Stereoscopic Image Quality Assessment Using Wavelet Decomposition and Natural Scene Statistics

Xianqiu Geng[(✉)], Liquan Shen, Yang Yao, and Ping An

School of Communication and Information Engineering,
Shanghai University, Shanghai 200072, China
`jsslq@163.com`

Abstract. Recently, stereoscopic image quality assessment (SIQA) has been attracted more attention in academia and industry nowadays. In this paper, a wavelet decomposition and natural scene statistics based no reference stereoscopic image quality assessment algorithm is proposed. Our motivation is based on the observation that the statistics of the wavelet coefficients can be effectively captured by a generalized Gaussian distribution (GGD), and the distributions of image with different distortion have different shape and spread. The fitting parameters of GGD are extracted as the features. In this paper, stereoscopic image relevant information including stereo pairs, cyclopean image and binocular disparity are regarded as the factors that affecting stereoscopic image quality, and they are involved in the process of feature extraction. Support vector regression (SVR) is utilized to learn a regression model to predict the quality of stereoscopic image. Experimental results demonstrate that the proposed algorithm achieves high consistency with subjective assessment on two public available 3D image quality assessment databases.

Keywords: Wavelet decomposition · Natural scene statistics · Generalized Gaussian distribution · Feature extraction · Support vector regression

1 Introduction

Recently, many researches have focused on the development of image quality assessment (IQA), and stereoscopic/3D-IQA has becoming a super-hot.

The existing stereoscopic image quality assessment (SIQA) methods can be classified into full-reference (FR) [1–3], reduced-reference (RR) [4], and no-reference (NR) methods [5–9]. Our work aims emphasis on no-reference stereoscopic image quality assessment (NR-SIQA), in which no reference image information is available. Akhter et al. [5] propose a NR-SIQA algorithm by extracting segmented local features of artifacts from stereo pairs and the estimated disparity map. Chen et al. [6] first use the binocular fusion model to compute cyclopean image from left and right images of stereo pairs. Features including 2D features and 3D features are extracted in spatial domain by natural scene statistics (NSS). In [7], Su et al. also consider the binocular combination model to generate a convergent cyclopean image from left and right images of stereo pairs. Spatial domain univariate NSS features, wavelet domain univariate NSS features,

and bivariate density and correlation NSS features are extracted from the convergent cyclopean image. Zhou et al. [8] utilize the complementary local patterns of binocular energy response and binocular rivalry response to simulate the binocular visual perception. The local patterns of the binocular responses' encoding maps are used to form various binocular quality-predictive features. In [9], Wang et al. construct feature vectors from binocular energy response and then use the machine learning method to learn a visual quality prediction model. However, the problem is that all these existing methods do not simultaneously take perceptual factors affecting 2D image quality and 3D stereo perception into consideration. Furthermore, 3D visual characteristics are partial simulated. In order to design a well-defined NR-SIQA method, quality relevant features from the independent left and right images could be useful. Moreover, features from transformed domain of cyclopean images are supplements. Others 3D visual perception information should be involved.

In this paper, we propose a perceptual NR-SIQA algorithm, in which quality relevant information including stereo pairs, cyclopean image and binocular disparity are considered. Features are extracted from the wavelet coefficients using natural scene statistics in the wavelet domain when stereo pairs, cyclopean image and binocular disparity map are transformed by the steerable pyramid decomposition. Support vector regression (SVR) is utilized to learn a regression model to predict the quality of stereoscopic image.

The remainder of this paper is organized as follows. The proposed algorithm is described in detail in Sect. 2. Experimental results are analyzed in Sect. 3, and finally conclusions are drawn in Sect. 4.

