A NO-REFERENCE STEREOSCOPIC QUALITY METRIC

Alessandro R. Silva [a], Max E. Vizcarra Melgar [b], Mylène C. Q. Farias [b]

[a]Instituto Federal de Goiás ,Anápolis, GO, Brazil
[b]University of Brasília , Department of Electrical Engineering Brasília, DF, Brazil

ABSTRACT

Although a lot of progress has been made in the development of 2D objective video quality metrics, the area of 3D video quality metrics is still in its infancy. Many of the proposed metrics are simply adaptations of 2D quality metrics that consider the depth channel as an extra color channel. In this paper, we propose a 3D no-reference objective quality metric that estimates 3D quality taking into account spatial distortions, excessive disparity, depth representation and temporal information of the video. The metric is resolution and frame-rate independent. To estimate the amount of spatial distortion in the video, the proposed metric uses a blockiness metric. The contribution of motion and excessive disparity to 3d quality is calculated using a non-linear relative disparity measure and a frame-rate proportional motion measure. The metric's performance is verified against the COSPAD1 database. The MOS predicted using the proposed metric obtained good correlation values with the subjective scores. The performance was on average better than the performance of two simple 2D full reference metrics: SIMM and PSNR.

Keywords: n o-reference, stereoscopic video, quality assessment, 3D video quality metrics.

1. INTRODUCTION

In the past years three-dimensional (3D) video has been perceived as the next development in video, given that it allows for a more natural and realistic user experience than conventional two-dimensional (2D) video. 3D video technology is not new and its first appearance can be traced back to 1903, when the Lumire brothers showed the first 3D short movies in the world fair in Paris.[1] Although these early films were relatively successful, it was not until 1950 that the movie industry started investing in 3D videos. But, despite its initial success, 3D videos failed to fulfill its expectations because of several technical problems.

With the transition from analogue to digital television services, there has been a new hope for a better 3D video technology. And, as in all communications services, the level of acceptability and popularity of the 3D applications in the next following years will be strongly correlated to the reliability of the service and the quality of the content provided.[2] As a consequence, a big effort of the scientific community has been dedicated to develop all aspects of production, distribution, and display of 3D.[3–5]

The most accurate way to determine the quality of 2D or 3D videos is to perform psychophysical experiments with human subjects. These experiments are expensive, time-consuming, and hard to incorporate into a design process or an automatic quality of service control. Therefore, algorithms (objective video quality metrics) that give a physical measure of the video quality are used to give an estimate of the quality being perceived by the user.[6–8]

The addition of a third dimension changes the overall user visual experience. In fact, for the 3D video content we can define the quality of experience(QoE) as a combination of several perceptual attributes, which include the overall spatial-temporal image quality (2D quality), the comfort level, the naturalness, and the realism of the 3D video.[9,10] All these attributes have to be taken into consideration when we measure the quality of 3D videos.

Although a lot of progress was made in the development of 2D objective video quality metrics, the area of 3D video quality metrics is still in its infancy.[6,8] Many of the proposed metrics are simply adaptations of 2D quality metrics that consider the depth channel as an extra color channel.[11–13] Some of the approaches estimate the errors in the right and left views and try to find a combination rule that better models 3D video quality, but fails to model how human viewers perceive the 3D content.[11] Also, most of the proposed metrics are full-reference (FR) metrics that use the original video to obtain an estimate of quality. Unfortunately, in real applications (e.g.

