Article

Learning to Measure Stereoscopic S3D Image Perceptual Quality on the Basis of Binocular Rivalry Response

Siyuan Huang ¹ and Wujie Zhou ^{1,2,*}¹ School of Information & Electronic Engineering, Zhejiang University of Science and Technology, Hangzhou 310023, China² Institute of Information & Communication Engineering, Zhejiang University, Hangzhou 310027, China

* Correspondence: 109029@zust.edu.cn

Received: 18 July 2019; Accepted: 16 September 2019; Published: 18 September 2019

Abstract: Blind perceptual quality measurement of stereoscopic 3D (S3D) images has become an important and challenging issue in the research field of S3D imaging. In this paper, a blind S3D image quality measurement (IQM) method that does not depend on examples of distorted S3D images and corresponding subjective scores is proposed. As the main contribution of this work, we replace human subjective scores with a quality codebook of binocular rivalry responses (BRRs); this allows blind S3D-IQM methods to be learned without evaluation performance loss. Our results, using the publicly accessible LIVE S3D dataset, confirm that our method is highly robust and efficient.

Keywords: stereoscopic 3D image; binocular rivalry response; image quality measurement; quality codebook

1. Introduction

In recent years, stereoscopic 3D (S3D) technologies have attracted considerable research interest. S3D measurement can be classified into S3D content measurement (e.g., visual discomfort, quality of experience, or unnatural experience) and S3D image quality measurement (IQM) (e.g., the measurement of compression distortions or transmission impairments for S3D images). However, research on S3D content measurement is still limited. In recent decades, blind S3D-IQM has been attracting attention in the research field of S3D imaging [1–3]. State-of-the-art blind S3D-IQM methods learn to measure S3D quality via regression from the subjective scores of training samples. For instance, in [4], Akhter et al. introduced a blind S3D-IQM method that extracts both 2D and S3D quality-aware features from an S3D image and a depth/disparity map. Next, a regression model is adapted to obtain a final quality value from the above-mentioned features. Similar studies were also conducted in [5]; Chen et al. proposed a blind S3D-IQM method based on a “cyclopean” view and disparity map. Zhou et al. introduced blind S3D-IQM based on the self-similarity of binocular features [6]. Other relevant existing studies can be found in [7–14]. These methods require human-scored S3D images for training; however, obtaining subjective scores through subjective evaluation is often expensive, cumbersome, and time-consuming, which limits their usability in practice. Thus, a significant problem is triggered: can we learn to measure S3D image quality blindly without relying on human-scored S3D images?

Quality measurement of 2D images has been extensively studied, and several excellent methods have been proposed [15–18]. For example, in [16], Xue et al. presented a quality-aware clustering metric, which can learn centroids to act as a quality-aware codebook for measuring distorted image quality. Inspired by [16], we propose a blind S3D-IQM method that learns a quality codebook of binocular rivalry responses (BRRs). The main contributions of this work are three-fold.

First, in the offline training phase, we learn a quality codebook from the BRRs of the original and distorted S3D images; this codebook will be used to replace human subjective scores to learn blind S3D-IQM methods.

Second, a BRR-weighted local binary pattern (LBP) histogram scheme is used to extract quality-predictive features; this scheme is effective for describing degradation patterns.

Third, in the online testing phase, a distance-based weighting scheme is used for perceptual quality pooling. Experimental results obtained using the publicly accessible LIVE S3D dataset indicate that the proposed method performs competitively and provides quality prediction performance and generalization ability.

2. Proposed Method

The block diagram for our blind S3D-IQM method is depicted in Figure 1. During the offline training phase, similarity scores and a weighted LBP histogram are generated from training S3D images; subsequently, a quality codebook is constructed. During the online testing phase, by implementing feature extraction for test S3D images, blind perceptual quality pooling can be easily achieved by a distance-based weighting scheme. Our method is described in detail as follows.

Figure 1. Block diagram for our blind S3D-IQM.

