

FULL-REFERENCE QUALITY ASSESSMENT OF STEREOPAIRS ACCOUNTING FOR RIVALRY

Ming-Jun Chen¹, Che-Chun Su¹, Do-Kyoung Kwon², Lawrence K. Cormack³ and Alan C. Bovik¹

Department of Electrical and Computer Engineering, The University of Texas at Austin¹,
Systems and Applications R&D Center, Texas Instrument²,
Department of Psychology, The University of Texas at Austin³

ABSTRACT

We develop a framework for assessing the quality of stereoscopic images that have been afflicted by possibly asymmetric distortions. An intermediate image is generated which when viewed stereoscopically is designed to have a perceived quality close to that of the cyclopean image. We hypothesize that performing stereoscopic QA on the intermediate image yields higher correlations with human subjective judgments. The experimental results confirm the hypothesis and show that the proposed framework significantly outperforms conventional 2D QA metrics when predicting the quality of stereoscopically viewed images that may have been asymmetrically distorted.

Index Terms— Binocular Rivalry, Image Quality, Stereoscopic Quality Assessment, Stereo Algorithm

1. INTRODUCTION

The theory of stereoscopic vision was invented in 1838 [1] and the production of 3D films can be dated back to 1903. Since then, numerous 3D films have been produced, but none was as successful as 2D films until Avatar in 2009. Avatar broke several box office records during its release and became the highest-grossing film of all time in the U.S. and Canada. The success of Avatar inspired 3D film production and technologies of 3D content capture and display. In 2011, mobile phones supporting 3D capturing and viewing have been released and broadcast 3D content over the internet is already available. With the release of 3D phones and 3D broadcasting services, it is a reasonable prediction that the amount of 3D content will increase exponentially in the near future. However, even though 3D movies are popular and prevalent today, our understanding of the quality of stereoscopic viewed 3D videos is still limited and needed to be advanced.

Research on 3D QA models can be divided into two classes based on whether computed disparity information is considered. The first group directly applies 2D QA model to the 3D QA problem. The methods in [2, 3] do not use disparity information and apply 2D QA algorithms to the left and right views independently, then combine (by various means)

the two scores into a predicted 3D quality score. The models in this class are based on the hypothesis that the quality of a binocularly viewed image may be deduced from the quality of the 2D images without accessing disparity or the third dimension. However, other studies [4] provide evidence that the quality of stereoscopically viewed images is generally different than a simple combination of the qualities of the 2D viewed images.

The second class of models takes depth information into account, typically by applying 2D quality assessment (QA) algorithms on both stereo images and also on the estimated disparity map [5, 6]. A 3D quality score is then generated using a combination of the various predicted 2D scores. The hypothesis underlying these QA models is that 3D viewing quality is correlated with depth quality. However, it is difficult to assess the quality of perceived depth or disparity, since ground truth disparity or depth is generally not available. Such models can only assess the depth quality using estimated disparity maps (computed from a pristine stereopair and/or from a distorted stereopair). Hence 3D QA performance may be substantially affected by the accuracy of the disparity estimation algorithm that is used. Moreover, benchmark tests on stereo algorithms [7] utilize high-quality stereo images, and the performance of stereo algorithms on distorted stereo images is rarely considered.

In this paper, we proposed a 3D QA framework that is motivated by the results of studies on masking and facilitation effects experienced when viewing stereoscopic images [8]. In particular, we make a model of the influence of binocular rivalry between the left-right views. Binocular rivalry is a perceptual effect that occurs when the two eyes view mismatched images at the same retinal location(s). Here, mismatch means that the stimuli received by the two eyes are sufficiently different from each other to cause match failures or to otherwise affect stereo perception. Failures of binocular matching trigger binocular rivalry, which is experienced in various ways, i.e., a sense of failed fusion or a bi-stable alternation between the left and right eye images. Figure 1 shows an example of binocular rivalry when mismatched stimuli are present. In Figure 1, in the interval (t_0, t_1) , the observer saw the stim-

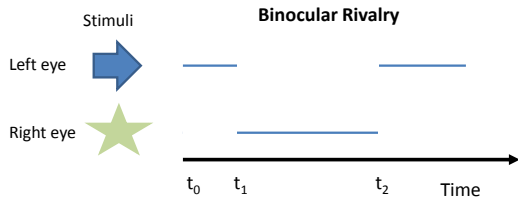


Fig. 1. Illustration of binocular rivalry: Two different stimuli are presented to the left eye (an arrow) and the right eye (a star). The blue line indicates that the stimulus is perceived by a human observer inside that time interval.