Three-Dimensional Image Processing, Measurement (3DIPM), and Applications 2015, edited by Robert Sitnik,
William Puech, Proc. of SPIE-IS&T Electronic Imaging, Vol. 9393, 93930B · © 2015 SPIE-IS&T
CCC code: 0277-786X/15/$18 · doi: 10.1117/12.2083306

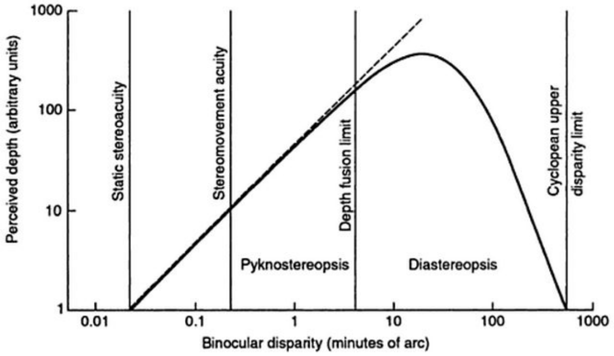


Figure 1. Graph of perceived depth versus binocular disparity, showing cyclopean limits of perceived depth and fusion [20]. The axis are in log units.

video transmission over IP) the original is hardly available and these types of metrics cannot be used. Its worth pointing out that FR methods are ideal to measure compression artifacts, but not 3D QoE. The reason for this is that the reference source could also suffer from stereoscopic artifacts and bad depth quality. Therefore, the development of blind metrics (no-reference – NR) is a necessity.

In this paper, we propose a new NR video quality metric for 3D videos. Our method takes in to account a non-linear relative disparity measure, a frame rate proportional motion measure, and a blockiness estimate. With this approach we evaluate the contribution of motion, excessive disparity and depth representation to 3d quality.

2. PROPOSED 3D VIDEO QUALITY METRIC

In general, objective quality metrics with best performances are metrics that try to incorporate relevant aspects of the HVS, such as color perception, contrast sensitivity, and pattern masking. In particular, to estimate the quality of 3D videos we must take into account the image quality (2D quality), the comfort level, the naturalness, and the realism of the 3D video (depth perception).[9,10] As mentioned earlier, a relatively big number of 3D video quality metrics have been proposed in the literature,[12–19] but only a subset of these metrics has considered factors other than 2D quality.

When stereoscopic cameras capture an object, it appears in different spatial positions in each of the views. The distance between these positions is called *binocular disparity.* 3D metrics that consider factors other than 2D quality often use disparity measure,[5,14,15] since disparity plays an important role in the quality of depth perception.

Frequently, 3D metrics make use of the absolute disparity[16,17] instead of the relative disparity, to estimate depth perception. This approach can lead to poor results if the disparities values of all objects in the scene are high and there are no low disparity objects to create the illusion of distance. Some metrics define depth perception as a non-decreasing function of the absolute disparity,[17,19] ignoring the effect of excessive disparity, which causes disconfort.

According with Hershenson,[20] depth perception does not vary linearly with the disparity. As depicted in Figure 1, bellow 0.5 min of arc approximately, binocular disparity produces effects that cannot be differentiated from a flat surface. Above this limit, depth perception increases almost linearly with disparity until the maximum depth perception is achieved at around 25 min. of arc. After this value, binocular rivalry occurs causing eyestrain. The NR 3D quality metric proposed in this paper takes into account two important parameters to predict the impact of disparity on depth perception. The first parameter measures the relation between viewing distance and screen size/resolution. The second parameter estimates the temporal information contribution. To calculate the screen parallax in pixels, the proposed metric uses a semi-global matching algorithm proposed by Hirschmuller.[21] This algorithm performs a pixel-wise matching based on a hierarchical calculation of mutual information. Its performance is equivalent to what is achieved by global methods, but better than what is achieved by local methods. Also, it is faster than popular global methods, like graph cuts and belief propagation.

To make the proposed metric independent of screen size and resolution, we estimate the disparity ($disp$) in minutes of arc using the following equation:[22]

$$disp = 2 \cdot tan^{-1}\left(\frac{pd}{2 \cdot D_{sc}} + \left|\frac{W_{sc} \cdot Paral}{2 \cdot D_{sc} \cdot Res_h}\right|\right) \tag{1}$$

where pd is the inter-pupillary distance, $Paral$ is the absolute screen parallax in pixels, D_{sc} is the distance to the screen, W_{sc} is the screen width, and Res_h is the horizontal resolution.

