Hindawi
Mathematical Problems in Engineering
Volume 2019, Article ID 1384921, 7 pages
https://doi.org/10.1155/2019/1384921

Research Article

Blind Stereo Image Quality Evaluation Based on Convolutional Network and Saliency Weighting

Wujie Zhou

School of Information and Electronic Engineering, Zhejiang University of Science and Technology, Hangzhou 310023, China

Correspondence should be addressed to Wujie Zhou; wujiezhou@163.com

Received 7 June 2019; Revised 26 July 2019; Accepted 23 August 2019; Published 9 September 2019

Academic Editor: Daniel Zaldivar

With the rapid development of stereo image applications, there is an increasing demand to develop a versatile tool to evaluate the perceived quality of stereo images. Therefore, in this study, a blind stereo image quality evaluation (SIQE) algorithm based on convolutional network and saliency weighting is proposed. The main network framework used by the algorithm is the quality map generation network, which is used to train the distortion image dataset and quality map label to obtain an optimal network framework. Finally, the left view, right view, and cyclopean view of the stereo image are used as inputs to the network frame, respectively, and then weighted fusion for the final stereo image quality score. The experimental results reveal that the proposed SIQE algorithm can improve the accuracy of the image quality prediction and prediction score to a certain extent and has good generalization ability.

1. Introduction

With the rapid development of stereo image applications, many related stereo image technologies and services have been introduced in our daily lives as well as in many professional fields [1–9]. A variety of distortions can occur during the collection, transmission, processing, and displaying of stereo images [10–19]. Therefore, it is of immense practical significance to establish a high-performance stereo image quality evaluation method. Stereo image quality evaluation (SIQE) is classified into objective and subjective evaluation. Subjective evaluation is the subjective evaluation of images directly by humans. Because the human visual system is the ultimate recipient of images, subjective evaluation is very persuasive. However, in practical applications, subjective evaluation becomes extremely time-consuming and laborious and is difficult to apply to real-time systems. Therefore, objective evaluation plays a dominant role in SIQE. There are three main types of objective stereo image quality evaluation methods: full-reference (FR) evaluation [4–8], reduced-reference (RR) evaluation [9], and no-reference (NR) evaluation [10–22]. The FR evaluation method compares the undistorted original image with the distorted image to obtain the difference between them. The reduced-reference evaluation method uses the partially undistorted original image information. The NR evaluation method does not use the undistorted original image at all. Because the original undistorted image is difficult to obtain in practical applications, the NR evaluation method has a higher research value.

Many SIQA methods are designed based on the typical 2D image quality evaluation methods [1–3]. The typical full-reference SIQE method was studied, which was proposed by Chen et al. [4], mainly uses a stereo image pair, disparity image, and Gabor filter response synthesis image. The central eye image uses an FR 2D image quality assessment method to predict the 3D quality score. Zhou et al. [11] proposed a new NR SIQE based on binocular self-similarity and deep neural networks. In [12], the stereo image block is first input to the convolutional neural network and then pooling. The final image quality score is obtained through the multilayer perceptron, where the initial parameters of the convolutional neural network are trained by a large number of natural images. Jiang et al. [13] performed SIQE by learning color visual characteristics based on nonnegative matrix factorization and considering binocular interaction.

The study [14] proposes an NR stereo image quality assessment based on the combination of wavelet decomposition and statistical models. The SIQE method proposed in [15, 16] and the method proposed in [15] are based on the evaluation method of binocular vision mechanism. The method proposed in [16] is based on the gradient dictionary color visual feature learning evaluation method. Wu et al. [17] proposed an evaluation method based on depth edge information and color signals, which also uses segmented self-encoders. Liu et al. [18] proposed an SIQE method based on classification and prediction. Jiang et al. proposed an SIQE algorithm for processing multiple distortions [19], which characterizes the local receptive field characteristics of the visual cortex by learning monocular and binocular local visual primitives for various distorted stereo images quality assessment and single distortion stereo images. Bensalma and Larabi [20] proposed perceptual quality metric for stereo images based on the binocular energy. Zhou et al. [21] proposed blind SIQE metric based on binocular combination and extreme learning machine.

