



Viva la Resolution: The Perceivable Differences between Image Resolutions for Light Field Displays

Peter A. Kara¹, Peter T. Kovacs^{2, 3}, Maria G. Martini¹, Attila Barsi², Kristof Lackner², Tibor Balogh²

¹ Wireless Multimedia and Networking Research Group, Kingston University, London, UK

² Holografika, Budapest, Hungary

³ Department of Signal Processing, Tampere University of Technology, Tampere, Finland

{p.kara, m.martini}@kingston.ac.uk, {p.kovacs, a.barsi, k.lackner,
t.balogh}@holografika.com, peter.t.kovacs@tut.fi

Abstract

As 3D display technologies are becoming more and more common in commercial, everyday usage, a special type of 3D known as projection-based light field displays is emerging as well. While holding many key characteristics such as field of view or angular resolution, traditional image resolution also plays a major role in the overall determination of user experience, similarly to 2D displays. The paper investigates the perceivable differences between display resolutions, and presents the acceptability of resolution degradation should it be visible. A total of 20 test participants provided subjective assessment in a series of pair comparisons between 5 resolutions for various stimuli. Beyond mean scores, the results are presented in terms of score distribution in this analysis, separately for each and every stimulus.

Index Terms: Quality of Experience, light field display, image quality, image resolution, subjective pair comparison

1. Introduction

Visual information is the primary source of information for humans, making displays crucial terminal devices in use today. While most of these displays can represent 2D (flat) images, displays with 3D capabilities are also available commercially. One can find 3D capabilities in televisions, mobile devices, and predominantly in cinemas. 3D perception is achieved by reproducing the binocular effect in the observer's eyes, coupled with motion parallax in more advanced displays. Autostereoscopic (glasses-free) displays can be implemented using various techniques, such as multiview displays based on lenticular lens sheets [1] or light field displays based on multi-projection [2–5].

Potential users of 3D displays rightfully expect to know what the capabilities of the displays are, like they know in case of 2D displays, described by metrics such as resolution, brightness or contrast. However, most of these metrics are not directly applicable to 3D displays, as the angular dependence of light color and intensity renders usual measurement methods useless, or at least how they shall be performed is unclear.

Our research studies the perception of image resolution of light field displays (projection-based light field displays in particular). The research question at hand is the perceivable difference between different resolutions for a given light field

display and the subjective tolerance of resolution reduction. We performed a series of subjective quality assessment measurements, where the assignment of the human observers was to evaluate the alterations in still image quality (gained through varying resolution) in a pair comparison task.

The paper is structured as follows: Section 2 presents prior relevant researches in the area, followed by the technical background of our own research in Section 3. The results and findings of our work are detailed in Section 4, and concluded in Section 5.

2. Related work

Projection-based light field displays emit light rays from a multitude of locations, originating from a large number of optical engines. As these light rays cross the projection screen, the hit points might not necessarily be at discrete positions on a regular grid; in practice, the light rays are hitting the screen at irregular positions, rendering the notion of resolution as used in case of 2D displays inapplicable. Also, the total angle in which light rays are emitted from the screen surface determine the viewing angle of the display, that is, the angular range from which viewers can observe the image from. Both of these factors have a strong effect on the perceived 3D image quality, as noted by Bowman et al. [6], who listed field of view and spatial resolution as important quality characteristics of display devices when used as part of 3D user interfaces.

Previous work in this topic discusses the estimation of resolution of lenticular lens-based multiview displays, such as the work of Boev et al. [7] who use frequency domain analysis of the response of the display given to specific sinusoidal test patterns. The work of Boher et al. [8] addresses the creation of measurement equipment that can be used to precisely characterize 3D displays. These measurement tools have been commercialized, however, they are designed to work with multiview displays, and are not applicable to light-field ones.

The International Display Measurement Standard [9] created by the Society for Information Display provides procedures and guidelines for measuring the resolution of light field displays from a single point of view, but it does not discuss how the measured resolution affects the perceived quality. We have also seen in previous works that the stereoscopic effect and stereoscopic image resolution both improve the feeling of presence when performing remote navigation tasks [10].

3. Research setup

3.1. Investigated resolutions

The pair comparisons were performed for the following source image resolutions (not to be confused with final resolution):

- 854x480 (WVGA)
- 1024x576 (PAL)
- 1280x720 (720p)
- 1920x1080 (1080p)
- 3840x2160 (2160p)

Final resolution is taken on a regular grid per optical module, and is the overall pixel count emitted from the optical modules. This is independent of the source resolution. The resulting light field, however, is irregular and uneven due to the optical properties of the screen and the manufacturing process. The calibration process required to compensate for physical inaccuracies and irregularities mixes up the pixel taken from the regular grid of the source image. The optical properties of the screen generate a light field that has continuous transitions horizontally (the pixels are blended horizontally with overlapping Gaussian distributions). Therefore resolution is viewer-position-dependent, irregular and not discretized. It is very hard to measure and it cannot be compared to resolution on 2D displays in the traditional sense.