(1) *Training dataset preparation:* We select 12 reference S3D images collected from Mobile S3DTV and MPEG (please refer to Figure 2) [19]. From each reference S3D image, we then generate distorted S3D images using four types of distortions, including JPEG2000 (JP2K), Gaussian blur (Gblur), white noise (WN), and JPEG. For each distortion, five quality levels are generated by controlling different parameters.

Figure 2. Left views of the training S3D images in this paper.

(2) *Quality measurement of the training S3D images*: Neuroscience has made rapid progress in comprehending the human visual mechanism and how S3D visual signals are transmitted to the human brain. BRR is a well-researched visual mechanism of competition in the visual cortex [20]. BRR is a human visual response whereby the binocular visual mechanism alternates between contradictory monocular images viewed by both eyes. Some findings in [20] showed that BRR can be strongly regulated by low-level sensory features, and a classic linear mathematical model is provided. Motivated by the above, 2D-Gabor is used for both views to simulate the rivalries. Consequently, BRR, which can be denoted by $R_{bin}(x, y, \sigma)$, is defined as

$$R_{bin}(x, y, \sigma) = W_l(x, y) \cdot M_l(x, y, \sigma) + W_r(x + d, y) \cdot M_r(x + d, y, \sigma) \quad (1)$$

where d denotes a disparity value estimated using an S3D matching algorithm [21] (Reference [22] reports that similar performance is attained when using either the ground truth or the estimated disparity). Subscripts “ l ” and “ r ” are the left and right views, respectively. $M_l(x, y, \sigma)$ and $M_r(x, y, \sigma)$ are the monocular responses of both views, which can be obtained using a difference of Gaussian (DoG) filter

$$h_\sigma(r, \sigma) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right) - \frac{1}{2\pi k^2\sigma^2} \exp\left(-\frac{r^2}{2k^2\sigma^2}\right) \quad (2)$$

where σ denotes the scales parameter and k is used to control the scales of the DoG. r denotes the l_2 norm. In this study, scales σ of the DoG are set to 0, 1, 1.6, 2.56, and 4.096 (the same values as the filter used in [2]), with $k = 1.6$. W_l and W_r denote the weights for imitating the BRR. They are represented as $W_l(x, y) = G_l(x, y)/(G_l(x, y) + G_r(x + d, y))$ and $W_r(x + d, y) = G_r(x + d, y)/(G_l(x, y) + G_r(x + d, y))$, respectively, where $G_l(x, y)$ and $G_r(x, y)$ denote the respective 2D-Gabor responses of both views. The 2D Gabor filter can be defined as:

$$G(x, y, \sigma_x, \sigma_y, \zeta_x, \zeta_y, \theta) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-(1/2)[((x \cos \theta + y \sin \theta)/\sigma_x)^2 + ((y \cos \theta - x \sin \theta)/\sigma_y)^2]} e^{i(x\zeta_x + y\zeta_y)} \quad (3)$$

where σ_x and σ_y denote the standard deviations along the x and y axes, ζ_x and ζ_y , and θ denotes spatial frequencies and orients the filter. The design of the 2D-Gabor filter was based on the research conducted by Su et al. [23].

The similarity score between the BRRs of the original S3D image and distorted S3D images is defined as:

$$S_n = \frac{1}{\Omega} \sum_{x, y} \left(\prod_{\sigma} \left(\frac{2 \langle R_{o,n}(x, y, \sigma), R_{d,n}(x, y, \sigma) \rangle + T}{\|R_{o,n}(x, y, \sigma)\|_2 \cdot \|R_{d,n}(x, y, \sigma)\|_2 + T} \right) \right) \quad (4)$$

where T is a small positive constant. In this study, we set $T = 0.085$ in our experiments. Ω indicates the whole spatial domain, $\|\cdot\|_2$ denotes the l_2 norm, and $\langle \cdot \rangle$ represents the inner product. Subscript “ n ” is the index of the n -th training distorted S3D images. Subscripts “ o ” and “ d ” are the original and distorted S3D images, respectively.