As a second step, a relative disparity value is measured using the followed procedure: for each fame we create a histogram of all disparities, equalize it, and compute the correlation ($disp_{corr}$) between these two histograms. The goal is to obtain a normal distribution histogram, using the whole range of values below the fusion limit. We apply a penalty on the value of the correlation relative to the percentage of the histogram above the fusion limit,as suggested by the model proposed by Hershenson.[20] This value is generally a better measure of depth estimation than the disparity mean. For example, consider a case where all pixel disparities have approximately the same positive value. If we use the disparity mean to estimate depth perception, we would obtain a constant positive value that could result in a false high value for depth. But if a histogram correlation is used we would obtain a small depth estimation value.

As mentioned earlier, movement plays an important role in human perception, providing important clues about the shape of 3D objects.[23] Motion tricks are frequently used to enhance the 3D experience making it unique, like for example when a moving object seems to come out the screen. Thats the reason why in this metric we consider motion as an important factor to the estimation of 3D quality, taking into account the movement in the z axis too.

To calculate the amount of motion of an object, we employed a motion estimation algorithm that use a polynomial expansion to approximate the neighborhood of each pixel. Then, it estimates displacement fields that generate the transform coefficients for each pixel. The z axis amount of movement is derived from the disparity map variance on the processed block. For each vector velocity map ($vvmap$), we use the following equation to obtain the amount of motion in the frame:

$$motion = \sum_{i=1}^{m}\sum_{j=1}^{n} \frac{\sqrt{vvmap(i,j) \cdot x^2 + vvmap(i,j) \cdot y^2 + vvmap(i,j) \cdot z^2}}{fr/a} \tag{2}$$

where fr is the video frame rate and a is a constant. In our tests, a has the same value as the window size, which is 15 pixels. It is worth pointing out that the frame rate is a parameter often neglected by video quality metrics. The final motion value for the video ($motion_v$) is given by the average $motion$ values for all video frames.

For 2D quality, it is possible to predict the overall annoyance of an impaired video estimating the strength of the artifacts present in the video by using one or more artifact metrics.[25] To estimate the spatial distortions of a 3D view, we used the blockiness metric proposed in our previous work.[25] In this algorithm, each frame is partitioned in blocks and, simultaneously, sampled in vertical and horizontal directions. Then, blockiness is estimated by comparing the cross-correlation of pixels inside and outside the borders of the block. The final blockiness value for the video ($block_v$) is given by the average blockiness values for all video frames. In our tests, we considered using other artifact metrics, but the results obtained were similar to what was obtained with using only this blockiness metric.

Finally, the overall 3D quality estimation is given by the following equation:

$$Q3D = a \cdot disp_{corr} - b \cdot \ln(block_v) + c \cdot motion_v^2 \tag{3}$$

To test our metric, we used the NAMA3DS1 COSPAD1 3D video database, which was created by Urvoy *et al.*[27] The 3D videos in this database have spatial resolution of 1920×1080. The display used in the experiment had a width (W_{sc}) of 101.83 cm and was distant (D_{sc}) 172cm from the observers. Since the study did not report the inter-pupillary distance (pd) of each participant, we used the know average of 6.3 cm.

We trained the metric using all the Hypothetical Reference Condition (HRC) of the database which corresponded to the Source Reference Sequences (SRC) 03 and 07. The choice of SRC07 and SRC03 was due to the fact that these scenes have high temporal information characteristics. Using these training sequences, the regression model returned the following constants for the $Q3D$ model: $a = 1.5136$, $b = 8.9842$, and $c = 0.031$. The metric value is thresholded at a maximum value of 5 and the log function prevents negative scores to occur. Also, since the axis is in log units, we get depth measures that better reflect the depth perception, as shown in Figure 1.

3. RESULTS AND DISCUSSION

We tested the proposed metric using the sequences from COSPAD1 database that were not used in the training phase, i.e. all HRCs corresponding to the SRCs 01-02, 04-06, and 08-10. The SRCs of the COSPAD1 database have different shooting distances, convergence angle, a variety of spatial and temporal information, and a range of disparities map configurations. Sample thumbnails of the SRCs used in this work are show in Figure 2.