3.2. Subjective pair comparison

Subjective assessment data was collected on a 5-point Degradation Category Rating Scale (DCR) [11]:

- Imperceptible
- Perceptible but not annoying
- Slightly annoying
- Annoying
- Very annoying

All possible combinations were evaluated, including self-comparisons, during which no perceptible difference is expected. Due to the type of the scale, the reference stimulus always had equivalent or higher resolution. With 5 selected resolutions, this means 15 pair comparisons in total (2160p was paired 5 equivalent or lower resolutions, 1080p was paired with 4 equivalent or lower resolutions etc.).

3.3. Research measurement environment

The measurements were carried out in a controlled environment, with lighting conditions of 25 lux. The images were displayed on a projection-based light field display, namely a HoloVizio C80 3D cinema system [12], which has 40 degrees of horizontal viewing angle and brightness of 1500 cd/m². A total of 20 test participants completed the evaluation task, with the average age of 26.

The participants were not fixed to a single viewing position, but could view the stimuli from different angles and distances, before making assessment. The closest they could go to the 4-meter-wide display was 4.6 meters, and this was also a mandatory distance from which each stimulus was to be observed, before increasing the viewing distance by 2 meters at most (see Figure 1).

3.4. Test stimuli

The previously specified 15 pair comparisons were tested on 3 source image sets – rendered directly in the given resolutions – so each test participant had to provide 45 DCR scores. These light field images were created by rendering the scene to a virtual camera array, using 180 camera positions. The source images – referred to as stimuli – were converted (resampled) to the display's light field. Stimulus A was a collection of simple shapes with plain-colored textures, stimulus B had more complex texture and stimulus C had a fine-detailed structure (see Figure 2).

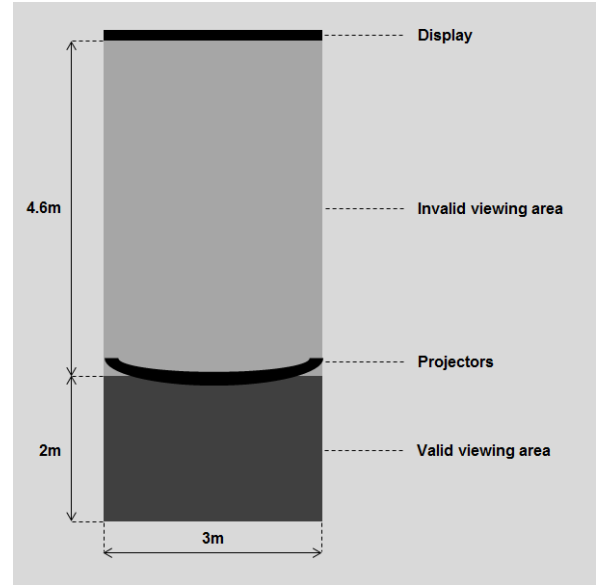


Figure 1: Test measurement setup.

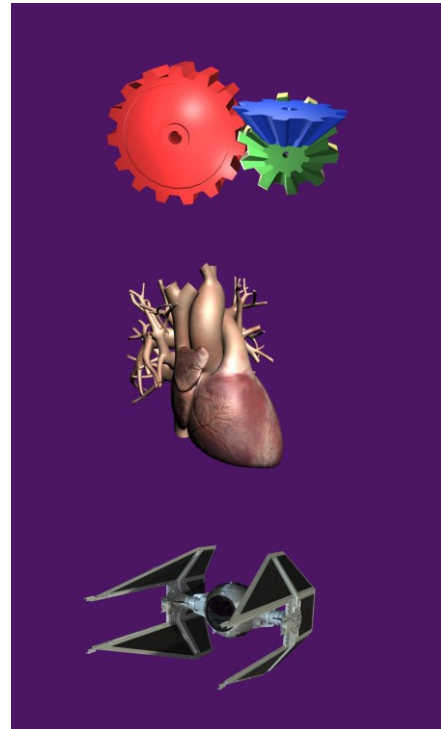


Figure 2: Stimulus A (top), B (middle) and C (bottom).