2   Proposed Algorithm

The flowchart of the proposed NR-SIQA algorithm is shown in Fig. 1.
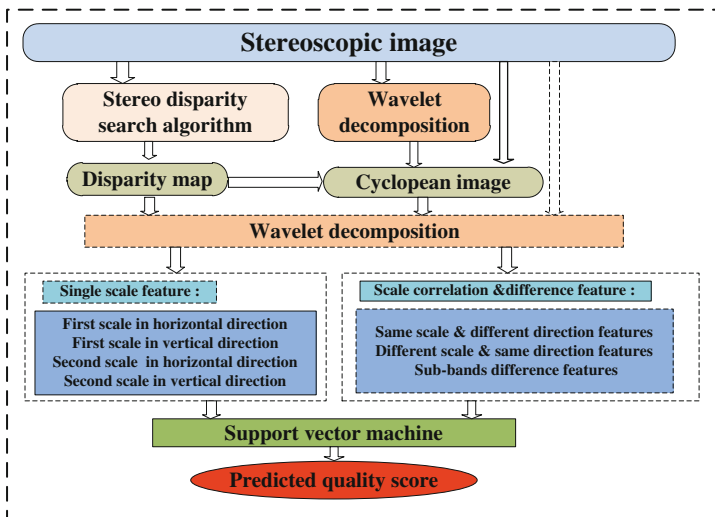

Fig. 1. Flowchart of the proposed algorithm

2.1 Binocular Disparity Search

The binocular disparity of one point in 3D image is the distance between two projected points in the left image and the right image, and the estimate of disparity for any point in left image is to find the same point in the right image.

In this paper, a Gaussian average SSIM based disparity search algorithm is proposed. SSIM index [10] measures the similarity between two image patch by

$$SSIM(1, r) = \frac{(2\mu_l\mu_r + c_1)(2\sigma_{lr} + c_2)}{(\mu_l^2 + \mu_r^2 + c_1)(\sigma_l^2 + \sigma_r^2 + c_2)} \tag{1}$$

where μ_l and μ_r are the average of left image patch and the right mage patch. σ_l^2 and σ_r^2 are the variance of left image and right image. σ_{lr} is the covariance of l and r. c_1 and c_2 are two parameters avoiding meaningless of the equation

In the process of disparity search, the leftmost part of the left image and the rightmost part of the right image should be discarded, since they cannot be captured by both cameras. For any point in left image with coordinate $[x_l, y_l]$, its related point in the right image is search from $[x_l - range, y_l]$ to $[x_l + range, y_l]$. In our experiment, the range is 32. The matched point in the right image has the largest similarity between the current point in the left image and the matched point in the search range of right image. The SSIM value between point in the left block and related point in the right image is calculated, and all the SSIM values are merged together with a Gaussian weighted sum. Figure 2 shows the stereo pairs and disparity image. Figures 2(a) and (b) denote the right view and left view of stereo pairs with high quality, respectively. The estimated disparity image of the stereo pairs is shown in Fig. 2(c). Figures 2(d) and (e) denote stereo pairs with gauss noise distortion, and the estimated disparity is shown in Fig. 2(f).

(a)   (b)   (c)

(d)   (e)   (f)
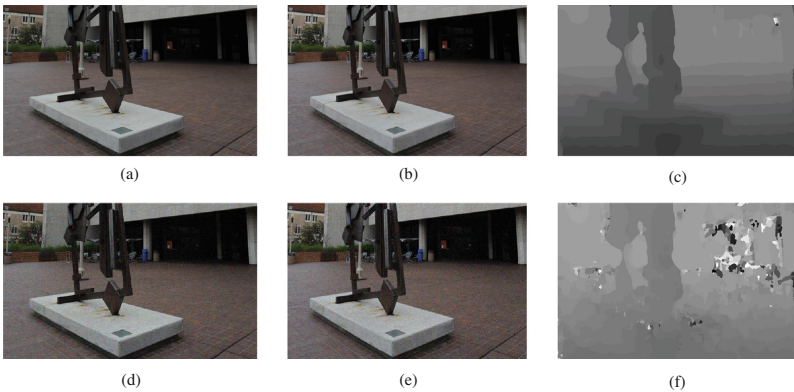
Fig. 2. Stereo pairs and disparity image

2.2 Cyclopean Image Generation

Cyclopean image is a single metal image of a scene created by the brain when combining two images from the two eyes. This process can be explained by binocular vision combination characteristic.

In this paper, the latest biological model, Gain-Control Theory model is utilized to simulate binocular fusion and explain cyclopean perception. The first step of cyclopean image generation is to decide the base image between the left image and the right image. In our implement, the image with better quality is selected as the base image, while the other one image is the aid image.

