

# No-reference stereoscopic image-quality metric accounting for left and right similarity map and spatial structure degradation

HUANLING WANG,<sup>1</sup> JIACHEN YANG,<sup>1,2,\*</sup> WEN LU,<sup>3</sup> BAIHUA LI,<sup>2</sup> KUN TIAN,<sup>4</sup> AND QINGGANG MENG<sup>2</sup>

<sup>1</sup>School of Electronic Information Engineering, Tianjin University, 92 Weijin Road, Tianjin 300072, China

<sup>2</sup>Department of Computer Science, School of Science at Loughborough University, Leicestershire, UK

<sup>3</sup>School of Electronic Engineering, Xidian University, Xi'an 710071, China

<sup>4</sup>National Key Laboratory of Science and Technology on Aerospace Intelligence Control, Beijing 100854, China

\*Corresponding author: yangjiachen@tju.edu.cn

---

Blind quality assessment of 3D images is used to confront more real challenges than 2D images. In this Letter, we develop a no-reference stereoscopic image quality assessment (SIQA) model based on the proposed left and right (LR)-similarity map and structural degradation. In the proposed method, local binary pattern features are extracted from the cyclopean image that are effective for describing the distortion of 3D images. More importantly, we first propose the LR-similarity map that can indicate the stereopair quality and demonstrate that the use of LR-similarity information results in a consistent improvement in the performance. The massive experimental results on the LIVE 3D and IRCCyN IQA databases demonstrate that the designed model is strongly correlated to subjective quality evaluations and competitive to the state-of-the-art SIQA algorithms.

---

In recent years, driven by the entertainment industry and the application of science, a variety of 3D applications have emerged [1–3]. Stereoscopic image quality assessment (SIQA) is the crux of the design and optimization of 3D systems. Compared to 2D image quality assessment (2D-IQA), 3D-IQA not only deals with at least twice the amount of 2D data, but also needs to consider more factors: visual comfort [4], 3D effects [5], and so on.

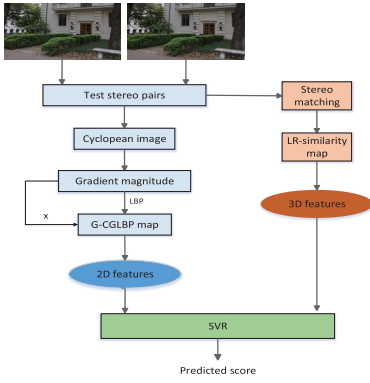
The existing mature 2D algorithms [6] can be directly applied to the left and right views to predict the overall scores of the stereopairs. However, extensive studies on human visual system (HVS) models have proved that 3D perception is not a simple combination of two views. For 3D-IQA, binocular vision properties [7,8], and 3D factors are very important and

challenging matters that need additional attention. Chen *et al.* [9] proposed the cyclopean view based on the theories of masking and facilitation effects experienced when viewing stereoscopic images. The cyclopean vision can be considered as a simulation of binocular synthesis vision perceived in the brain when humans observe left and right views. Accordingly, the evaluation of 3D quality on cyclopean is of great significance. At present, the cyclopean image model has brought about a multitude of promising 3D-IQA algorithms. In addition, Shao *et al.* [10] developed a blind stereoscopic image quality assessment (BSIQA) model based on a binocular feature combination, which led to a more natural and convenient representation of binocular visual perception.

Currently, machine learning has been applied to some IQA metrics to predict quality. Distortion Identification-based Image Verity and Integrity Evaluation (DIIVINE) [11] and Blind Image Quality Index (BIQI) [12] draw natural scene statistical features from 2D images, and then employ a regression model to predict the final objective quality. However, these metrics cannot effectively predict the quality of stereopairs. The predominant challenge for machine-learning-based SIQA is feature representation, in particular, taking the binocular vision and depth information into account.