(3) *Quality-predictive features extraction*: Based on Marr’s theory [24], local structural features based on the visual cortex (area V1) are related to image quality. LBPs are basic feature extractors that are commonly used to represent an image’s structural and textural information [25]. First, we compute the rotation invariant uniform LBP map, $LBP_{P,R,BRR,n}^{riu2}$, for BRR by comparing BRR intensity against its eight neighbors (in consideration of computational complexity, R and P were set to 1 and 8, respectively). To effectively capture textural and structural features, the BRR-weighted LBP histogram is applied to extract quality-predictive features of BRR, which are defined as

$$h_n(k, \sigma) = \frac{1}{\Omega} \sum_{x, y} R_{d,n}(x, y, \sigma) \times f(LBP_{P,R,BRR,n}^{riu2}(x, y, \sigma), k) \quad (5)$$

where $f(i, j) = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{otherwise} \end{cases}, i \& j \in [0, P + 1]$.

(4) *BRR quality codebook*: Thus far, we have gained a set of human-scored S3D images for training. All the quality scores (denoted by $\{S_n\}$) and all the feature vectors (denoted by $\{h_n(k, \sigma)\}$) of the distorted S3D images for training constitute a BRR quality codebook to reflect different quality levels of visual signals and associated quality-predictive features.

(5) *Blind perceptual quality pooling*: After obtaining a BRR quality codebook, we use the codes to predict the quality score of a test S3D image, relying on the hypothesis that S3D images with similar quality-aware features should share similar visual qualities. First, let $h_m^d(k, \sigma)$ represent the quality-aware features of the n -th test S3D image. We define the distance $H_{m,n}$ between $h_m^d(k, \sigma)$ and $h_n(k, \sigma)$ as the product of the chi-square distance between each pair of quality-predictive features.

$$H_{m,n}[h_m^d(k, \sigma), h_n(k, \sigma)] = \sum_{n=1}^N \frac{(h_m^d(k, \sigma) - h_n(k, \sigma))^2}{h_m^d(k, \sigma) + h_n(k, \sigma)} \quad (6)$$

where N is the total number of training S3D images. Let $\{\hat{H}_{m,1}, \hat{H}_{m,2}, \dots, \hat{H}_{m,9}, \hat{H}_{m,10}\}$ represent the chi-square distances of 10 of the nearest quality-predictive features, which are calculated using the n -th test set and the selected training set. A distance-based weighting method is then utilized to measure the perceptual visual quality of the test set. The normalized weight for the 10 similarity scores of the selected training set, which is denoted by $w_{m,n'}$, can be represented as

$$w_{m,n'} = (\hat{H}_{m,n'})^{-1} / \sum_{n'=1}^{10} (\hat{H}_{m,n'})^{-1}. \quad (7)$$

By distributing larger weights to the similarity values of the selected neighboring training S3D images, the final quality value of the n -th test S3D image is calculated as

$$Q_m = \sum_{n'=1}^{10} w_{m,n'} \cdot S_{n'}. \quad (8)$$

3. Results

In the experiment, the effectiveness of our method was verified using the LIVE S3D [5] dataset, which includes 365 symmetrically distorted S3D images (in the future, we plan to understand how asymmetrical distortion affects perceptual quality (for the case of blur artifacts, the perceptual quality of asymmetric blurred S3D images is mainly dominated by the view that contains more information. For the case of JPEG artifacts, the binocular perceptual quality has a tendency towards the lower quality view) and extension of the proposed method to facilitate asymmetrical distortion S3D-IQM). In order to facilitate the experiment (refer to [16]), we utilized only four distortion types: JP2K, WN, JPEG, and Gblur. The images in the dataset were assigned difference mean opinion scores (DMOS), which represent subjective judgments.

