

Automatic 3DTV Quality Assessment Based On Depth Perception Analysis

J.A. Rodrigo¹, J. P. López², D. Jiménez³, J.M. Menéndez⁴

^{1,2,3,4}Escuela Técnica Superior de Ingenieros de Telecomunicación, Universidad Politécnica de Madrid, Madrid, Spain

¹jrf@gatv.ssr.upm.es, ²jl@gatv.ssr.upm.es, ³djb@gatv.ssr.upm.es, ⁴jmm@gatv.ssr.upm.es

Abstract: Quality assessment is a key factor for stereoscopic 3D video content as some observers are affected by visual discomfort in the eye when viewing 3D video, especially when combining positive and negative parallax with fast motion. In this paper, we propose techniques to assess objective quality related to motion and depth maps, which facilitate depth perception analysis. Subjective tests were carried out in order to understand the source of the problem. Motion is an important feature affecting 3D experience but also often the cause of visual discomfort. The automatic algorithm developed tries to quantify the impact on viewer experience when common cases of discomfort occur, such as high-motion sequences, scene changes with abrupt parallax changes, or complete absence of stereoscopy, with a goal of preventing the viewer from having a bad stereoscopic experience.

Keywords: 3DTV, depth maps, zone of comfort, VQA, motion estimation, parallax.

1 INTRODUCTION

3D content is claiming importance in media environment. The success of new 3D services is a reality due to the improvement in technology, but visual comfort analysis is demanded. Quality issues are currently a bigger concern in 3D media than they were in traditional media. Although there are some impact factors and initial measurement methods in this field, there is still no common way and procedure to compare 3D video content and integrated solutions and obtain an evaluation of quality.

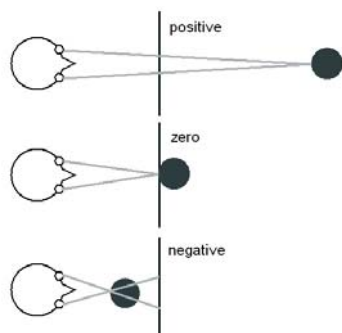


Figure 1. Parallax comparison

Stereoscopic 3D video perception is based on the fact that two different video signals are captured in order to feed to each of the viewer's eyes; recreating the experience of watching a real world scene, where two different images are captured by each eye and the difference between them depends on the position of the elements in the world related to the viewer's position. This means that the system is feeding the observer with a

disparity depth cue. Parallax created by disparity is determined by the virtual perceived location of the objects in a scene, as shown in Figure 1.

Watching 3DTV is significantly different from a natural view, as the point of view is prefixed by the fixed point of view of the camera lenses that have captured the scene, and is therefore the focus. Furthermore, in natural viewing, the eyes focus (accommodate) and converge to the same distance, but when looking at a 3D object displayed on a screen, a viewer's eyes must focus on the screen for a while, and at the same time, they converge on a point in space that may be located beyond the screen, on the screen, or in front of the screen. This is known as the vergence-accommodation conflict. This conflict limits the amount of parallax that a viewer can tolerate without suffering visual discomfort, also known as the Zone of Comfort [13].

This paper aims to study the effects of stereoscopic disparity in quality assessment through the analysis of depth maps of a sequence and its temporal evolution. We try to quantify objectively the effects of parallax, depth and motion, exporting the common situations in which discomfort is substantial, from opinions of observers derived from empirical and subjective tests.

Following, related studies are presented in Section 1 and the description of subjective assessment developed is compiled in Section 3. The proposed method is defined in Section 4. The test results are given in Section 5 and conclusions in Section 6.

2 RELATED WORK

Video quality assessment is a difficult process which plays a major role in various processing applications [1]. A lot of work has been developed in this field, defining metrics and algorithms to predict the quality of a video sequence. An overview of the extensive and most interesting work in quality assessment is collected in [2], [3] and [4]. This work is always related to subjective quality assessment and most of the published adhere to the procedures contained in the Recommendation ITU-R BT.500 [5].