In this Letter, we propose a novel framework for BSIQA. Figure 1 illustrates the procedure of our model. As the first step, we extract 2D features from the cyclopean. In view of the fact that cyclopean image considers the binocular visual characteristics and achieves excellent performance in the evaluation of 3D image quality, the method in Ref. [9] is employed to synthesize the cyclopean in our study. According to the study that image structures carry the essential visual information, the gradient magnitude map is calculated from the cyclopean.

The local binary pattern (LBP) operator is an excellent measure of the spatial structure of the local image texture [13]. Thus, we compute the LBP histogram to represent the texture invariance. By applying the LBP operator on the cyclopean gradient magnitude map (CGLBP), the CGLBP at one location is derived as



**Fig. 1.** Proposed blind quality assessment framework for stereopairs.

$$\text{CGLBP}_{P,R} = \sum_{i=0}^{P-1} s(g_p - g_c) 2^i, \quad (1)$$

where  $P$  is the total number of neighbors ( $P = 8$ ),  $R$  is the radius of the neighborhood ( $R = 1$ ),  $g_p$  and  $g_c$  are gray values at the center location and its neighbor, and  $s(\cdot)$  is the threshold function as follows:

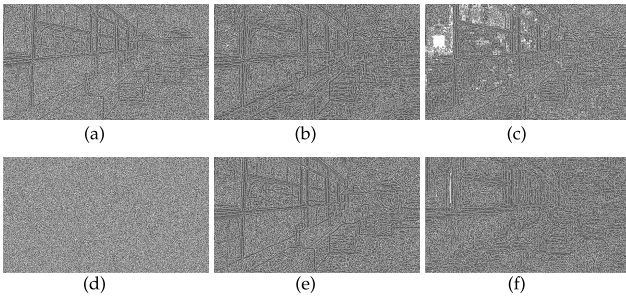
$$s(t) = \begin{cases} 1, & t \geq 0 \\ 0, & t < 0 \end{cases}. \quad (2)$$

In the interest of achieving admirable discrimination, a rotation invariant texture description is defined as

$$\text{CGLBP}_{P,R}^{\text{riu}2} = \begin{cases} \sum_{p=0}^{P-1} s(g_p - g_c) & \text{if } U(\text{CGLBP}_{P,R}) \leq 2 \\ P + 1 & \text{otherwise,} \end{cases} \quad (3)$$

$$U(\text{CGLBP}_{P,R}) = |s(g_{P-1} - g_c) - s(g_0 - g_c)| + \sum_{p=1}^{P-1} |s(g_p - g_c) - s(g_{p-1} - g_c)|, \quad (4)$$

where  $U$  is the uniform measure, and superscript superscript<sup>riu2</sup> reflects the use of rotation invariant “uniform” patterns that have  $U$  value of at most 2. It can be seen that  $\text{CGLBP}_{P,R}^{\text{riu}2}$  has  $P + 2$  distinct output patterns in all. As shown in Fig. 2, the CGLBP patterns change with different distortions according to their own characteristics, which make it an effective measure to capture the degradation by various distortions.



**Fig. 2.** CGLBP maps under different distortions: (a) pristine; (b) JP2K, difference mean opinion scores (DMOS) = 30.0645; (c) JPEG, DMOS = 17.3438; (d) WN, DMOS = 60.9635; (e) blur, DMOS = 13.8411; (f) FF, DMOS = 54.0365.