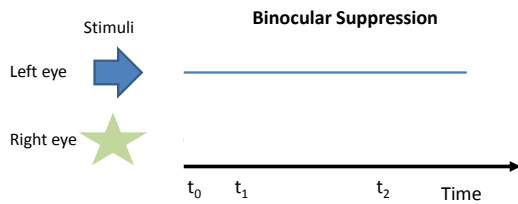


Fig. 2. Illustration of binocular suppression: Two different stimuli are presented to the left eye (an arrow) and the right eye (a star). An observer only sees the arrow when s/he experiences binocular suppression.

ulus from the left eye (the arrow). Then, the stimulus from the right eye (the star) dominated until time t_2 , after which the observer again saw an arrow. This fluctuation continues when an observer is experiencing binocular rivalry. The fluctuation period may vary from a fraction of a second to several seconds, and it may depend on the color, shape, and texture of the stimuli. Binocular suppression [9] is a special case of binocular rivalry. When binocular suppression is experienced, no rivalrous fluctuations occur between the two images when viewing the mismatched stereo stimulus. Instead, only one of the images is seen while the other is hidden from conscious awareness. Figure 2 shows an example of binocular suppression.

The rest of the paper was organized as follows. The overall 3D QA framework is described in Section 2 including the derivation of a practical 3D QA algorithm from the models. Section 3 describes the experiments conducted on the 3D QA database and analyzes the model performance. Finally, Section 4 is the conclusion.

2. A FRAMEWORK FOR QUALITY ASSESSMENT OF DISTORTED STEREO IMAGES

The logical goal of a 3D stereoscopic QA model is to estimate the quality of the true cyclopean image formed within

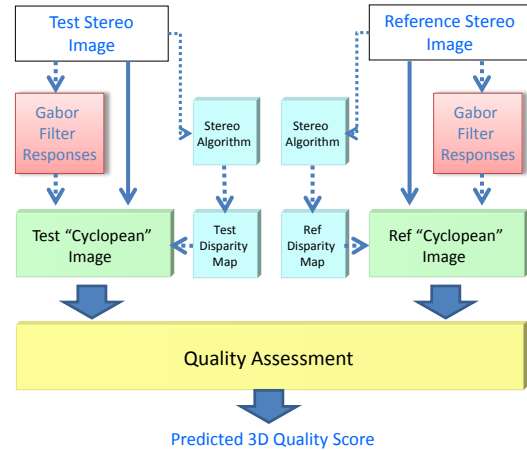


Fig. 3. The proposed framework for 3D quality assessment.

an observer’s mind when presented with a stereo image are presented with a stereopair. Of course, simulating the true *cyclopean* image associated with a given stereopair is a daunting task, since it would require accounting for the display geometry, the presumed fixation, vergence, and accommodation. This task is herculean, and is compounded by the fact that it is still unclear how a cyclopean image is formed! Towards a limited approximation of this goal, however, we seek to synthesize an internal image having a quality level that is close to the quality of the true *cyclopean* image. By way of notation, henceforth we still use the term “cyclopean” image to represent the synthesized image and *cyclopean* image to mean the one formed in the observer’s mind. By performing 3D quality assessment on the “cyclopean” image we hope to produce accurate estimates of 3D quality perceived on the *cyclopean* image.

The concept underneath the model framework is shown in Figure 3. Given a stereo image, an estimated disparity map is generated by a stereo algorithm, while Gabor filter responses are generated on the stereo images using a bandpass filter bank. A “cyclopean” image is synthesized from the stereo image pair, the estimated disparity map, and the Gabor filter responses. A “cyclopean” image is created from the reference stereopair and another “cyclopean” image is calculated from the test stereopair. Finally, full reference 2D QA models are applied to the two “cyclopean” images to predict 3D quality scores.

2.1. Stereo Algorithms

Research on stereo algorithm design has been a topic of intense inquiry for decades. However, there is no consensus on the type of stereo matching algorithm that should be used in 3D QA other than it be of low complexity. Further, there is scarce literature on the performance of stereo algorithms

operating under different distortion regimens. Therefore, we deploy a variety of these efficient stereo depth-finding algorithms differing considerably in their operational constants along with the framework we described above to assess perceived 3D quality.