In 3D media, new factors related to the optic effect of stereoscopy are concerned in order to assess quality; such as visual discomfort or perceptual inconsistencies between depth cues, as stated in [12]. Much work has also been developed in this field relating depth and motion, such as [6], where filtering is used to reduce visual discomfort on screens.

In [7], an overview describing the main topics relevant to comfort in viewing stereoscopic television is developed, analyzed after subjective tests, related to accommodation-vergence conflict, parallax distribution, binocular mismatches,

Corresponding author: J. A. Rodrigo, Universidad Politécnica de Madrid, Madrid, Spain, +34 913 367 344, jrf@gatv.ssr.upm.es

depth, and cognitive inconsistencies. In [8], it is reported that depth and motion are closely related in terms of calculating visual discomfort. And [9] offers a visual comfort model for detecting a salient object's motion features in depth of field.

An interesting subjective evaluation of visual discomfort is developed in [10], where parallax limits and regions of comfort, dependent on the screen size, disparity and viewing time, are obtained. Other artifacts such as stereo window violation (SWV) and temporal continuity of the disparity (TCD) have been studied in [11] where guidelines to create comfortable and faster stereoscopic films are included.

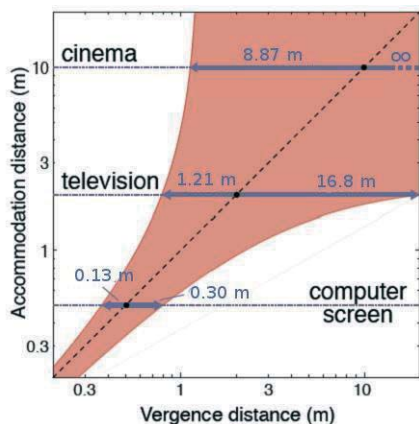


Figure 2. Shibata's Zone of Comfort

The Zone of Comfort (ZoC) was first introduced by Percival [14]. He suggested the limits to vergence-accommodation postures that could be achieved without causing discomfort. More recent studies such as Shibata et al. [15] concludes that the ZoC may differ from Percival's, when the experiments are based on stereoscopic vision rather than on vision through spectacles. In stereo vision the vergence-accommodation conflict constantly changes, while in a lens or spectacles system it is maintained fixed [13]. Figure 2 shows Shibata's ZoC for different accommodation distances in stereo vision. According to this diagram, images with positive parallax have little-to-no capability to induce discomfort, while negative parallax is most likely to cause discomfort if not controlled.

In order to adapt a stereoscopic ZoC to 3D video, it is necessary to take into account motion and time of exposure in a stereo scene. The ZoC will be further reduced when these elements appear. The time to converge and accommodate in this case is relevant, thus there is a need to adapt the concept of ZoC. In [10] the variation of time of exposure is studied in order to determine its effects on visual discomfort.

Determining an image parallax range uses an associated depth map. There are several ways to obtain it from a stereo image, depending on computing complexity and accuracy restrictions. As a rule of thumb it can be stated that complexity is proportional to accuracy, thus, low-complexity algorithms such as Sum of Absolute Differences (SAD) can typically perform well, as stated in [16]. SAD-based algorithms are between the least-complex and more often-used. Census-based algorithms are common in real-time hardware-based systems and may work better in homogeneous zones of the

image. Its complexity increases when used in software-based systems because of its bit-based nature. Some systems mix both algorithms in order to obtain the best results from each one of them.

3 SUBJECTIVE 3D EVALUATION

Tests have been run over a set of 3D video sequences to understand and analyze different features which generate visual discomfort or quality reduction. A group of 16 observers were asked to rank the sequences taking into account their 3D quality. Results were compared to the objective data obtained through our developed tools to decide which features would be a possible cause of visual discomfort and how to modify them to obtain good 3D experiences. All tests were carried out on a 46" screen with passive glasses at the recommended distance.