In order to compare the proposed method against existing S3D-IQM methods, two widely used performance criteria were calculated by applying the DMOS and the predicted quality scores. The first was a Pearson's linear correlation coefficient (PLCC), which was computed to evaluate the prediction accuracy. The second was a Spearman's rank order correlation coefficient (SROCC), which measured the prediction monotonicity. For an ideal IQM method, values close to 1 were needed for SROCC and PLCC.

To demonstrate the robustness and effectiveness of the proposed method, we compared it with several existing representative IQM methods, including three 2D-IQM methods—SSIM (full reference, FR) [26], IL_NIQE (blind, not depending on DMOSs) [15], and Xue's method (blind, not depending

on DMOSs) [16]—and seven S3D-IQM methods—Lin’s method (FR) [2], Akhter’s method (blind, depending on DMOSs) [4], Chen’s method (blind, depending on DMOSs) [5], Yue’s method (blind, depending on DMOSs) [9], Zhou’s method (blind, depending on DMOSs) [12], Shao’s method (blind, not depending on DMOSs) [13], and Zhou’s method (blind, not depending on DMOSs) [14]. All IQM methods are based on the grayscale of S3D images. The SROCC and PLCC results are listed in Table 1; the criterion that provides the best evaluation performance is emphasized in boldface (all the blind methods that do not depend on DMOSs). As shown in Table 1, compared with other methods, it is evident that our method correlates well with the DMOSs, owing to its sufficient application of human visual perception. It thereby delivers a performance that is superior to most blind and FR methods, including SSIM and IL_NIQE, and the methods proposed by Xue, Lin, Shao, and Zhou. Thus, our method is an accurate, effective, and consistent S3D-IQM method, even though it does not use human-scored S3D images.

Table 1. Overall performance evaluation of nine methods.

Criteria	FR		Blind								
			Depending on Difference Mean Opinion Scores (DMOS)					Not Depending on DMOS			
	SSIM	Lin [2]	Akhter [4]	Chen [5]	Yue [9]	Zhou [12]	IL_NIQE	Xue [16]	Shao [13]	Zhou [14]	Proposed
PLCC	0.899	0.873	0.640	0.901	0.937	0.973	0.896	0.871	0.877	0.887	0.925
SROCC	0.882	0.830	0.395	0.899	0.914	0.965	0.876	0.873	0.866	0.892	0.887

Another experiment was conducted to further investigate the evaluation ability of the IQM methods for specific distortion types. In this experiment, we tested the performance of the state-of-the-art methods on individual distortion types. The SROCC and PLCC results are listed in Table 2. For each individual distortion type, the top two results are emphasized in boldface (all blind methods that do not depend on DMOS). As shown in Table 2, our method placed among the top methods most often. Although some methods may provide good performance for some distortion types, the proposed method performs better than, or on par with, the best method for each individual distortion type.

Table 2. Performance of nine methods for each individual distortion type.

Distortion	Criteria	FR		Blind								
				Depending on DMOS					Not Depending on DMOS			
		SSIM	Lin [2]	Akhter [4]	Chen [5]	Yue [9]	Zhou [12]	IL_NIQE	Xue [16]	Shao [13]	Zhou [14]	Proposed
JP2K	PLCC	0.865	0.838	0.905	0.907	0.934	0.988	0.854	0.919	0.901	0.848	0.939
	SROCC	0.857	0.839	0.866	0.863	0.832	0.961	0.861	0.886	0.870	0.837	0.887
JPEG	PLCC	0.485	0.214	0.729	0.695	0.744	0.916	0.533	0.722	0.456	0.626	0.673
	SROCC	0.435	0.199	0.675	0.617	0.595	0.912	0.544	0.682	0.429	0.638	0.612
WN	PLCC	0.937	0.928	0.904	0.917	0.962	0.988	0.927	0.858	0.950	0.925	0.943
	SROCC	0.940	0.928	0.914	0.919	0.932	0.965	0.920	0.938	0.914	0.931	0.909
Gblur	PLCC	0.920	0.948	0.617	0.917	0.971	0.974	0.904	0.923	0.919	0.899	0.976
	SROCC	0.882	0.935	0.555	0.878	0.857	0.855	0.873	0.871	0.932	0.833	0.903

Next, we discuss the influence of disparity estimation performance on the proposed method. The performance of the proposed method using ground truth, SAD [27], and Klaus [21] are listed in Table 3. Table 3 shows that a similar performance is attained when using either the ground truth or the estimated disparity.