If the left image is selected as base image, the synthesized cyclopean image I(x, y) is calculated as follows,

$$
\begin{aligned}
I(x, y) = {}& \omega_L(x, y) \cdot I_L(x, y) \\
& + \omega_R\big(x - D_L(x, y), y\big) \cdot I_R\big(x - D_L(x, y), y\big)
\end{aligned}
\tag{2}
$$

where

$$
\omega_L = \frac{E_L(x, y)}{E_L(x, y) + E_R(x - D_L(x, y), y)}; \quad \omega_R = \frac{E_R(x - D_L(x, y), y)}{E_L(x, y) + E_R(x - D_L(x, y), y)}
$$

While if the right image is selected as base image, the synthesized cyclopean image I(x, y) can be calculated as,

$$
\begin{aligned}
I(x, y) = {}& \omega_L\big(x + D_R(x, y), y\big) \cdot I_L\big(x + D_R(x, y), y\big) \\
& + \omega_R(x, y) \cdot I_R(x, y)
\end{aligned}
\tag{3}
$$

where

$$
\omega_L = \frac{E_L\big(x + D_R(x, y), y\big)}{E_L\big(x + D_R(x, y), y\big) + E_R(x, y)}; \quad \omega_R = \frac{E_R(x, y)}{E_L\big(x + D_R(x, y), y\big) + E_R(x, y)}
$$

In Eqs. (2) and (3), $D_L(x, y)$ and $D_R(x, y)$ are the estimated disparity map computed using the left image and right image as the base image, respectively. ω_L and ω_R are weighting maps. E_L and E_R denote the sum of the energies of wavelet coefficient computed using a steerable pyramid. The stereo pairs and synthesized cyclopean image are shown in Fig. 3. Figures 3(a) and (b) show the stereo pairs with no distortion existing, and the synthesized cyclopean image computed from Fig. 3(a) and (b) is shown in Fig. 3(c). Stereo pairs with gauss distortion are shown in Fig. 3(d) and (e), respectively. Figure 3(f) shows the cyclopean image with gauss distortion.

(a)   (b)   (c)

(d)   (e)   (f)
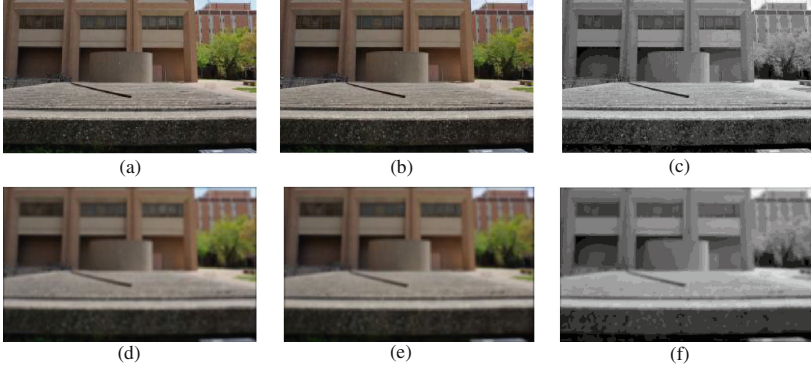
Fig. 3. Stereo pairs and cyclopean image

2.3   Features Extraction

The strategy that adopted for feature extraction is based on natural scene statistics (NSS), which has been proved to be effective in image quality assessment [11].

To extract features, the estimated disparity image, synthesized cyclopean image and image with better quality from the stereo pairs are processed by wavelet decomposition to form oriented band-pass responses. The motivation is based on the fact that the scale-space-orientation decomposition achieves a high performance in modeling the visual signal processing mechanism, which occurs in the primary visual cortex of human visual system (HVS).

In our implement, we perform wavelet decomposition over 2 scales and 6 orientations. Thus, 12 sub-bands across scales and orientations labeled $S_\beta^\theta \left(\beta \in \{1, 2\} \text{ and } \theta \in \left\{ 0°, 30°, 60°, 90°, 120°, 150° \right\} \right)$ are obtained for each image. Then, a series of statistics features will be extracted from the obtained sub-band coefficients as follow.

2.3.1   Single Sub-band-Based NSS Feature

To extract single scale feature, it is not necessary to use all the sub-bands, since some of the sub-bands may be correlated with others. In this paper, four sub-bands (i.e., $S_1^{0°}, S_1^{90°}, S_2^{0°}$ and $S_2^{90°}$) are chosen to extract single sub-band based feature.