Figure 3. Example of sequence "Itaca 3D"

Sequences used for the assessment included a 3-minute sequence called "Modernism" (Figure 4), created by Mediapro, in which different scenes appeared with different levels of motion and depth, sequences "Rain Fruits" and "Fountain" from EBU were also used [17], as well as synthetic test sequences "Palco HD" and "Itaca 3D" (Figure 3), which include parallax and object distances variation experiments created by us. All these sequences are high-definition resolution (1920x1080) with no compression, available in side-by-side formats.



Figure 4. Examples of frames in sequence "Modernism"

3.1 Cases of Study

As a conclusion of opinions obtained after subjective assessment with sequences with variations of parallax (positive and negative), motion and scene changes; different cases of study can be isolated. Different experiments were developed:

- Pairs of sequences with transitions from different types of parallax, negative and positive, to detect the impact over abrupt stereoscopic changes. A "Positive parallax" sequence (P.P) is considered when it has not remarkable negative parallax and pixels with positive parallax represent more than 25% of an image. On the other side, "Negative Parallax" sequences (N.P.) are sequences whose images posses more than 15% of pixels in negative parallax (assuming an environment of positive parallax). See Figure 5 for results.

- Negative parallax sequence with different levels of motion: low, medium and high motion. Test statistics are collected in Figure 6.

- Sequences with window violation (W.V) produced in different sides of the image, in lateral or top/bottom regions of the image. See subjective results in Figure 7.

- Long sequence with soft variation of parallax, at the end the sequence starts from the beginning producing an abrupt parallax change. Results from this experiment are in Figure 8.

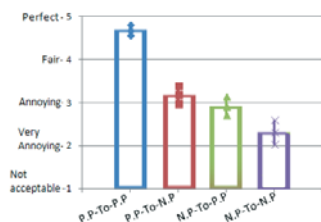


Figure 5. Transitions between types of parallax

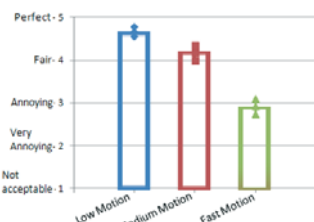


Figure 6. Results in impact related to motion

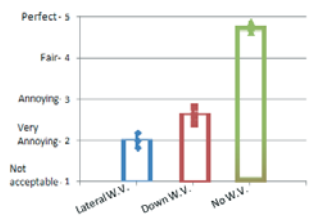


Figure 7. Sequences with Window Violation (W.V)

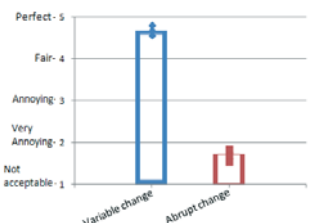


Figure 8. Progressive or abrupt parallax variations

4 WORK IMPLEMENTATION

In this section, the work developed is described in two subsections. First (Figure 9.a), the tools implemented in order to obtain quality through depth map histograms, calculating degradations related to each individual frame, are described in detail. In the second part (Figure 9.b), the work for static images is extended to sequences, analyzing video motion and the effect of depth when there are variations in parallax, derived from depth maps. Also some cases of study are analyzed when combining depth and motion.

4.1 Quality in Static Images Using Depth Map Histograms

To resolve the validity of a stereoscopic image it is required to determine whether it delivers visual discomfort or annoyance. The developed algorithm obtains parallax information through the computation of a depth map.

First of all, the depth map histogram is compared to the suited ZoC, in order to check if it doesn't fall out of its boundaries. Vergence-Accommodation conflict needs to be confined between ZoC limits to prevent visual discomfort.

In order to evaluate the disparity results it is necessary to understand the relation between disparity in pixels and virtual perception of depth. Figure 5 shows the trigonometric relations between the observer's location, an object's depth perception and its disparity measured in pixels. The relation

between d and x is not a fixed value, because the distance between the eyes doesn't change with the screen's size, thus it has to be assumed a size for the screen.