In order gain insights into the influence of stereo algorithms on the performance of 3D QA models, three stereo algorithms were selected based on their complexity and performance. In general, better stereo algorithms (based on results on the Middlebury database [7]) have higher computational complexity, and we balanced this tradeoff in the choice of stereo matching models. The first algorithm has the lowest complexity. It uses a very simple sum-of-absolute difference (SAD) luminance matching functional without a smoothness constraint. The disparity value of a pixel in a stereopair is uniquely computed by minimizing the SAD between this pixel and its horizontal shifted pixels in the other view with ties broken by selecting the lower disparity solution. The second algorithm [10] has the highest complexity among the three models. This segmentation-based stereo algorithm delivers highly competitive results on the Middlebury database [7]. The third is a SSIM based stereo algorithm that uses SSIM scores to choose the best matches. The disparity map of a stereopair is generated by maximizing the SSIM scores between the stereopair along the horizontal direction, again resolving ties by a minimum disparity criterion.

2.2. Gabor Filter Bank

As discussed earlier, when the two images of a stereopair present different degrees or characteristics of distortion, the subjective quality of the stereoscopically viewed 3D image generally cannot be predicted from the average quality of the two individual images. Binocular rivalry is a reasonable explanation for this observation. Levelt [11] conducted a series of experiments that clearly demonstrated that binocular rivalry/suppression was strongly governed by low-level sensory factors. He used the term stimulus strength, and noted that stimuli that were higher in contrast, or had more contours, tend to dominate the rivalry. Inspired by this result, we use the energy of Gabor filter bank responses on the left and right images to model stimulus strength and to simulate rivalrous selection of “cyclopean” image quality.

The Gabor filter bank extracts features from the luminance and chrominance channels. These filters closely model frequency-orientation decompositions in primary visual cortex and capture energy in a highly localized manner in both space and frequency [12]. A complex 2-D Gabor filter is defined

$$G(x, y, \sigma_x, \sigma_y, \zeta_x, \zeta_y, \theta) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}[(\frac{R_1}{\sigma_x})^2 + (\frac{R_2}{\sigma_y})^2]} e^{i(x\zeta_x + y\zeta_y)} \quad (1)$$

where $R_1 = x \cos \theta + y \sin \theta$, and $R_2 = -\sin \theta + y \cos \theta$,

σ_x and σ_y are the standard deviations of an elliptical Gaussian envelope along x and y axes, and ζ_x and ζ_y are spatial frequencies, and θ orients the filter. The design of the Gabor filter bank was based on the work conducted by Su, *et al.* [13]. The local energy is estimated by summing Gabor filter magnitude responses over four orientations (horizontal, both diagonals, and vertical (90 degrees) at a spatial frequency of 3.67 cycles/degree.

2.3. Cyclopean Image

A linear model was proposed by Levelt [11] to explain the experience of *binocular rivalry* in perceived Cyclopean image when a stereo stimulus is presented. The model he proposed is:

$$C = w_l E_l + w_r E_r \quad (2)$$

where E_l and E_r are the stimuli to the left and the right eye respectively, w_l and w_r are weighting coefficients for the left and right eye that are used to describe the process of binocular rivalry, where $w_l + w_r = 1$, and C is the *cyclopean* image.

Given that a foveally presented monocular stimulus generally does not disappear spontaneously, he hypothesized that the duration of a period of dominance period of an eye does not depend on the strength of the stimulus presented to that eye, but rather on the stimulus strength presented to the other eye. Therefore he concludes that the experience of binocular rivalry is not correlated to the absolute stimulus strength of each view, but is instead related to the relative stimulus strengths of two views. He also proposed a model whereby the weighting coefficients are positively correlated with the stimulus strengths, which we embody in a biologically plausible model whereby the local energies of the responses of a bank of Gabor filters are used to weight the left and right image stimuli. Since binocular rivalry is a local phenomena, broadening Levelts model in this manner is a natural way to simulate a synthesized cyclopean image. In our model, as in Levelts; the stereo views used to synthesize to the cyclopean view are disparity-compensated. Thus the localized linear model that we use to synthesize a “cyclopean” image is :

$$CI(x, y) = W_L(x, y) \times I_L(x, y) + W_R((x + d), y) \times I_R((x + d), y) \quad (3)$$

where CI is the simulated “cyclopean” image, I_L and I_R are the left and right images respectively, and d is a disparity index that corresponds pixels from I_L to those in I_R . The weights W_L and W_R are computed from the normalized Gabor filter magnitude responses:

$$W_L(x, y) = \frac{GE_L(x, y)}{GE_L(x, y) + GE_R((x + d), y)} \quad (4)$$

$$W_R(x + d, y) = \frac{GE_R((x + d), y)}{GE_L(x, y) + GE_R((x + d), y)} \quad (5)$$