The single sub-band coefficients statistics of the cyclopean image in Fig. 3(c) for various distortions are shown in Fig. 4. It demonstrates that the probability density distribution of the sub-band coefficients exhibits a Gaussian-like appearance, which can be effectively captured by a generalized Gaussian distribution (GGD). The density function of GGD with zero mean is given by,

$$f(x; \alpha, \sigma^2) = \frac{\alpha}{2\beta\Gamma(1/\alpha)} \exp\left(-\left(\frac{|x|}{\beta} \right)^\alpha \right) \tag{4}$$

where $\beta = \sigma \sqrt{\dfrac{\Gamma(1/\alpha)}{\Gamma(3/\alpha)}}$, and $\Gamma(.)$ is the gamma function: $\Gamma(\alpha) = \int_0^\infty t^{a-1} e^{-t} dt \ a > 0$. In the model of GGD, the shape parameter α controls the real shape of the distribution, and σ^2 controls the variance. We use the parameter α and σ^2 as the quality relevant features, which can be reliably estimated using the moment-matching based approach in [12]. Features can be extracted from the wavelet coefficients of cyclopean image, disparity image and stereo pairs.
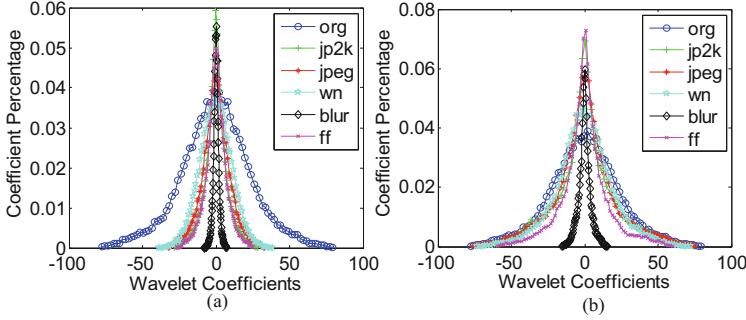

Fig. 4. Single sub-band statistics of the cyclopean image

2.3.2 Spatial Correlation-Based NSS Feature

There is a high correlation between sub-bands over the same scale and different orientation, as well as sub-bands over different scale and same orientation.

In Fig. 5(a), we plot the probability density distribution of coefficients formed by $[S_1^{0°} : S_1^{90°}]$ to demonstrate the sub-bands correlation over the same scale and different orientation. To show the correlation over different scale and same orientation, the probability density distribution of coefficients formed by $[S_1^{0°} : S_2^{0°}]$ is shown in Fig. 5(b).
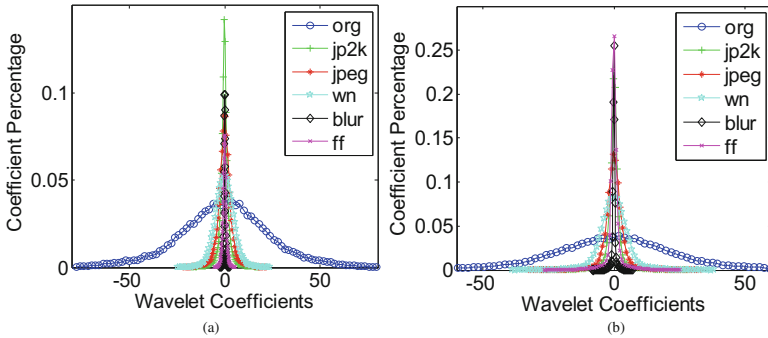

Fig. 5. Sub-bands statistics of spatial correlation

To extract features, the GGD model can also be utilized to fit the Gaussian-like distortion. The parameter (α, σ^2) of GGD is extracted as features from the wavelet coefficients of cyclopean image, disparity image and stereo pairs.

2.3.3 Spatial Difference-Based NSS Feature

In our work, we calculate the difference between two sub-bands coefficient at the same scale across different orientation as,

$$D_1^{(\theta_i - \theta_j)} = S_1^{\theta_i} - S_1^{\theta_j}; D_2^{(\theta_i - \theta_j)} = S_2^{\theta_i} - S_2^{\theta_j} \tag{5}$$

where $\theta_i \in \{0^\circ\}$ and $\theta_j \in \{0^\circ, 30^\circ, 60^\circ, 90^\circ, 120^\circ, 150^\circ\}$.