The HRCs of the COSPAD1 database were created using H.264 and JPEG2000 and typical image processing algorithms, like downsampling and image sharpening. These are the HRC database specifications:

- H.264 compression – quantization parameters (QP) equal to 32, 38, and 44;
- MJPEG – 2, 8, 16, and 32 Mb/s;
- Reduction of resolution – downsampling by 4;
- Image sharpening – edge enhancement;
- Reduction of resolution + image sharpening;

From these HRCs, we only used the HRCs corresponding to H264 compression and image sharpenning.

The Pearson correlation coefficient (PCC) and the Spearman correlation coefficient (SCC) corresponding to using the proposed metric to estimate the quality of these HRCs are presented in Table 1 and Table 2, respectively. In these tables, we also present the results obtained using the 2D FR quality metrics SSIM and PSNR. For the SSIM and PSNR metrics, we averaged the values obtained for the left and right views.

Results show that, for each SRC, the proposed metric has a good correlation with the Mean Observer Scores (MOSs) of the database. The only exception was SRC02, for which the metric obtained a Pearson correlation of 0.851 and a Spearman correlation of 0.818. The Pearson and Spearman correlation coefficients obtained for the whole test set was 0.832 and 0.893, respectively. The average Pearson and Spearman correlation coefficients across all SRCs is 0.930 and 0.899, respectively. Figure 3 shows the scatter plot of the MOSs versus the corresponding Q3D values.

Table 1. Pearson correlation coefficients for metrics tested on COSPAD1 database[27].

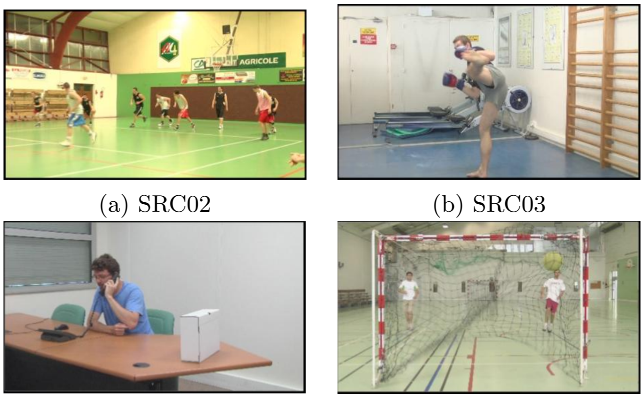
Source References	Proposed 3D NR Metric	SSIM	PSNR
SRC01	0.933	0.871	0.678
SRC02	0.851	0.860	0.649
SRC04	0.922	0.921	0.803
SRC05	0.965	0.821	0.827
SRC06	0.912	0.758	0.860
SRC08	0.934	0.928	0.719
SRC09	0.941	0.872	0.761
SRC10	0.989	0.889	0.647
AVG	0.930	0.865	0.743
All SRCs	0.832	0.712	0.630

4. CONCLUSIONS

In this paper, we proposed a new NR 3D video quality metric. The metric is a combination rule of a blockiness metric, a disparity metric, and a motion measure. The metric presents good results, outperforming the results of

Table 2. Spearman Correlation coefficients for metrics tested on COSPAD1 database.[27]

Source References	Proposed 3D NR Metric	SSIM	PSNR
SRC01	0.900	0.900	0.900
SRC02	0.818	0.975	0.821
SRC04	0.900	0.900	0.900
SRC05	0.900	0.900	0.900
SRC06	0.900	0.900	0.900
SRC08	0.900	0.900	0.900
SRC09	0.900	0.900	0.900
SRC10	0.975	0.900	0.900
AVG	0.899	0.889	0.647
All SRCs	0.893	0.909	0.890



(a) SRC02

(b) SRC03

(c) SRC07

(d) SRC08

Figure 2. Sample frame thumbnails of SRCs used in this work.

two 2D FR metrics. We believe that the good results obtained by the proposed NR metric are due to the motion estimate and the resolution independent disparity measure. In sources with low spatial quality, motion and depth measures contribute to an increase of the metric value, better matching the MOS values. Such behavior is evident for SRC08, where the motion and depth compensated the low scores given by the blockiness metric.