Table 1. SROCC scores obtained by averaging left and right QA scores (center column) and using the 3D “cyclopean” model (right column)

	2D Baseline	“Cyclopean Model”
PSNR	0.672	0.762
SSIM	0.796	0.856
MS-SSIM	0.78	0.901

where GE_L and GE_R are the summation of the convolution responses of the left and right images to filters of the form (1). Because of the normalization in (4, 5), increased Gabor energy of either (the left or right) stimulus suppresses the contribution of the other view when there is binocular rivalry. Finally, the task of 3D QA is performed by applying a full reference 2D QA algorithm on the reference “cyclopean” image and on the test “cyclopean” image.

3. EXPERIMENT AND DISCUSSION

A human study (33 subjects) was conducted to construct a subjective data set to be used in assessing algorithms of this type. The data set has 8 pristine stereo image pairs that have ground truth disparity maps (measured by high-precision laser range scanner) and 360 distorted stimuli. Five distortion types (white noise, blur, JPEG compression, JPEG2000 (JK2K) compression, and fast-fading (FF) distortions) were included. Stimuli were distorted both asymmetrically and symmetrically. We followed the recommendation for a single stimulus continuous quality scale (SSCQS) [14] to collect the 3D subjective image quality of each distorted stereoscopic image. Finally, difference mean opinion scores (DMOS) were obtained for each distorted stereo image pair.

We studied three widely-used full-reference 2D QA metrics (PSNR, SSIM [15], and MS-SSIM [16]) as candidate 2D QA methods to be used in the 3D QA framework. This is the final stage of predicting the quality of the *cyclopean* image. We used Spearman’s rank ordered correlation coefficient (SROCC) to measure the performance of 3D QA models. Higher SROCC values indicate good correlation with human perception.

3.1. Performance Using Ground Truth Disparity Map

We begin the performance analysis by using ground truth depth, which minimizes the effects of flaws in the stereo matching algorithms. The performance numbers are shown in Table 1. Also included are the performance numbers arrived at using the same 2D FR QA algorithms, simply applied to the left and right views and the QA scores averaged. The “cyclopean” QA algorithm does significantly better than the 2D baseline QA algorithms on the mixed data set containing

both symmetric and asymmetric distorted data.

It is clear from Table 1 that MS-SSIM delivers the best performance among the three 2D QA algorithms when embedded in the “cyclopean” model. To obtain deeper insights into how the performance of the “cyclopean” 3D QA model is improved by accounting for binocular rivalry, its performance on the separated symmetric and asymmetric distorted stereopairs is reported in Table 2. The performance numbers in Table 2 indicate that the “cyclopean” model did not boost performance on symmetric distorted stereoscopic images. However, performance was greatly enhanced on the asymmetric distorted stereopairs. Furthermore, Table 2 indicates that the task of predicting the quality of asymmetric distorted stereopairs is more difficult than that of predicting the quality of symmetric distorted data.

3.2. Influence of Stereo Matching Algorithms

The preceding discussion describing the stereoscopic “cyclopean” QA model assumed that highly accurate ground truth depth values are available. Next, we study stereoscopic QA performance when estimated depth is used as computed by stereo algorithms. The performance of the “cyclopean” model using ground truth disparity, estimated disparity, and no disparity information are reported in Table 3. Table 3 shows that there is no significant difference in the performance attained using the ground truth and estimated disparities, although the performance of the very simple SAD-based stereo algorithm is slightly lower than the other two stereo algorithms. All three significantly outperform the no-disparity case indicating that estimated disparities provide useful information when predicting the quality of the stereo 3D images in the database. These results suggest that we should not use bad-pixel rate to evaluate stereo algorithms in the context of 3D image quality assessment algorithm design.

Table 3. SROCC relative to human subjective scores attained by “cyclopean” model using different disparity maps.