The probability density distribution of $D_1^{(0^\circ - 90^\circ)}$ and $D_2^{(0^\circ - 90^\circ)}$ computed from the cyclopean image are shown in Fig. 6(a) and (b). Similarly, to extract features from spatial difference, the GGD model is used to fit the distributions of the difference. Thus, the parameter (α, σ^2) of GGD computed from cyclopean image, disparity image and stereo pairs are extracted as features.
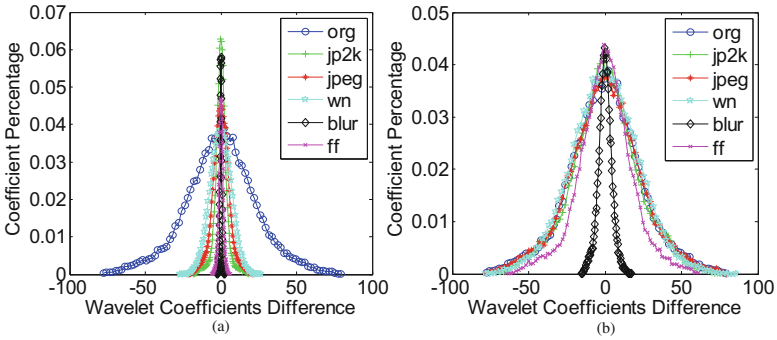
Fig. 6. Sub-band statistics of spatial difference

2.4 Quality Prediction

Machine learning is applied to map these features to quality score. Specifically, in the train phase, the best map between features and subjective quality scores (MOSs) included in the 3D image databases is obtained by the regression module. In the test phase, correlation between predicted objective scores obtained by SIQA algorithm and subjective scores is received. Multiple iterations of the above training and testing procedure are performed by varying the splitting of data over the training and test sets.

3   Experimental Results and Analysis

3.1   Databases and Evaluation Criteria

The proposed 3D-IQA method is evaluated on two publicly available subject-rated image quality databases: LIVE 3D IQA Database Phase I [13] and LIVE 3D IQA Database Phase II [6]. The LIVE 3D IQA Database Phase I consists of 20 reference images and 365 distorted images. Five types of distortions: JPEG and JPEG2000 (JP2K) compression, Gaussian blur (GB), White noise (WN) and a Rayleigh fast-fading (FF) are symmetrically applied to the left and right reference images at various levels. The LIVE 3D IQA Database Phase II consists of 120 symmetrically and 240 asymmetrically distorted images generated from 8 reference images, with the same distortions as Database Phase I.

Three criterions including Pearson linear correlation coefficient (PLCC), Spearman rank order correlation coefficient (SROCC) and root mean squared error (RMSE) are used to evaluate the SIQA metrics. PLCC and RMSE are used to evaluate prediction accuracy of SIQA metrics, and SROCC is used for prediction monotonicity. Before computing these performance criterions, a nonlinear regression analysis is utilized to provide a nonlinear mapping between the objective scores and subjective mean opinion scores (MOSs). For the nonlinear regression, a five-parameter logistic function [14] is used,

$$f(x) = \beta_1 \cdot \left(\frac{1}{2} - \frac{1}{1 + e^{\beta_2(x-\beta_3)}} \right) + \beta_4 x + \beta_5 \tag{6}$$

where $\beta_i, i = 1, 2 \dots 5$, are parameters determined by the best fit of subjective scores and the objective scores. Higher SROCC and PLCC and lower RMSE values demonstrate better objective SIQA metrics.

3.2   Overall Performance Comparison

The overall performance comparison of proposed algorithm and other NR-SIQA methods on both LIVE 3D image database Phase I and Phase II is shown in Table 1. The top metric has been highlighted in boldface. It can be seen from Table 1 that the proposed algorithm significantly outperforms (i.e., PLCC and SROCC are highest, while RMSE is lowest) all the considered other metrics on both the two databases. Meanwhile, the proposed metric and other NR-SIQA metrics achieve a better performance on LIVE 3D image quality database phase I than those on database phase II. The reason is that LIVE 3D image quality database phase I contains only symmetric distortions, while phase II contains both symmetric and asymmetric distortions. It is much more difficult to assess the stereoscopic images with asymmetric distortions. That is because of the limited understanding of human visual system when human eyes watching asymmetric stereoscopic images.