Assuming a display of diagonal size $D = 46''$ and with an aspect ratio (AR) of 16/9 and 1920x1080 resolution, screen width (W) would be 101 cm, therefore $R=101\text{cm}/1920\text{pixels}=0.053 \text{ cm/pixel}$. Assuming $E=6.5 \text{ cm}$ and $d= 2.4 \text{ m}$, that would leave us with a parallax range limited to $[-125, 107]$ for Shibata's ZoC, measured in pixels. Anything relevant that the algorithm finds out of those bounds will be considered as a cause of visual discomfort.

The other feature measured the window violation which is another suspect of causing visual annoyance. Window violation occurs when an object with negative parallax doesn't fit the screen and, therefore, is cut by the screen edges. Having negative parallax, it is supposed to be out of the screen, which means that screen edges shouldn't be able to hide its view. This generates an incoherent depth cue situation.

In order to measure this feature, the algorithm will examine the depth map's limits looking for negative parallaxes, which will be computed as a factor of visual annoyance.

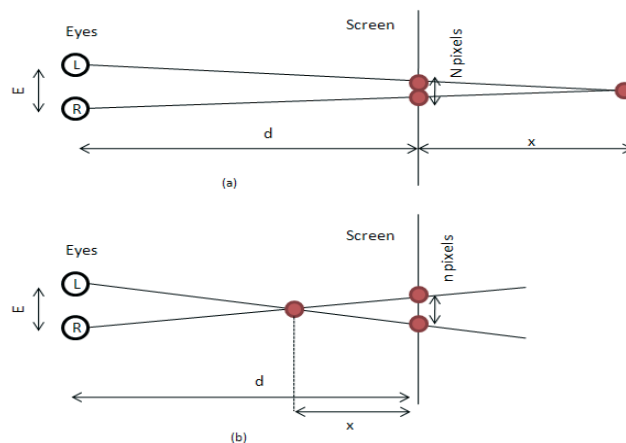


Figure 9. Positive (a) and negative (b) parallax estimation

4.1.1 Depth Map Calculation

To compute depth maps from stereoscopic images, the system performs a SAD-based algorithm. We need to obtain general depth characteristics of a scene and its evolution, though pixel depth accuracy in the whole image is not necessary. SAD-based algorithms work well enough to fulfill our goal and are less computationally demanding.

The weakest detections with SAD algorithms occur in homogeneous zones, where the capability to discern between possible pair candidates is low. In order to alleviate these probable errors, the system creates a difference between both views in order to calculate depths only over those pixels that will differ from one image to another, reducing homogeneous zones and, therefore, noise in resulting depth maps. Discarded pixels won't be taken into account. Figure 10 shows the original depth map (left) and the filtered depth map (right). In the original depth map there are several errors in the background zone, where the sky is homogeneous. In the left image this zone isn't calculated and therefore not taken into

account to classify the image's general depth, improving the absence of errors in the histogram.

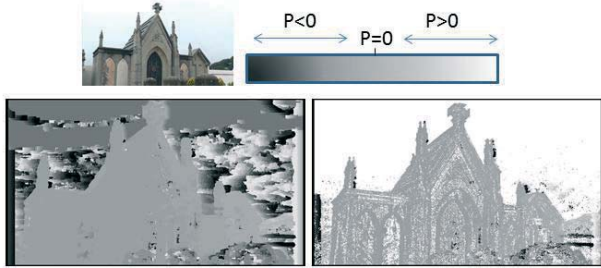


Figure 10. Advanced Depth Map

Figure 11 shows the histogram calculated for the previous depth map. All the elements in the scene have positive parallax. There is a very small amount of negative parallax pixels which represent noise (bad information) that result from the depth map algorithm calculation.