Stereo Algorithm	SROCC
Ground Truth	0.901
SAD	0.876
SSIM	0.893
Klaus	0.890
No depth information	0.817

4. CONCLUSION AND FUTURE WORK

We presented a new framework for conducting automatic objective 3D QA that delivers highly competitive performance, with a clear advantage when left-right distortion asymmetries

Table 2. SROCC scores relative to human subjective scores. Obtained using averaged left-right QA scores (2D baseline) and the “Cyclopean” model on symmetric and asymmetric distorted stereopairs

	Symmetric		Asymmetric	
	2D Baseline	“Cyclopean” Model	2D Baseline	“Cyclopean” Model
PSNR	0.781	0.819	0.596	0.698
SSIM	0.826	0.85	0.742	0.827
MSSSIM	0.912	0.929	0.687	0.854

are present. The design of the framework is motivated by studies on the perception of distorted stereoscopic images, and theories of binocular rivalry. The “cyclopean” 3D QA model that we derived was tested on the LIVE Asymmetric 3D Image Quality Database, and found to significantly outperform conventional 2D QA models and well-known 3D QA models. The impact of the stereo algorithm used to conduct 3D QA was discussed. We also found that a low-complexity SSIM-based stereo algorithm performs quite well for estimating disparity in the “cyclopean” algorithm in the sense that a high level of 3D QA performance is maintained.

5. REFERENCES

- [1] C. Wheatstone, “Contributions to the physiology of vision. part the first. on some remarkable, and hitherto unobserved, phenomena of binocular vision,” *Phil. Trans. R. Soc. Lond.*, vol. 128, pp. 371–394, 1838.
- [2] S. L. P. Yasakethu, C. T. E. R. Hewage, W. A. C. Fernando, and A. M. Kondoz, “Quality analysis for 3D video using 2D video quality models,” *IEEE Trans. Consumer Electronics*, vol. 54, no. 4, pp. 1969–1976, 2008.
- [3] P. Gorley and N. Holliman, “Stereoscopic image quality metrics and compression,” *Proc. SPIE*, vol. 6803, p. 05, 2008.
- [4] S. L. B. Meegan, D. V. and W. J. Tam, “Unequal weighting of monocular inputs in binocular combination: implications for the compression of stereoscopic imagery,” *J. Exp. P.: Appl.*, vol. 7, pp. 143–53, November 2001.
- [5] A. Benoit, P. Le Callet, P. Campisi, and R. Cousseau, “Quality assessment of stereoscopic images,” *EURASIP Journal on Image and Video Processing*, vol. 2008, pp. 1–13, 2009.
- [6] J. You, L. Xing, A. Perkis, and X. Wang, “Perceptual quality assessment for stereoscopic images based on 2D image quality metrics and disparity analysis,” *Proc. Int’l Workshop Video Process Quality Metrics Consumer Elect.*, 2010.
- [7] D. Scharstein and R. Szeliski, “A taxonomy and evaluation of dense two-frame stereo correspondence algorithms,” *International Journal of Computer Vision*, vol. 47, pp. 7–42, 2002.
- [8] M.-J. Chen, A. C. Bovik, and L. K. Cormack, “Study on distortion conspicuity in stereoscopically viewed 3D images,” *IEEE 10th IVMSP Workshop*, pp. 24–29, 2011.
- [9] R. Blake, D. H. Westendorf, and R. Overton, “What is suppressed during binocular rivalry?” *Perception*, vol. 9, no. 2, pp. 223–31, 1980.
- [10] A. Klaus, M. Sormann, and K. Karner, “Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure,” *Proc. Intl Conf. Pattern Recognition*, vol. 3, pp. 15–18, 2006.
- [11] W. J. M. Levelt, *On Binocular Rivalry*. The Hague; Paris: Mouton, 1968.
- [12] D. J. Field, “Relations between the statistics of natural images and the response properties of cortical-cells,” *J. Opt. Soc. Am.*, vol. 4, no. 12, pp. 2379–2394, 1987.
- [13] C.-C. Su, A. C. Bovik, and L. K. Cormack, “Natural scene statistics of color and range,” *IEEE Int. Conf. Image Processing*, Sep. 2011.
- [14] I. T. U. R. Assembly and U. International Telecommunication, *Methodology for the Subjective Assessment of the Quality of Television Pictures*. Geneva, Switzerland: International Telecommunication Union, 2003.
- [15] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE Trans. Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [16] Z. Wang, E. P. Simoncelli, and A. C. Bovik, “Multi-scale structural similarity for image quality assessment,” *Asilomar Conf. on Signals, Syst. and Computers*, vol. 2, pp. 1398–1402, 2003.