Table 3. The performance of the proposed method using different stereo methods.

Stereo Method	Pearson's Linear Correlation Coefficient (PLCC)	Spearman's Rank Order Correlation Coefficient (SROCC)
Ground Truth	0.927	0.891
SAD	0.921	0.883
Klaus	0.925	0.887

4. Conclusions

In this paper, we proposed a new blind S3D-IQM method that does not depend on DMOS during learning. The main contribution of this study is that we constructed a quality-aware codebook to obtain blind IQM using a perceptual pooling strategy. Experimental results show that the proposed method can produce assessments that are statistically much more consistent with DMOS.

Further work will mainly focus on: (1) other binocular visual perception (e.g., binocular fusion, binocular suppression) can be considered for improving the performance of the proposed method, (2) the effect of disparity information needs to be further addressed and, besides disparity, many other factors, such as monocular depth cues, visual comfort, visual direction, etc., should be considered in modeling S3D-IQM without relying on human-scored S3D images.

Author Contributions: S.H. conceived and designed the experiments, analyzed the data, and wrote the paper. W.Z. supervised the work, helped with designing the conceptual framework, and edited the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (Grant No. 61,502,429) and the Zhejiang Provincial Natural Science Foundation of China (Grant No. LY18F020012).

Conflicts of Interest: The authors declare no conflict of interest.

References

- Jiang, Q.; Shao, F.; Lin, W.; Jiang, G. Learning a referenceless stereopair quality engine with deep nonnegativity constrained sparse autoencoder. *Pattern Recognit.* **2018**, *76*, 242–255. [\[CrossRef\]](#)
- Lin, Y.H.; Wu, J.L. Quality assessment of stereoscopic 3D image compression by binocular integration behaviors. *IEEE Trans. Image Process.* **2014**, *23*, 1527–1542. [\[PubMed\]](#)
- Li, Q.; Lin, W.; Fang, Y. BSD: Blind image quality assessment based on structural degradation. *Neurocomputing* **2017**, *236*, 93–103. [\[CrossRef\]](#)
- Akhter, R.; Baltes, J.; Sazzed, Z.; Horita, Y. No-reference stereoscopic image quality assessment. In *Stereoscopic Displays and Applications XXI*; International Society for Optics and Photonics: Bellingham, WA, USA, 2010; Volume 7524, p. 75240T.
- Chen, M.-J.; Cormack, L.K.; Bovik, A.C. No-Reference quality assessment of natural stereopairs. *IEEE Trans. Image Process.* **2013**, *22*, 3379–3391. [\[CrossRef\]](#) [\[PubMed\]](#)
- Zhou, W.; Zhang, S.; Pan, T.; Yu, L.; Qiu, W.; Zhou, Y.; Luo, T. Blind 3D image quality assessment based on self-similarity of binocular features. *Neurocomputing* **2017**, *224*, 128–134. [\[CrossRef\]](#)
- Jiang, G.; He, M.; Yu, M.; Shao, F.; Peng, Z. Perceptual stereoscopic image quality assessment method with tensor decomposition and manifold learning. *IET Image Process.* **2018**, *12*, 810–818. [\[CrossRef\]](#)
- Chen, Z.; Zhou, W.; Li, W. Blind stereoscopic video quality assessment: From depth perception to overall experience. *IEEE Trans. Image Process.* **2018**, *27*, 721–734. [\[CrossRef\]](#)
- Yue, G.; Hou, C.; Jiang, Q.; Yang, Y. Blind stereoscopic 3D image quality assessment via analysis of naturalness, structure, and binocular asymmetry. *Signal Process.* **2018**, *150*, 204–214. [\[CrossRef\]](#)
- Yang, J.; Jiang, B.; Wang, Y.; Lu, W.; Meng, Q. Sparse representation based stereoscopic image quality assessment accounting for perceptual cognitive process. *Inf. Sci.* **2018**, *430*, 1–16. [\[CrossRef\]](#)
- Liu, L.; Yang, B.; Huang, H. No-reference stereopair quality assessment based on singular value decomposition. *Neurocomputing* **2018**, *275*, 1823–1835. [\[CrossRef\]](#)
- Zhou, W.; Chen, Z.; Li, W. Dual-Stream Interactive Networks for No-Reference Stereoscopic Image Quality Assessment. *IEEE Trans. Image Process.* **2019**, *28*, 3946–3958. [\[CrossRef\]](#) [\[PubMed\]](#)

