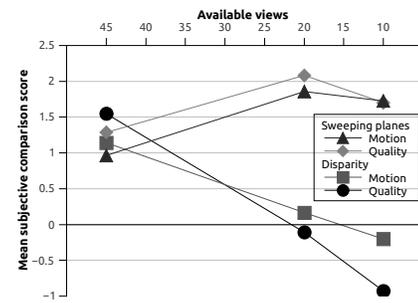


Fig. 9: Interpolated stimuli (IS) compared to the rendered interpolation inputs (R).

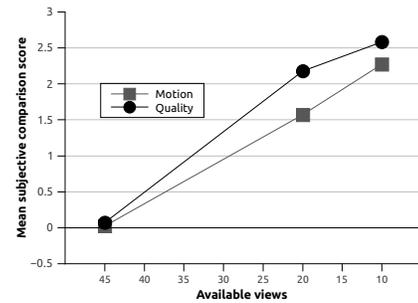


Fig. 10: Sweeping planes based interpolation (S) compared to disparity based interpolation (D).

perience, both motion and quality (see Figure 9). Based on these results, we can also conclude that disparity based interpolation requires higher angular resolution image sets to be beneficial.

The direct comparison of the techniques clearly shows the difference at the lower angular resolution image sets, as the sweeping planes based interpolation performed better (see Figure 10). For 45-view inputs, even though they do not differ directly, disparity based interpolation improved both aspects more (see Figure 9).

6. CONCLUSIONS

In this paper, we presented our research results on interpolation techniques for light field visualization. Our findings indicate that the sweeping planes based interpolation can preserve image quality at low input angular resolutions, while significantly improving the smoothness of the horizontal motion parallax. The disparity based interpolation can also benefit the visual experience, however, it requires a sufficiently dense input image set. From the two techniques, the sweeping planes technique performed better in case of sparse inputs, but otherwise there is no apparent difference in performance. In future dissemination of this research, we aim to provide a detailed analysis of the results, for each source stimuli. A potential continuation of the research is the development and analysis of further techniques, specially designed for light field displays, and the performance comparison of 3D reconstruction and interpolation algorithms. The computational requirements of such operations are also worth investigating, as a long-term goal in this area is to perform real-time content enhancement for future light field services.

7. REFERENCES

- [1] Shenchang Eric Chen and Lance Williams, "View interpolation for image synthesis," in *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*. ACM, 1993, pp. 279–288.
- [2] Tien-ying Kuo, JongWon Kim, and C-CJ Kuo, "Motion-compensated frame interpolation scheme for H. 263 codec," in *Proceedings of the 1999 IEEE International Symposium on Circuits and Systems, 1999. ISCAS'99*. IEEE, 1999, vol. 4, pp. 491–494.
- [3] Yan Ye and Marta Karczewicz, "Improved H. 264 intra coding based on bi-directional intra prediction, directional transform, and adaptive coefficient scanning," in *15th IEEE International Conference on Image Processing, 2008. ICIP 2008*. IEEE, 2008, pp. 2116–2119.
- [4] Antoine Dricot, Joel Jung, Marco Cagnazzo, Béatrice Pesquet, Frédéric Dufaux, Péter Tamás Kovács, and Vamsi Kiran 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.
- [5] Peter A Kara, Peter T Kovács, Suren Vagharshakyan, Maria G Martini, Attila Barsi, Tibor Balogh, Aleksandra Chuchvara, and Ahmed Chehaibi, "The Effect of Light Field Reconstruction and Angular Resolution Reduction on the Quality of Experience," in *12th International Conference on Signal Image Technology & Internet Based Systems (SITIS) 3rd International Workshop on Quality of Multimedia Services (QUAMUS)*, 2016.
- [6] Suren Vagharshakyan, Robert Bregovic, and Atanas Gotchev, *Image Based Rendering Technique via Sparse Representation in Shearlet Domain*, pp. 1379–1383, The Institute of Electrical and Electronics Engineers, Inc., 9 2015.
- [7] Emilie Bosc, Romuald Pepion, Patrick Le Callet, Martin Koppel, Patrick Ndjiki-Nya, Muriel Pressigout, and Luce Morin, "Towards a new quality metric for 3-d synthesized view assessment," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 7, pp. 1332–1343, 2011.
- [8] Joel Carranza, Christian Theobalt, Marcus A Magnor, and Hans-Peter Seidel, "Free-viewpoint video of human actors," in *ACM transactions on graphics (TOG)*. ACM, 2003, vol. 22, pp. 569–577.
- [9] Naho Inamoto and Hideo Saito, "Virtual viewpoint replay for a soccer match by view interpolation from multiple cameras," *IEEE Transactions on Multimedia*, vol. 9, no. 6, pp. 1155–1166, 2007.
- [10] Alvaro Collet, Ming Chuang, Pat Sweeney, Don Gillett, Dennis Evseev, David Calabrese, Hugues Hoppe, Adam Kirk, and Steve Sullivan, "High-quality streamable free-viewpoint video," *ACM Transactions on Graphics (TOG)*, vol. 34, no. 4, pp. 69, 2015.
- [11] Stephen T Barnard and William B Thompson, "Disparity analysis of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, , no. 4, pp. 333–340, 1980.
- [12] Ruigang Yang, Greg Welch, and Gary Bishop, "Real-time consensus-based scene reconstruction using commodity graphics hardware," in *Computer Graphics Forum*. Wiley Online Library, 2003, vol. 22, pp. 207–216.
- [13] Sergey Smirnov, Mihail Georgiev, and Atanas Gotchev, "Comparison of cost aggregation techniques for free-viewpoint image interpolation based on plane sweeping," *Ninth International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, 2015.
- [14] Peter A Kara, Maria G Martini, Peter T Kovács, Sandor Imre, Attila Barsi, Kristof Lackner, and Tibor Balogh, "Perceived quality of angular resolution for light field displays and the validity of subjective assessment," in *International Conference on 3D Imaging (IC3D)*. IEEE, 2016.
- [15] ITU-R Rec., "BT.500-13: Methodology for the subjective assessment of the quality of television pictures," 2012.