$\text{CGLBP}_{P,R}^{\text{riu}2}$  is an excellent measure of the spatial patterns, but, in accordance with the definition, it is a grayscale invariant measure and abandons contrast. Cyclopean gradient magnitude is an impactful measure to encode contrast information and HVS is highly sensitive to it. The change of local contrast has a significant influence on image quality [14], so we want to incorporate the contrast of the local image in our model as well. We merge the cyclopean gradient magnitude in  $\text{CGLBP}_{P,R}^{\text{riu}2}$  as the weight coefficient ( $G - \text{CGLBP}_{P,R}^{\text{riu}2}$ ). At present, the final  $G - \text{CGLBP}_{P,R}^{\text{riu}2}$  is expected to be a very powerful quality evaluation measure and can be represented as

$$G - \text{CGLBP}_{P,R}^{\text{riu}2}(k) = \sum_{i=1}^N w_i g(\text{CGLBP}_{P,R}^{\text{riu}2}(i, k)), \quad (5)$$

$$g(x, y) = \begin{cases} 1, & x = y \\ 0, & \text{otherwise,} \end{cases} \quad (6)$$

where  $k \in (0, P + 1)$ ,  $N$  denote the total number of pixels in an image,  $g(\cdot)$  is the discriminant function, and  $w_i$  is the weight coefficient assigned to each pixel. This expression is based on the property that the HVS is expert in capturing texture features. Now, the final 2D features can be expressed as the frequency of each pattern. Considering the HVS’s perceive ability for image details depends on the sampling density of the image signal, we implement these features in four scales, yielding 40 dimensional features.

As the second step, 3D feature extraction is performed on the left and right (LR)-similarity map. Taking into consideration that the feature extracted from the disparity map has less improvement on the overall performance of SIQA, for the first time, to the best of our knowledge, we propose and construct the LR-similarity map to represent the relationship between the left and right images. The LR-similarity map produced by the stereo matching algorithm is a powerful feature for the task of SIQA. In our implementation, in order to improve the computational efficiency, the LR-similarity map is based on structural similarity index measurement (SSIM)-based stereo matching algorithm, which is also used in the cyclopean production. The energy cost function is defined as

$$f = \frac{(2\mu_l\mu_r + C_1)(2\sigma_l + C_2)}{(\mu_l^2 + \mu_r^2 + C_1)(\sigma_l^2 + \sigma_r^2 + C_2)}, \quad (7)$$

where  $l$  is the left-view image, and  $r$  is the right-view image. According to the principle of winner take all, the point with maximum energy cost function in the matching image (right view) is considered to be the matching point of the original image (left view), and the corresponding value is considered to be the similarity degree of the two pixels. Ultimately, the LR-similarity map is obtained by normalizing all the values of similarity. We assume that the degree of similarity, to a certain extent, can indicate the degree and type of distortion. The greater the distortion of the image is, the smaller the degree of similarity in the corresponding point becomes. Figure 3 demonstrates the LR-similarity maps under the same distortion levels as in Fig. 2.

By comparing (b)–(f) with (a) in Fig. 3, we can find that the distortion type and degree affect the LR-similarity maps. And in a LR-similarity map, the parts of the worst similarity are clustered around the edge. Image spatial entropy indicates the amount of information and spatial distribution of information

within an image. Therefore, the spatial entropy is extracted from the LR-similarity map. We divide each map into  $8 \times 8$  blocks, and then compute spatial entropy within each block. The spatial entropy is defined as

$$E = -\sum_x p(x) \log_2 p(x), \quad (8)$$

where  $x$  are the pixel values within a block, and  $p(x)$  are the relative frequency density. To explore the behavior of the local spatial entropy values against the LR-similarity maps that are produced from the same images with different degrees and types of distortions, we conducted a series of validation experiments.

We can find in Fig. 4 that the LR-similarity map extracted from the undistorted image has a spatial entropy histogram with a mean of about 3.5, and which is mildly “left-skewed.” However, the introduction of distortion will change its mean and skew. For example, “JP2K,” “JPEG,” “WN,” and “FF” tend to increase the mean sharply and induce the histogram to be typically “left-skewed.” While “blur” tends to reduce the mean and the degree of “left-skewed.”