3.5. Subjective test structure

Each source was first introduced to the test participant in the reference resolution of 3840x2160 (which was not paired with any other and was not to be assessed), and then the test cases were evaluated in a random order. The test cases were clustered, meaning that the random pair comparisons were organized by source. The observer was notified every time the displayed image had changed, since images in the comparison pairs were not displayed next to each other simultaneously, but were shown directly after each other.

A single image was shown for 10 seconds, and subjective evaluation of quality degradation had approximately the same duration, so the total duration of an assessment measurement was around 25 minutes for a test participant.

The test participants were not given any information about the stimuli; not only image resolution, but other parameters – e.g. brightness or angular resolution – were unavailable to the participants as well.

4. Results

4.1. Overall analysis

Based on the information provided in the research setup, a total of 900 DCR scores were collected (20 participants providing 15 scores each for 3 stimuli). The mean scores (see Figure 3) offer a quick glimpse at the results of the pair comparisons, however, their interpretation should not be confused with the Mean Opinion Score (MOS) of Absolute Category Rating (ACR) scores [11]. The diagonal values indicate that self-comparisons suffered an extent of distortion – which is approximately the same for all, a tad lower at higher resolutions – meaning that some participants experienced perceptible differences, even though the compared stimuli were identical. It can also be seen that shifting between the top 3 does not reach “slight annoyance” in general.

A better insight can be obtained observing the score distribution (see Figure 4). Each comparison was assessed 60 times during the measurement, since 20 test participants evaluated 3 stimuli. The first thing that can be observed – as suggested by the mean scores – is that self-comparison is far from optimal; more than one fourth of these test cases were given scores reflecting perceptible degradation. This can be explained by the fact that one third of all comparisons were identical (5 self-comparisons), and also some comparisons contained no noticeable difference, depending on the individual of course. In these cases perception should not be able to distinguish images, however, since the test participant expects to see differences, the preconception overwrites perception and allows degradation to be seen [13-14], according to the dissonance reduction during cognitive dissonance [15]. The level of such degradation may be above Just Noticeable Difference (JND), and can even result annoyance (“Slightly annoying” and “Annoying” scores).

Throughout the measurement, 40 out of 60 times the test participants were not able to distinguish 2160p and 1080p, and this number is 20 when compared with 720p. The comparison with the lower resolution invokes annoyance to considerable extents, but even in the evaluation of the difference between 2160p and PAL, nearly half of the scores are “Imperceptible” and “Perceptible but not annoying”.

	2160p	1080p	720p	PAL	WVGA
2160p	4.68	4.65	4.02	3.33	2.53
1080p	4.65	4.76	4.45	3.75	2.78
720p	4.02	4.45	4.6	4.08	3.27
PAL	3.33	3.75	4.08	4.6	3.66
WVGA	2.53	2.78	3.27	3.66	4.56

Figure 3: Mean scores of the pair comparisons.

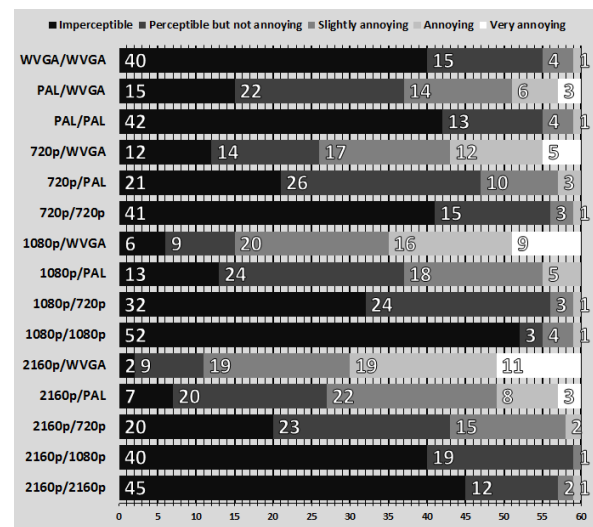


Figure 4: Score distribution of the subjective assessment.

Comparing 2160p with itself and with 1080p resulted in very similar scores, meaning that the cognitive biases in the self-comparison of 2160p and the actually perceivable difference between 2160p and 1080p (plus any bias) have the same effect on experienced quality.

It is interesting to see that the comparison of 2160p and 720p, and of 720p and PAL have a rather similar score distribution. Also, it is apparent that the higher the resolution is, the harder it is to distinguish adjacent resolutions and the more tolerable the perceived difference is.