Table 1. Overall performance on LIVE Phase I and Phase II

Database	Live Phase I			Live Phase II		
Criteria	SROCC	PLCC	RMSE	SROCC	PLCC	RMSE
Akhter [5]	0.383	0.626	14.827	0.543	0.568	9.249
Chen [6]	0.891	0.895	7.247	0.880	0.895	5.102
Su [7]	–	–	–	0.905	0.913	4.657
Zhou [8]	0.887	0.928	6.025	0.823	0.861	5.779
Wang [9]	0.828	0.885	7.238	0.794	0.784	7.236
Proposed	**0.934**	**0.945**	**5.433**	**0.915**	**0.928**	**4.215**

We also evaluate our proposed algorithm on each type of distortion. To make a comprehensive comparison, several state-of-the-art FR-SIQA metrics (i.e., Benoit [1], You [2], and Chen [3]) are also considered. Performances on LIVE 3D image database Phase I and Phase II are listed in Tables 2 and 3, respectively. The top two metrics are highlighted in boldface, and italicized algorithms are FR-SIQA algorithms. It can be seen that the proposed metric belongs to the top two metrics, which mean that the proposed method can predict the stereoscopic image quality consistently across different types of distortions. For FR-SIQA metrics, Benoit's and You's methods are based on 2D-IQA algorithms and simultaneously consider disparity information. Chen's algorithm in [3] uses the SSIM index to assess the quality of cyclopean image. Even though the reference information is available in these FR-SIQA metrics, the performance is not better than our method. For NR-SIQA metrics, the scheme in [6] extracts NSS-model-based features of cyclopean in spatial domain, but others 3D visual perception are not considered adequately. Therefore, the performance is worse than our proposed algorithm.

Table 2. Performance on each type of distortion on LIVE Phase I

		FR-SIQA			NR-SIQA			
		Benoit [1]	You [2]	Chen [3]	Akhter [5]	Chen [6]	Su [7]	**Proposed**
PLCC	JPEG	*0.640*	*0.487*	*0.603*	**0.905**	**0.907**	–	0.861
	JP2K	*0.939*	***0.939***	*0.912*	0.904	0.917	–	**0.946**
	WN	*0.925*	*0.925*	***0.942***	0.729	0.695	–	**0.975**
	GB	*0.948*	***0.948***	*0.942*	0.617	0.917	–	**0.980**
	FF	*0.747*	*0.747*	***0.776***	0.503	0.735	–	**0.856**
SROCC	JPEG	*0.603*	*0.439*	*0.530*	**0.866**	**0.863**	–	0.846
	JP2K	*0.910*	*0.860*	*0.888*	0.914	**0.919**	–	**0.942**
	WN	*0.930*	*0.940*	***0.948***	0.675	0.617	–	**0.960**
	GB	***0.931***	*0.882*	***0.925***	0.555	0.878	–	0.910
	FF	*0.699*	*0.588*	***0.707***	0.640	0.652	–	**0.747**
RMSE	JPEG	***5.022***	*5.709*	*5.216*	5.438	5.402	–	**3.076**
	JP2K	***4.426***	*6.206*	*5.320*	7.092	6.433	–	**3.378**
	WN	*6.307*	*5.621*	*5.581*	**4.273**	4.532	–	**3.601**
	GB	***4.571***	*5.679*	*4.822*	11.387	5.898	–	**2.786**
	FF	*8.257*	*8.429*	***7.837***	9.332	8.322	–	**6.255**