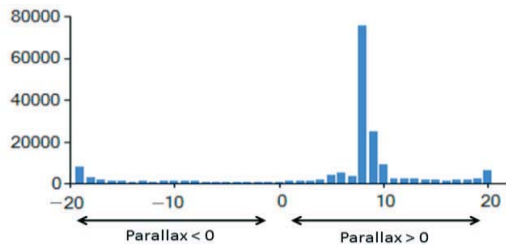


Figure 11. Example of parallax histogram

4.2 Depth and Motion QoE Decision Algorithm

In 3D stereoscopic video, motion is a basic element to take into account when assessing quality, as it is a primal reason for visual discomfort and is related to high depth levels which combine areas with negative and positive parallax.

The steps to follow for formulating a QoE decision are the ones which follow. Firstly, the complete sequence is processed to obtain the motion vectors in order to find the scene changes. In the exact moment where scene change occurs, the motion vectors are calculated between consequent frames to obtain the level of motion in that specific scene.

Depending on the level of motion, the scene is classified as slow, medium or fast motion. This is necessary to decide the necessity of calculating new depth maps for various frames if it is fast motion, or assuming same disparity for a collection of similar frames, saving time of computation.

After deciding the key-frames (one or more), depth maps are obtained for each of these frames by making use of the difference in images between left and right view, which is used as a mask to simplify the process. Depth map is calculated as explained in the previous section about static images. The comparison between parallax histograms, derived from each key-frame, allow us to make a statistic about the variations in objects depth and, consequently, quantify the probability of visual discomfort appearance.

4.2.1 Motion Vectors and Motion Estimation

The work derived from the static image process is related to motion. It is necessary to evaluate the motion level in a video

sequence, to conclude how much this motion affects the perception of the third dimension in stereoscopic video. For this purpose, the motion vectors calculation is obtained.

The whole sequence is processed in order to detect frames where motion is produced. In consecutive static frames or areas with low motion, the depth map is assumed to be the same for that sequence of frames. When medium or fast motion happens, more depth map information is necessary to compare results. Motion is calculated through average motion vectors in sequence (Figure 12). Motion is calculated as the average valid motion vectors (without discarding incoherent ones), always related to the variance of motion lengths.

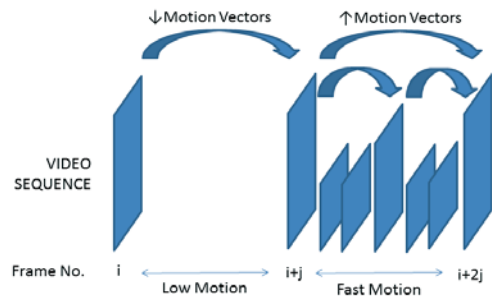


Figure 12. Video sequence motion analysis

For motion vectors calculation, only the left stereoscopic image is selected, and a grid is created to detect the block motion, in the case of the example a grid with 3 lines and 5 columns allow us to obtain 15 different motion vectors. The blocks between 9x9 and 15x15 pixels are searched in the next frame left image, homogeneous blocks are discarded to avoid false detections. The motion must be coherent in distance, so vectors with length values over two times the variance are also discarded. The final average motion vector length either reveals if the image is static or the corresponding motion level (low, medium or fast) related to the objects.

The last case is when a scene change occurs. Then depth maps from both previous and next are processed. This is a concrete case of motion vector abrupt variation, in which the variance of the vectors length is higher than when fast motion happens. As manifested from observers, the abrupt changes of negative parallax in a positive/negative parallax environment provoke a high visual discomfort in the observer's eye. Discomfort is usually produced in environments with significant-variance of negative parallax and motion, even with low and medium motion, and especially in fast motion sequences.

5 TEST RESULTS

With the results obtained from subjective assessment, studies were developed in static images and motion video sequences.

5.1 Results in Static Images

Tests were run over still images to classify stereo features without dealing with motion effects. Tests were focused on ZoC measurements, window violations and depth distribution.