Since we believe that there exists a strong relationship between spectral entropy values and the LR-similarity maps under different distortion types and degree, the block discrete cosine transform (DCT) coefficient matrix is also computed on each  $8 \times 8$  block. Then, the DCT coefficients are normalized to produce a spectral probability map  $P(i, j)$ . We define the spectral entropy as

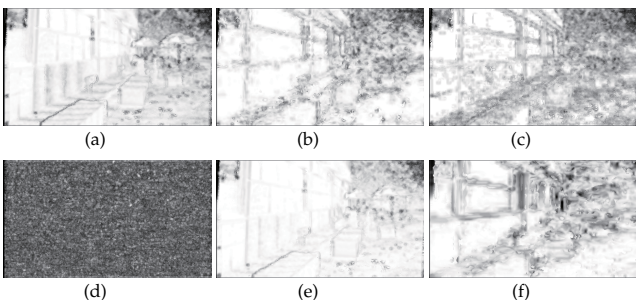
$$E = -\sum_i \sum_j P(i, j) \log_2 P(i, j), \quad (9)$$

where  $1 \leq i \leq 8$ ,  $1 \leq j \leq 8$ , and  $i, j \neq 1$ .

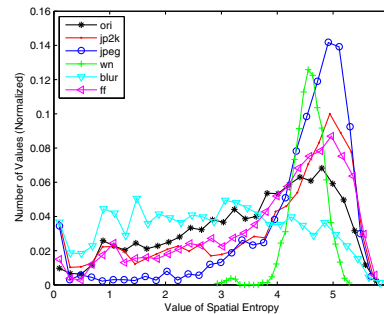
Figure 5 shows the value of spectral entropy varies with the distortion degree and type. Clearly, we may find that the introduction of distortion will also alter the spectral entropy of the original LR-similarity map. For instance, “JPEG,” “JP2K,” “blur,” and “FF” slightly change the trend of “right-skewed” and the mean of spectral entropy values. However, “WN” will introduce much high-frequency information, so spectral entropy values of LR-similarity maps under “WN” increase sharply.

Overall, entropy features are strongly indicative of the LR-similarity maps under different distortion. Therefore, we use the mean and skew as descriptive quality features.

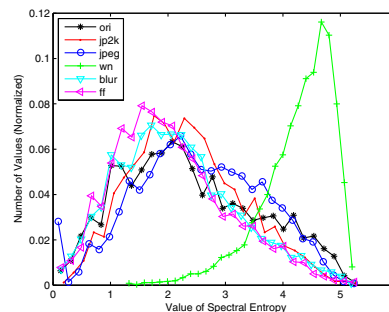
The final stage is quality prediction. The support vector regression (SVR) is employed as the mapping function from the 2D and 3D feature vectors to the quality score.



**Fig. 3.** LR-similarity maps under different distortions, where brightness indicates the degree of similarity.



**Fig. 4.** Histograms of spatial entropy values for LR-similarity maps in Fig. 3.



**Fig. 5.** Histograms of spectral entropy values for LR-similarity maps in Fig. 3.

We appraise the performance of our model on two publicly 3D IQA databases: LIVE 3D IQA database (LIVE Phase-I and Phase-II) and the IRCCyN database. All indices reported are the results of fitting using a standard nonlinear five-parameter logistic function. In the experiment, a database is randomly separated into nonoverlapping training subset and testing subset. In each train–test process, 80% of the database is elected for training and the rest is for testing. The aforementioned procedure is executed 1000 times, and the median performance are reported.

We compare the performance of our algorithm on LIVE 3D database with seven IQA approaches: DIIVINE [11], BIQI [12], Chen’s scheme [9], Benoit’s scheme [15], Lin’s scheme [16], Shao’s scheme [17], and Chen’s scheme [18]. Since the first two metrics are 2D-IQA metrics, feature vectors are extracted separately from the left and right views, and then averaged to obtain the final feature vector for support vector machine (SVM) to train a regression function. Chen’s [9], Lin’s [16], and Shao’s scheme [17] are full-reference (FR) metrics and do not need training, we report their results on the entire database. While Chen’s scheme [18] and our proposed model are no-reference (NR) metrics and need training, their results are reported in accordance with the principles of 80% for training and 20% for testing.