Table 3. Performance on each type of distortion on LIVE Phase II

		FR-SIQA			NR-SIQA			
		Benoit [1]	You [2]	Chen [3]	Akhter [5]	Chen [6]	Su [7]	**Proposed**
PLCC	JPEG	*0.853*	*0.830*	*0.862*	0.776	**0.899**	**0.888**	0.854
	JP2K	*0.784*	*0.905*	*0.834*	0.722	**0.947**	0.847	**0.933**
	WN	*0.926*	*0.912*	***0.957***	0.786	0.901	**0.953**	0.941
	GB	*0.535*	*0.784*	*0.963*	0.795	0.941	**0.968**	0.966
	FF	*0.807*	*0.915*	*0.901*	0.674	0.932	**0.944**	0.923
SROCC	JPEG	***0.867***	*0.795*	*0.843*	0.724	**0.867**	0.818	0.786
	JP2K	*0.751*	*0.894*	*0.814*	0.714	**0.950**	0.845	**0.907**
	WN	*0.923*	*0.909*	***0.940***	0.649	0.867	**0.946**	0.902
	GB	*0.455*	*0.813*	*0.905*	0.682	0.900	0.903	**0.906**
	FF	*0.773*	*0.891*	*0.884*	0.559	**0.933**	0.899	**0.923**
RMSE	JPEG	***3.787***	*4.086*	*3.865*	6.189	4.298	4.169	**3.337**
	JP2K	*6.096*	*4.186*	*5.562*	7.416	**3.513**	5.482	**3.224**
	WN	*4.028*	*4.396*	*3.368*	4.535	**3.342**	3.547	**3.289**
	GB	*11.763*	*8.649*	***3.747***	8.450	4.725	4.453	**3.177**
	FF	*6.894*	*4.649*	*4.966*	8.505	**4.180**	4.199	**4.120**

4   Conclusion

In this paper, we propose a novel no-reference quality assessment for stereoscopic image using wavelet decomposition and natural scene statistics. We find that the statistics of the wavelet coefficients can be effectively captured by a generalized Gaussian distribution (GGD). Therefore, the parameters of GGD are extracted as the quality relevant features. To extract feature, stereo pairs, cyclopean image and binocular disparity image are considered. Finally, machine learning is applied to map these features to quality score. Experimental results demonstrate that the proposed algorithm achieves high consistency with subjective assessment on two public available 3D image quality assessment databases.

References

1. Benoit, A., Le Callet, P., Campisi, P., Cousseau, R.: Quality assessment of stereoscopic images. EURASIP J. Image Video Process. (2008)
2. You, J., Xing, L., Perkis, A., Wang, X.: Perceptual quality assessment for stereoscopic images based on 2D image quality metrics and disparity analysis. In: VPQM, Scottsdale, Arizona, USA (2010)

3. Chen, M.J., Su, C.C., Kwon, D.K., et al.: Full-reference quality assessment of stereo pairs accounting for rivalry. Sign. Process. Image Commun. **28**(9), 1143–1155 (2013)
4. Qi, F., Zhao, D., Gao, W.: Reduced reference stereoscopic image quality assessment based on binocular perceptual information. IEEE Trans. Multimedia **17**(12), 2338–2344 (2015)
5. Akhter, R., Baltes, J., Parvez Sazzad, Z.M., Horita, Y.: No reference stereoscopic image quality assessment. In: Proceedings of SPIE, vol. 7524, p. 75240T, February 2010
6. Chen, M.J., Cormack, L.K., Bovik, A.C.: No-reference quality assessment of natural stereopairs. IEEE Trans. Image Process. **22**(9), 3379–3391 (2013)
7. Su, C.C., Cormack, L.K., Bovik, A.C.: Oriented correlation models of distorted natural images with application to natural stereopair quality evaluation. IEEE Trans. Image Process. **24**(5), 1685–1699 (2015)
8. Zhou, W., Yu, L.: Binocular responses for no-reference 3D Image quality assessment. IEEE Trans. Multimedia **18**(6), 1077–1084 (2016)
9. Wang, S., Shao, F., Jiang, G.: Supporting binocular visual quality prediction using machine learning. In: IEEE International Conference on Multimedia and Expo Workshops (ICMEW 2014), Chengdu, pp. 1–6 (2014)
10. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)
11. Mittal, A., Moorthy, A.K., Bovik, A.C.: No-reference image quality assessment in the spatial domain. IEEE Trans. Image Process. **21**(12), 4695–4708 (2012)
12. Sharifi, K., Leon-Garcia, A.: Estimation of shape parameter for generalized Gaussian distributions in subband decompositions of video. IEEE Trans. Circuits Syst. Video Technol. **5**(1), 52–56 (1995)
13. Moorthy, A.K., et al.: Subjective evaluation of stereoscopic image quality. Sign. Process. Image Commun. **28**(8), 870–883 (2013)
14. Sheikh, H.R., Sabir, M.F., Bovik, A.C.: A statistical evaluation of recent full reference image quality assessment algorithms. IEEE Trans. Image Process. **15**(11), 3440–3451 (2006)