4.2. Individual analysis per source

The results for stimulus A (see Figure 5) and B (see Figure 6) show some similarities in general, especially at higher resolutions; in fact, the scores of the 2 highest resolution comparisons (2160p/2160p and 2160p/1080p) are almost exactly the same (a difference of a single quality comparison rating). Regarding these test cases, it is quite intriguing to see that comparing with 1080p actually received better scores than the self-comparison of 2160p. Based on these scores, it can be stated that there is no perceivable difference between 2160p and 1080p, at least for these sources.

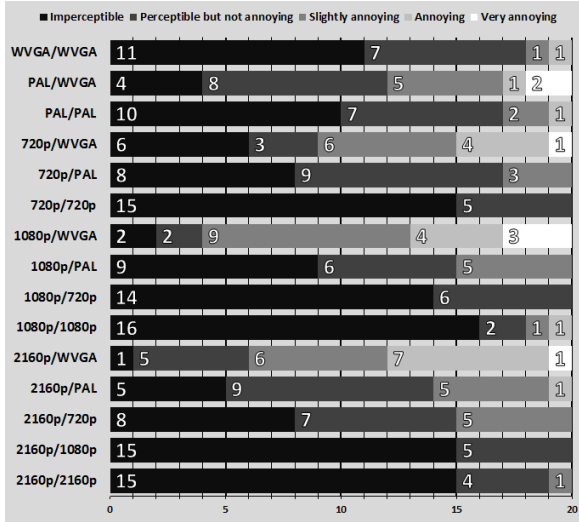


Figure 5: Score distribution of stimulus A.

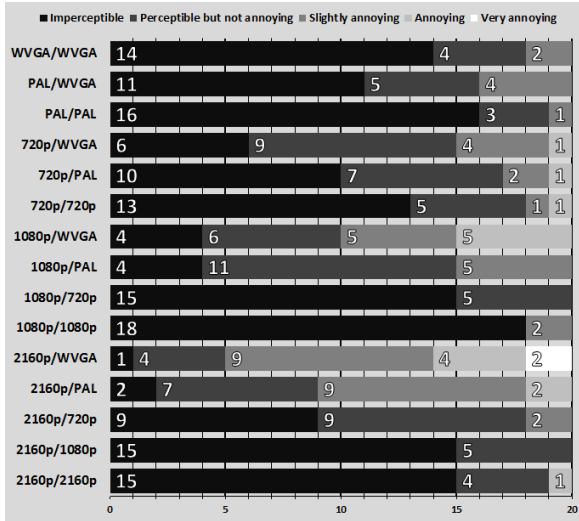


Figure 6: Score distribution of stimulus B.

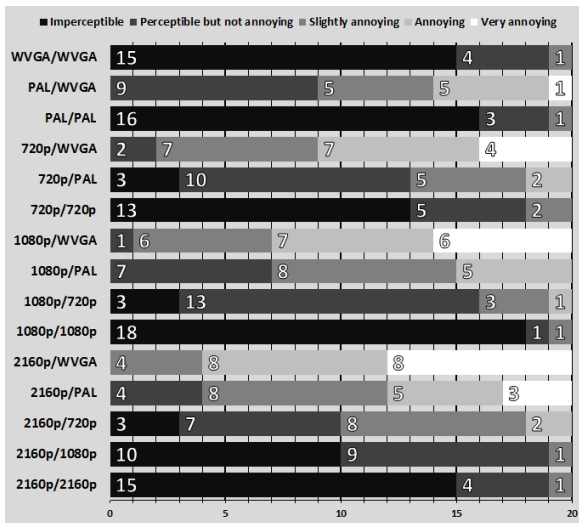


Figure 7: Score distribution of stimulus C.

More significant dissimilarities between A and B can only be found at lower resolutions, with differences between resolutions easier to detect during the comparisons of stimulus A. This implies that the uniform straight object edges in stimulus A gained more attention during resolution reduction than the textures of stimulus B.

As for stimulus C (see Figure 7), even though the self-comparison of 2160p is the same as for A and B, results differ when compared with 1080p. In general the scores given for stimulus C present a notable increase in sensitivity towards resolution reduction compared to the other stimuli. While in case of 2160p/1080p and 1080p/720p, 75% of the test participants deemed the difference to be “Imperceptible” for stimulus A and B, half of the scores for test case 2160p/1080p was perceptible for stimulus C, and only 3 out of 20 times was the reduction in 1080p/720p undetected. The mean scores of the measurement (see Figure 3) indicate that the comparison of 2160p and 720p is assessed to be “Perceptible but not annoying”, half of the scores of stimulus C have some extent of annoyance. The reason for the more penalizing evaluation of stimulus C is the detailed structure of the object; fine structural details get blurry or vanish during resolution reduction. The tendencies and relations between test cases follow the patterns of stimulus A and B, but in general the differences are easier to detect and cause more irritation at the user side.