To evaluate the effects of parallax out of ZoC we have rendered virtual images such as the one showed in Figure 3. When disparity was forced to be near Shibata's ZoC the

perception of the observers was negative, even when disparity fell below 70% of ZoC range. In order to secure a good comprehension of the scene, the threshold was fixed at 2/3 of the total of Shibata's ZoC. Further away the vergence-accommodation conflict was found to be nearly unsolvable or, at least, it took a lot of time to be solved. This effect of time will be dealt with in subsequent sections. Outside of the ZoC violations were easily detected by the algorithm analyzing the resultant parallax histogram. Other still images from the sequence (Figure 4) were used to quantify window violation cue conflict. During that sequence, the text is turning and, from time to time, some of that text crosses the screen's limits. As the text has a negative parallax, it should never touch the borders. From the tests results, it was determined that window violation cue conflict became difficult to overcome when at least 20% of the screen edge was filled with negative parallax pixels. Again, we were able to detect window violations measuring positive parallax pixels over the edges from computed depth maps results.

The last still image test was related to QoE rather than annoyance or discomfort. In this case, a set of images were ranked for their 3D effectiveness. Results were compared to their depth map histogram distribution. Figure 13 to Figure 16 show depth maps and histograms of the images submitted to test. Table 1 holds variance statistics for all the images tested. Note that histogram value for -80 pixels always shows a peak. This peak is considered as noise related to depth map calculation techniques and will not be taken into account when statistics are calculated.

The 3D perceiving the church and the cemetery images, observers usually prefer the second image because there's a wider range of depth. This is statistically measured as a bigger positive parallax variance. The "Library" and the "table" were found to be the preferred images due to its variety of depths, from positive to negative parallax.

Table 1: Histogram variances

Image	Positive Variance (pix)	Negative Variance (pix)
1	10	-
2	19,6	-
3	12,8	23,2
4	15,6	16,4

All these tests revealed an interest in depth variance and negative parallax because of their higher immersive capabilities. Our developed tools confirmed these features in each one of the images through statistical depth analysis, which led us to believe the system is well-suited to detect possible indications of 3D quality of experience through objective analysis of still images.

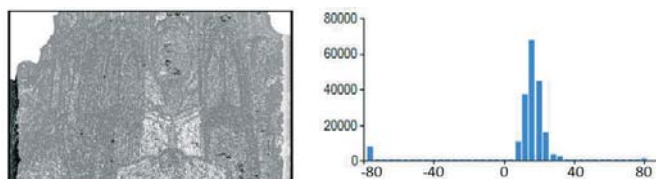


Figure 13. Church depth map and histogram

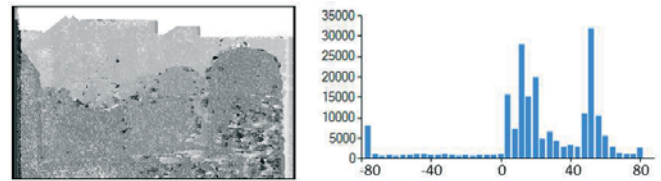


Figure 14. Cemetery depth map and histogram

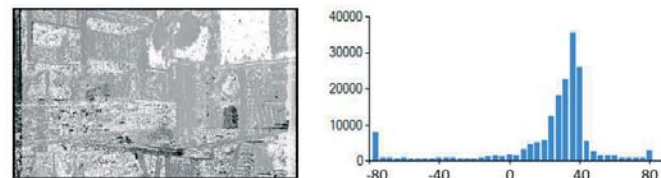


Figure 15. Library depth map and histogram

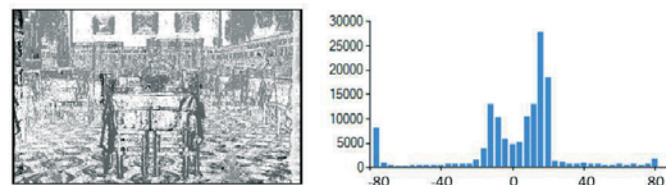


Figure 16. Table depth map and histogram

5.2 Depth and Motion QoE Decision Algorithm

First of all, the scene changes were analyzed with different variations of negative parallax in an environment of positive high-variance. As seen in Figure 17 the depth map and histogram are calculated, and their related statistics evaluated, to detect the scene changes.