Table 1 illustrates the performance results on LIVE Phase-I and Phase-II data sets. Table 2 gives the detailed values of each individual distortion type on LIVE Phase-I. One can see that, even though the performance of the proposed model on some specific types of distortion is lower than other schemes, the overall performance of the proposed model is higher than the other metrics.



**Table 1. Comparison with State of the Art over Live 3D IQA Database (Best Performance Is Marked in Bold)<sup>a</sup>**

Metric	LIVE 3D Phase-I			LIVE 3D Phase-II		
	PLCC	SRCC	RMSE	PLCC	SRCC	RMSE
DIIVINE [11]	0.8419	0.7885	8.7660	0.6654	0.5821	8.3273
BIQI [12]	0.9210	0.8892	6.3672	0.6955	0.6036	8.0391
Ref. [9]	0.916	0.9153	8.7697	0.9067	0.9017	4.7603
Ref. [15]	0.8829	0.8862	10.2681	0.6938	0.7140	12.5746
Ref. [16]	0.8645	0.8559	10.9898	0.6584	0.6375	8.4956
Ref. [17]	0.9367	0.9365	5.7426	0.8601	0.8387	5.7581
Ref. [18]	0.8950	0.8910	7.2470	0.8950	0.8800	5.1020
Only 2D features	0.9397	0.9360	5.5438	0.9177	0.8977	4.4429
2D + 3D features	<b>0.9505</b>	<b>0.9457</b>	<b>5.0389</b>	<b>0.9292</b>	<b>0.9170</b>	<b>4.1406</b>

<sup>a</sup>For an efficient metric, its Pearson linear correlation coefficient (PLCC) and Spearman rank order correlation coefficient (SRCC) values should be high while its root mean squared error (RMSE) values should be low.

**Table 2. Detailed PLCC Performance Comparison on LIVE Phase-I**

	JPEG	JP2K	WN	Blur	FF	All
DIIVINE [11]	0.6315	0.6870	0.9379	0.9034	0.7739	0.8419
BIQI [12]	0.6901	0.8777	0.9546	0.9411	0.6884	0.9210
Ref. [9]	0.6356	0.8381	0.9351	0.9417	0.7579	0.916
Ref. [15]	0.5579	0.8897	0.9360	0.9256	0.7514	0.8645
Ref. [16]	0.2866	0.8381	0.928	0.9475	0.7086	0.8645
Ref. [17]	0.7636	0.9518	0.9271	<b>0.9600</b>	<b>0.8765</b>	0.9367
Ref. [18]	<b>0.9170</b>	0.907	0.695	0.917	0.735	0.895
2D + 3D features	0.8529	<b>0.9647</b>	<b>0.9662</b>	0.9574	0.7810	<b>0.9505</b>

To demonstrate the effectiveness of our proposed 3D features, we combine the 3D features with the 2D-extended features, and the SVM is employed to train regression function. Table 3 shows that the 3D features can greatly improve the performance of the 2D-extended algorithm. To further verify the performance of our model, we also report the performance of our metrics on the IRCCyN database. Table 4 displays the performance of our model on the IRCCyN database.

Accordingly, we can come to the conclusion that our proposed metric is powerful for predicting the 3D image quality, and the 3D information of the LR-similarity can enormously improve the performance of SIQA algorithms.