In general, the measurement provided a great diversity in DCR scores; specific test cases were rated with all possible assessment values, meaning that while the difference was very annoying for one person, it remained unseen for another.

5. Conclusions

The paper has presented the results obtained from the subjective assessment of perceivable differences during image resolution reductions for light field displays. The performed image comparisons with the considered rendered content indicate that on such displays 2160p and 1080p resolutions are either hardly distinguishable or distinguishable but completely tolerable. We can also conclude that 720p is an adequately good resolution for specific use cases, since the difference can be rather hard to tell from higher resolutions. Our findings also include that even the lowest resolutions can be acceptable, as only half of the test observers found it considerably annoying in comparison with the highest available resolution. Further results, separately for source images, show that the structural complexity of displayed objects is significantly more determining in the subjective evaluation of resolution degradation than textures, increasing sensitivity towards differences and thus enables a higher level of overall annoyance. Our future work shall take angular resolution into consideration as well, which is responsible for continuous motion parallax.

6. Acknowledgements

The work in this paper was funded from the European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 643072, Network QoE-Net. The research leading to these results has received funding from the PROLIGHT-IAPP Marie Curie Action of the People programme of the European Union’s Seventh Framework Programme, REA grant agreement 32449.

7. References

- [1] C. van Berkel, D. W. Parker, A. R. Franklin, "Multiview 3D-LCD," *Proc. SPIE, Stereoscopic Displays and Virtual Reality Systems III*, vol. 2653, 1996, pp. 32-39.
- [2] T. Balogh, "The HoloVizio system," *Proc. SPIE, Stereoscopic Displays and Applications XIII*, vol. 6055, 2006.
- [3] G. Wetzstein, D. Lanman, M. Hirsch, R. Raskar, "Tensor Displays: Compressive Light Field Synthesis using Multilayer Displays with Directional Backlighting," *Proc. SIGGRAPH*, 2012.
- [4] J. H. Lee, J. Park, D. Nam, S. Y. Choi, D. S. Park, C. Y. Kim, "Optimal Projector Configuration Design for 300-Mpixel Light-Field 3D Display," *Optics Express*, vol. 21, issue 22, 2013, pp. 26820-26835.
- [5] M. Kawakita, S. Iwasawa, R. Lopez-Gulliver, M. Makino, M. Chikama, M. P. Tehrani, "Glasses-free 200-view 3D Video System for Highly Realistic Communication," *Proc. Digital Holography and Three-Dimensional Imaging*, OSA Technical Digest, paper DM2A.1, 2013.
- [6] D. A. Bowman, E. Kruijff, J. J. LaViola Jr., I. Poupyrev, "3D User Interfaces: theory and practice," Addison Wesley Longman Publishing Co., Inc., Redwood City, CA, USA, 2004.
- [7] A. Boev, R. Bregovic, A. Gotchev, "Measuring and modeling per-element angular visibility in multi-view displays," *Journal of the Society for Information Display*, vol. 18, issue 9, 2010, pp. 686-697.
- [8] P. Boher, T. Leroux, T. Bignon, V. Colomb-Patton, "A new way to characterize autostereoscopic 3D displays using Fourier optics instrument," *Proc. SPIE, Stereoscopic Displays and Applications XX*, vol. 7237, 2009.
- [9] Society for Information Display, "The International Display Measurement Standard v1.03," 2012.
- [10] S. Lee, G. Kim, "Effects of haptic feedback, stereoscopy, and image resolution on performance and presence in remote navigation," *International Journal of Human-Computer Studies*, vol. 66, issue 10, 2008, pp. 701-717.
- [11] ITU-T Rec. "P.910: Subjective video quality assessment methods for multimedia applications," 2008.
- [12] HoloVizio C80 3D cinema system
<http://www.holografika.com/Products/NEW-HoloVizio-C80.html>
(retrieved June 2016)
- [13] A. Sackl, P. Zwickl, S. Egger, P. Reichl, "The role of cognitive dissonance for QoE evaluation of multimedia services," *IEEE Globecom Workshops (GC Wkshps)*, 2012, pp. 1352-1356.
- [14] P. A. Kara, A. Sackl, M. Mourão, L. Bokor, "What your phone makes you see: Investigation of the effect of end-user devices on the assessment of perceived multimedia quality," *7th International Workshop on Quality of Multimedia Experience (QoMEX)*, Costa Navarino, Messinia, Greece, 2015, pp. 1-6.
- [15] L. Festinger, "A theory of cognitive dissonance," Stanford, CA: Stanford University Press, 1957.