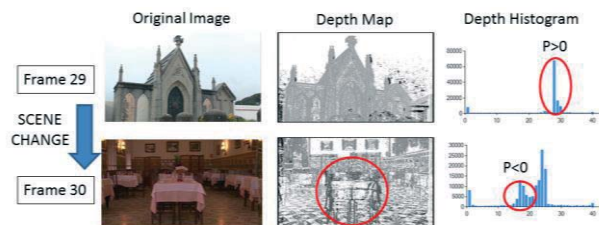


Figure 17. Example of scene change detection

Table 2: Positive and negative parallax from histogram analysis

Frame	Positive Parallax (% pixels)	Negative Parallax (% pixels)
29	44,26%	2,60%-
30	31,58%	18,15%

The variation of negative parallax from frame 29 to the next one is more than 15%, taking into account that, although the negative parallax in frame 29 nears zero, the positive parallax is very significant, with a score of more than 25%, which means that there is a high probability of detecting visual discomfort, as observers manifested in subjective tests, which need time to focus the objects in negative areas. Similar results have been obtained with scene changes with variations of negative parallax higher than 10%. Tests related to motion with high parallax variation offers similar results.



Figure 18. Frames with high negative parallax

Fast motion is detected in some sequences when the motion vectors reveal a movement higher than 2 pixels per frame, as happens in Figure 18, which shows the camera making a “travelling” fast movement.

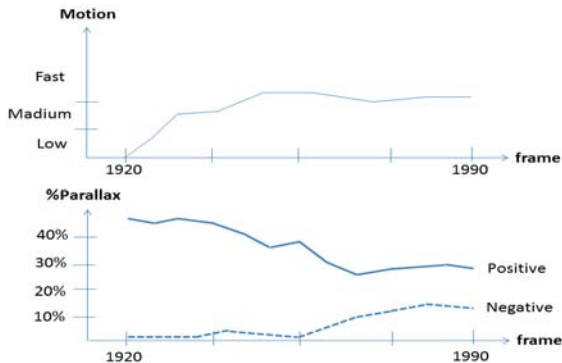


Figure 19. Evolution of motion and parallax percentage between frames 1920 and 1990

Figure 19 shows both negative and positive parallax percentage in parallel to motion description. It is remarkable that an abrupt increase in negative parallax is not enough for visual discomfort to be detected. It is necessary to create an environment with parallax variance and motion. The probability of discomfort is higher with faster motion.

6 CONCLUSIONS

Depth and motion are main factors in perceived quality of experience. Information provided by depth maps and estimated motion vectors is useful to avoid effects that can cause visual discomfort and fatigue in observers when contemplating 3d stereoscopic contents.

Subjective assessment allowed us to isolate the main features to be detected, in order to perform an algorithm which could translate user’s opinions into an automatic objective system.

The presence of objects with negative parallax on a static image, and especially when motion is detected in the video sequence which contains that image, requires quantifying the probability of the observer’s annoyance. This information can be obtained through depth maps, motion vectors and parallax histograms. In graphics comparing parallax and motion evolution the relation between both parameters in the final experience of users is remarkable. Previous Zone of Comfort (Zoc) studies have been found to be greatly affected by motion and time of viewing, diminishing its range significantly. Parallax getting near the ZoC edges (especially negative) has been proven to be undesirable when fast motion or high parallax variance appeared.

Tests that have been developed showed good results when applying the techniques to video sequences that contain effects which could be considered annoying for the human eye. Results obtained offer guidelines for stereoscopic video creation, extracting probabilities of visual discomfort and fatigue and reaching consensus between 3D sensationalism and annoyance to the observer’s eye. Nevertheless, the user has the final decision to accept or reject a determined content.