13. Shao, F.; Lin, W.; Wang, S.; Jiang, G.; Yu, M.; Dai, Q. Learning receptive fields and quality lookups for blind quality assessment of stereoscopic images. *IEEE Trans. Cybern.* **2016**, *46*, 730–743. [[CrossRef](#)] [[PubMed](#)]
14. Zhou, W.; Yu, L.; Qiu, W.; Luo, T.; Wang, Z.; Wu, M.W. Utilizing binocular vision to facilitate completely blind 3D image quality measurement. *Signal Process.* **2016**, *129*, 130–136. [[CrossRef](#)]
15. Zhang, L.; Zhang, L.; Bovik, A.C. A feature-enriched completely blind image quality evaluator. *IEEE Trans. Image Process.* **2015**, *24*, 2579–2591. [[CrossRef](#)] [[PubMed](#)]
16. Xue, W.; Zhang, L.; Mou, X. Learning without human scores for blind image quality assessment. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 995–1002.
17. Zhou, W.; Yu, L.; Qiu, W.; Zhou, Y.; Wu, M. Local gradient patterns (LGP): An effective local statistical features extraction scheme for no-reference image quality assessment. *Inf. Sci.* **2017**, *397*, 1–14. [[CrossRef](#)]
18. Zhou, W.; Yu, L.; Zhou, Y.; Qiu, W.; Wu, M.-W.; Luo, T. Local and global feature learning for blind quality evaluation of screen content and natural scene images. *IEEE Trans. Image Process.* **2018**, *27*, 2086–2095. [[CrossRef](#)]
19. Mobile 3DTV Content Delivery Optimization over DVB-H System. Available online: <http://sp.cs.tut.fi/mobile3dtv/stereo-video/> (accessed on 16 March 2011).
20. Levelt, W.J. The alternation process in binocular rivalry. *Br. J. Psychol.* **1966**, *57*, 225–238. [[CrossRef](#)]
21. Klaus, A.; Sormann, M.; Karner, K. Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure. In Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06), Hong Kong, China, 20–24 August 2006.
22. Chen, M.J.; Su, C.C.; Kwon, D.K.; Cormack, L.K.; Bovik, A.C. Full-reference quality assessment of stereopairs accounting for rivalry. *Signal Process. Image Commun.* **2013**, *28*, 1143–1155. [[CrossRef](#)]
23. Su, C.-C.; Bovik, A.C.; Cormack, L.K. Natural scene statistics of color and range. In Proceedings of the 18th IEEE International Conference on Image Processing, Brussels, Belgium, 11–14 September 2011; pp. 257–260.
24. Marr, D.; Hildreth, E. Theory of edge detection. *Proc. R. Soc. Lond. B Biol. Sci.* **1980**, *207*, 187–217.
25. Ojala, T.; Pietikinen, M.; Menp, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 971–987. [[CrossRef](#)]
26. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [[CrossRef](#)] [[PubMed](#)]
27. Scharstein, D.; Szeliski, R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. J. Comput. Vis.* **2002**, *47*, 7–42. [[CrossRef](#)]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).