**Table 3. Performance of 2D-Extended Metrics on LIVE 3D IQA Database**

Metric	LIVE 3D Phase-I			LIVE 3D Phase-II		
	PLCC	SRCC	RMSE	PLCC	SRCC	RMSE
DIIVINE [11]	0.8419	0.7885	8.7660	0.6654	0.5821	8.3273
DIIVINE + 3D features	0.9026	0.8895	6.9757	0.7520	0.7271	7.3728
BIQI [12]	0.9210	0.8892	6.3672	0.6955	0.6036	8.0391
BIQI + 3D features	0.9447	0.9330	5.3338	0.8046	0.8086	6.6544

**Table 4. Performance Comparison on the IRCCyN Database**

Metric	PLCC	SRCC	RMSE
Ref. [9]	0.6779	0.6376	20.2737
Ref. [19]	0.7700	0.7400	16.17
STRIQE [20]	0.8504	0.8413	11.60
2D + 3D features	<b>0.9326</b>	<b>0.9133</b>	<b>7.6757</b>

In conclusion, we have presented a new framework for BSIQA accounting for the LR-similarity map and the sensitivity of HVS to the structure. Our results demonstrate that the proposed model is promising in handling the quality assessment problem of stereoscopic images. The extraordinary contribution of this work is that we construct the first effective LR-similarity map and the experimental results prove the validity of LR-similarity map. This is just the results of our preliminary research, future work will be focused on further studying the LR-similarity map and exploring more effective features.

**Funding.** National Natural Science Foundation of China (NSFC) (61471260, 61271324); Natural Science Foundation of Tianjin (16JCYBJC16000).

## REFERENCES

1. D. Zhao, B. Su, G. Chen, and H. Liao, *Opt. Express* **23**, 9812 (2015).
2. J. Wang, Y. Song, Z. Li, A. Kempf, and A. He, *Opt. Lett.* **40**, 1231 (2015).
3. Y. Gong, D. Meng, and E. J. Seibel, *Opt. Express* **23**, 10771 (2015).
4. Q. Jiang, F. Shao, G. Jiang, M. Yu, Z. Peng, and C. Yu, *Signal Process.* **38**, 57 (2015).
5. J. Kim, P. V. Johnson, and M. S. Banks, *Opt. Express* **22**, 31924 (2014).
6. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, *IEEE Trans. Image Process.* **13**, 600 (2004).
7. W. J. Zhou, L. Yu, and M. W. Wu, *Opt. Express* **23**, 23710 (2015).
8. J. Yang, Y. Liu, Z. Gao, R. Chu, and Z. Song, *J. Visual Commun. Image Represent.* **31**, 138 (2015).
9. M. J. Chen, C. C. Su, D. K. Kwon, L. K. Cormack, and A. C. Bovik, *Signal Process.* **28**, 1143 (2013).
10. F. Shao, K. Li, W. Lin, G. Jiang, and M. Yu, *IEEE Signal Process. Lett.* **22**, 1548 (2015).
11. A. K. Moorthy and A. C. Bovik, *IEEE Trans. Image Process.* **20**, 3350 (2011).
12. A. K. Moorthy and A. C. Bovik, *IEEE Signal Process. Lett.* **17**, 513 (2010).
13. T. Ojala, M. Pietikainen, and T. Maenpaa, *IEEE Trans. Pattern Anal. Mach. Intell.* **24**, 971 (2002).
14. Q. Li, W. Lin, and Y. Fang, *IEEE Signal Process. Lett.* **23**, 541 (2016).
15. A. Benoit, P. Le Callet, P. Campisi, and R. Cousseau, *EURASIP J. Image Video Process.* **2008**, 1 (2009).
16. Y. H. Lin and J. L. Wu, *IEEE Trans. Image Process.* **23**, 5586 (2014).
17. F. Shao, K. Li, G. Jiang, M. Yu, and C. Yu, *Appl. Opt.* **54**, 9671 (2015).
18. M.-J. Chen, L. K. Cormack, and A. C. Bovik, *IEEE Trans. Image Process.* **22**, 3379 (2013).
19. S. Ryu and K. Sohn, *IEEE Trans. Circuits Syst. Video Technol.* **24**, 591 (2014).
20. S. K. Md, B. Appina, and S. S. Channappayya, *IEEE Signal Process. Lett.* **22**, 1985 (2015).