Acknowledgement

This paper is based on work performed in the framework of the project 3D-Contournet with research in techniques to assess quality in stereoscopic video. The work is also related to Immersive TV public funding project, headed by Indra Company and in collaboration with Mediapro, with the objective of developing an immersive environment with the use of CAVE and stereoscopic screens, and by the project TEC2012-38402-C04-01 HORFI, as well. We would like to acknowledge Jordi Alonso and people from Mediapro for lending 3D stereoscopic video contents with variations of parallax, available for the test development.

References

- [1] Z. Wang, A.C. Bovik, and L. Lu, “Why is image quality assessment so difficult?,” IEEE International Conference on Acoustics, Speech, & Signal Processing, vol. 4, pp. 3313-3316, May 2002.
- [2] S. Winkler. “Digital Video Quality: Vision Models and Metrics”. Ed. Wiley. March 2013. ISBN-13: 978-0470024041
- [3] H.R. Wu, K.R. Rao, “Digital Video Image Quality and Perceptual Coding (Signal Processing and Communications)”. CRC Press. November 2005. ISBN-13: 978-0824727772.
- [4] Z. Wang, A. Bovik. “Modern Image Quality Assessment (Synthesis Lectures on Image, Video, & Multimedia Processing)”. Morgan & Claypool Publishers. February, 2006. ISBN-13: 978-1598290226
- [5] “Methodology for the Subjective Assessment of the Quality of Television Pictures,” Recommendation ITU-R BT.500-11, ITU Telecom. Standardization Sector of ITU, 2002.
- [6] Jung, Y.J.; Sohn, H.; Lee, Seong-il; Speranza, F.; Ro, Y.M., "Visual importance- and discomfort region-selective low-pass filtering for reducing visual discomfort in stereoscopic displays," Circuits and Systems for Video Technology, IEEE Transactions on , vol.PP, no.99.
- [7] Wa James Tam; Speranza, F.; Yano, S.; Shimono, K.; Ono, H., "Stereoscopic 3D-TV: Visual Comfort," Broadcasting, IEEE Transactions on , vol.57, no.2, pp.335,346, June 2011
- [8] F. Speranza, W.J. Tam, R. Renaud, and N. Hur, “Effect of disparity and motion on visual comfort of stereoscopic images” in Proc. of SPIE, vol 6055, pp. 94-103, 2006.
- [9] Y.J. Jung, S. Lee., H. Sohn, H.W. Park and Y.M. Ro. “Visual comfort assessment metric based on salient object motion information in stereoscopic video”. Journal of Electron Imaging, vol. 21, Issue 1, Feb. 2012.
- [10] Sang-Hyun Cho; Hang-Bong Kang, "Subjective evaluation of visual discomfort caused from stereoscopic 3D video using perceptual importance map," TENCON 2012 - 2012 IEEE Region 10 Conference , vol., no., pp.1.6, 19-22 Nov. 2012
- [11] Kun-Lung Tseng; Wei-Jia Huang; An-Chun Luo; Wei-Hao Huang; Yin-Chun Yeh; Wen-Chao Chen, "Automatically optimizing stereo camera system based on 3D cinematography principles," 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D.
- [12] J.A. Rodrigo, D. Jiménez and J.M. Menéndez, “Real-Time 3-D HDTV Depth Cue Conflict Optimization” ICCE Berlin, September 2011, p. 8.
- [13] M.S. Banks, J.C.A. Read, R. S. Allison, and S. J. Watt “Stereoscopy and the Human Visual System”. Motion Imaging Journal (May/June 2012).
- [14] A. S. Percival, “The Relation of Convergence to Accommodation and its Practical Bearing”. Ophthalmol. Rev. 11: 313-328, 1892.
- [15] T. Shibata, J. Kim, D. M. Hoffman, and M.S. Banks, “The Zone of Comfort: Predicting Visual Discomfort with Stereo Displays”, J. Vision, 11: 1-28, 2011.
- [16] P. Leclercq, J. Morris. “Assessing Stereo Algorithm Accuracy”, IVCNZ ’02: Proceedings of Image and Vision Computing, 2002.
- [17] EBU Test Sequences. <http://tech.ebu.ch/testsequences>