Unfortunately, currently there is a lack of 3D video content containing both spatial and stereoscopic impairments. The COSPAD1 database (with corresponding MOS) was the only database we found that could be used for our purpose. Nevertheless, limited the testing and training was possible given the number of SRCs available

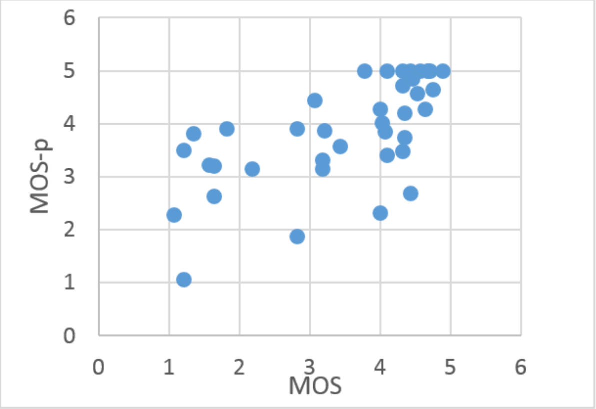


Figure 3. Scatterplot showing the relation of MOS and the predicted MOS using the proposed metric ($Q3D$).

and the absence of realistic 3D artifacts. We believe that the proposed metric can achieve a better performance when common 3D artifacts are represented in the database.

REFERENCES

[1] Lenny Lipton, *Foundations of the Stereoscopic Cinema: A Study in Depth*, Van Nostrand Reinhold Inc.,U.S., 1982.

[2] Nicolas Staelens, Koen Casier, Wendy Van den Broeck, Brecht Vermeulen, and Piet Demeester, "Determining customer's willingness to pay during in-lab and real-life video quality evaluation," in *Proceedings of the VPQM VI*, 2012, pp. 1–6.

[3] A.A. Alatan, Y. Yemez, U. Gudukbay, X. Zabulis, K. Muller, C.E. Erdem, C. Weigel, and A. Smolic, "Scene representation technologies for 3dtv: A survey," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 17, no. 11, pp. 1587–1605, Nov 2007.

[4] Aljoscha Smolic, Karsten Mueller, Nikolce Stefanoski, Joern Ostermann, Atanas Gotchev, Gozde B. Akar, Georgios Triantafyllidis, and Alper Koz, "Coding Algorithms for 3DTVA Survey," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 11, pp. 1606–1621, Nov. 2007.

[5] H Urey, K V Chellappan, E Erden, and P Surman, "State of the Art in Stereoscopic and Autostereoscopic Displays," *Proceedings of the IEEE*, vol. 99, no. 4, pp. 540–555, Apr. 2011.

[6] Anush Krishna Moorthy and Alan Conrad Bovik, "Visual quality assessment algorithms: what does the future hold?," *Multimedia Tools and Applications*, vol. 51, no. 2, pp. 675–696, Oct. 2010.

[7] Mylène C Q Farias, "Visual-quality estimation using objective metrics," *Jornal of the SID 19/11*, pp. 764–770, 2011.

[8] Quan Huynh-Thu, Patrick Le Callet, and Marcus Barkowsky, "Video quality assessment: From 2d to 3d - challenges and future trends.," in *ICIP*. 2010, pp. 4025–4028, IEEE.

[9] Pieter Seuntiens and Ingrid Vogels, "Visual Experience of 3DTV with pixelated Ambilight," *Teleoperators and Virtual Environments - Presence*, 2007.

[10] Marc T. M. Lambooij, Wijnand A. IJsselsteijn, and Ingrid Heynderickx, "Visual discomfort in stereoscopic displays: a review," Feb. 2007, vol. 6490, pp. 64900I–64900I–13.

[11] Atanas Boev, Maija Poikela, Atanas Gotchev, and Anil Aksay, "Modelling of the stereoscopic hvs," *Report on Mobile 3DTV* {*http://sp. cs. tut. fi/mobile3dtv/results/*}, 2009.

[12] C T E R Hewage, S T Worrall, S Dogan, and A M Kondoz, "Prediction of stereoscopic video quality using objective quality models of 2-D video," *Electronics Letters*, vol. 44, no. 16, pp. 44–45, 2008.

[13] Chaminda T. E. R. Hewage, Stewart T. Worrall, Safak Dogan, Stephane Villette, Ahmet M. Kondoz, and Student Member, "Quality Evaluation of Color Plus Depth Map-Based Stereoscopic Video," *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 2, pp. 304–318, Apr. 2009.

[14] Lutz Goldmann, Touradj Ebrahimi, Pierre Lebreton, and Alexander Raake, "Towards a Descriptive Depth Index for 3D Content: Measuring Perspective Depth Cues," in *Proceedings of the VPQM VI*, 2012.

[15] Pierre Lebreton, Alexander Raake, Marcus Barkowsky, and Patrick Le Callet, "Evaluating Depth Perception of 3D Stereoscopic Videos," *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 6, pp. 710–720, Oct. 2012.

[16] S.L L P Yasakethu, W.A.C. A C Fernando, B. Kamolrat, A. Kondoz, and Senior Member, "Analyzing perceptual attributes of 3d video," *IEEE Transactions on Consumer Electronics*, vol. 55, no. 2, pp. 864–872, May 2009.

[17] Jungdong Seo, Xingang Liu, Donghyun Kim, and Kwanghoon Sohn, "An Objective Video Quality Metric for Compressed Stereoscopic Video," *Circuits, Systems, and Signal Processing*, vol. 31, no. 3, pp. 1089–1107, Nov. 2011.

[18] Donghyun Kim, Seungchul Ryu, and Kwanghoon Sohn, "Depth perception and motion cue based 3D video quality assessment," in *IEEE BMSB*. June 2012, pp. 1–4, IEEE.

[19] Varuna De Silva, Hemantha Kodikara Arachchi, Erhan Ekmekcioglu, and Ahmet Kondoz, "Toward an impairment metric for stereoscopic video: a full-reference video quality metric to assess compressed stereoscopic video.," *IEEE transactions on image processing*, vol. 22, no. 9, pp. 3392–404, Sept. 2013.

[20] Maurice Hershenson, *Visual Space Perception: A Primer*, MIT Press, 1999.

[21] H. Hirschmuller, “Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information,” in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. 2005, vol. 2, pp. 807–814, IEEE.

[22] Daniele Siragusano, “Target Screensize for Stereoscopic Feature Film,” *SMPTE Conferences*, vol. 2010, no. 7, pp. 1–13, July 2010.

[23] Kalpana Seshadrinathan and Alan Conrad Bovik, “Motion tuned spatio-temporal quality assessment of natural videos.,” *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 19, no. 2, pp. 335–50, Feb. 2010.

[24] Gunnar Farnebäck, “Two-frame motion estimation based on polynomial expansion,” in *Proceedings of the 13th Scandinavian Conference on Image Analysis*, Berlin, Heidelberg, 2003, SCIA'03, pp. 363–370, Springer-Verlag.

[25] M.C.Q. Farias and S.K. Mitra, “No-reference video quality metric based on artifact measurements,” in *IEEE ICIP 2005*. 2005, vol. 3, pp. III–141, IEEE.

[26] T. Vlachos, “Detection of blocking artifacts in compressed video,” *Electronics Letters*, vol. 36, no. 13, pp. 1106, June 2000.

[27] Matthieu Urvoy, Marcus Barkowsky, Romain Cousseau, Yao Koudota, Vincent Ricordel, Patrick Le Callet, Jesus Gutierrez, and Narciso Garcia, “NAMA3DS1-COSPAD1: Subjective video quality assessment database on coding conditions introducing freely available high quality 3D stereoscopic sequences,” in *QoMEX*, July 2012, pp. 1–6.