Although the above methods have achieved certain effects on the stereo image evaluation problem, these evaluation methods do not consider versatility or image saliency weighting. Moreover, few studies consider the left and right views weight assignment problems of stereo images. Therefore, in this study, an NR SIQE method based on convolutional networks and saliency weighting is proposed. The main contributions of this work are as follows.

First, the training dataset used in the proposed method is a self-made distortion image dataset, and the quality map obtained by the high-performance FR image quality evaluation method is used to render corresponding labels.

Second, the main network framework used by the algorithm is the quality map generation network, which is used to train the distortion image dataset and quality map label to obtain an optimal network framework.

Finally, the left view, the right view, and the cyclopean view of the stereo image are, respectively, used as the input of the quality map generation optimal network framework, and the corresponding quality maps are predicted. Weighted fusion is sequentially performed to obtain the final stereo image quality score.

2. Model Methods

Figure 1 illustrates the overall frame structure of the proposed method. The inputs to the network frame are a left-view distortion image, right-view distortion image, and cyclopean view of the left and right views. The output is a predicted value of the distortion stereo image quality score. As can be seen from the figure, the overall framework can be divided into three submodules, the quality map generation network, the saliency weighting module, and the weighted summation module of the left, right, and cyclopean views. The specifications of each module are described in detail in the following sections.

2.1. Quality Map Generation Network. In the proposed method, the quality map generation network is a main component of the overall framework. The requirement for the quality map generative network is outputting a quality map of the same size with the input image. We construct an improved framework of U-Net, an extension of fully convolutional networks, as a base of quality map generative network because U-Net integrates the hierarchical representations in subsampling layers with the corresponding features in upsampling layers. Hence, these layers increase the resolution of the output. In order to localize, high-resolution features from the contracting path are combined with the upsampled output. A successive convolution layer can then learn to assemble a more precise output based on this information.

A major component of the quality map generation network is the convolutional layer. In the encoder part, the basic structure consists of a two-layer convolution plus a layer of pooling as a small module. The number of convolution kernels in each convolutional layer is depicted in Figure 2. The size of the input distortion image is $w \times h \times 3$, the size of the output quality map is $w \times h$, and the size of the convolution kernel is 3×3. The activation function uses a linear correction unit (ReLU) function. Let w_i^l denote the parameters of lth filter kernel in the ith convolution layer and b_i^l denote its corresponding bias. Then, the lth feature map produced in the ith convolution layer is represented by

$$y_i^l = s\left(w_i^l y_i^{l-1} + b_i^l\right), \tag{1}$$

where h_{i-1} denote $i-1$th feature maps outputted from previous convolution layer and h_0 corresponds to the input image. $s(\cdot)$ is the ReLU function.

The supervisory label used in the network is based on the image structure similarity (SSIM) quality map, and the effect of locally obtaining the SSIM quality map is better than that of globally obtaining it.

In this work, we first select 100 source images from a dataset [23] for the training set. The source images have different scenes (all reference images of the dataset are shown in Figure 2, resized to a fixed-size of 528×400). Four commonly observed distortion types, namely, JPEG 2000 (JP2K) compression, JPEG compression, Gaussian blur (GB), and white noise (WN) are used to generate the distorted images. Finally, the SSIM metric is employed to generate the ground-truth objective quality/similarity maps as training labels. In the proposed method, the SSIM quality map used is the local mapping matrix. Let x and y be two image signals. The original and distorted image signals are obtained, thereby obtaining a quality map based on the similarity, and the generation formula of the SSIM quality map is defined as

$$\mathrm{SSIM}(x, y) = \frac{\left(2\mu_x\mu_y + C_1\right)\left(2\sigma_{xy} + C_2\right)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)}, \tag{2}$$

where μ_x is the mean of image x, μ_y is the mean of image y, σ_x is the standard deviation of image x, σ_y is the standard deviation of image y, and σ_{xy} is the covariance of x and y and C_1 and C_2 denote small positive constant for increasing stability when the denominator approaches zero. In this paper, we set $C_1 = C_2 = 0.085$ for our experiments.

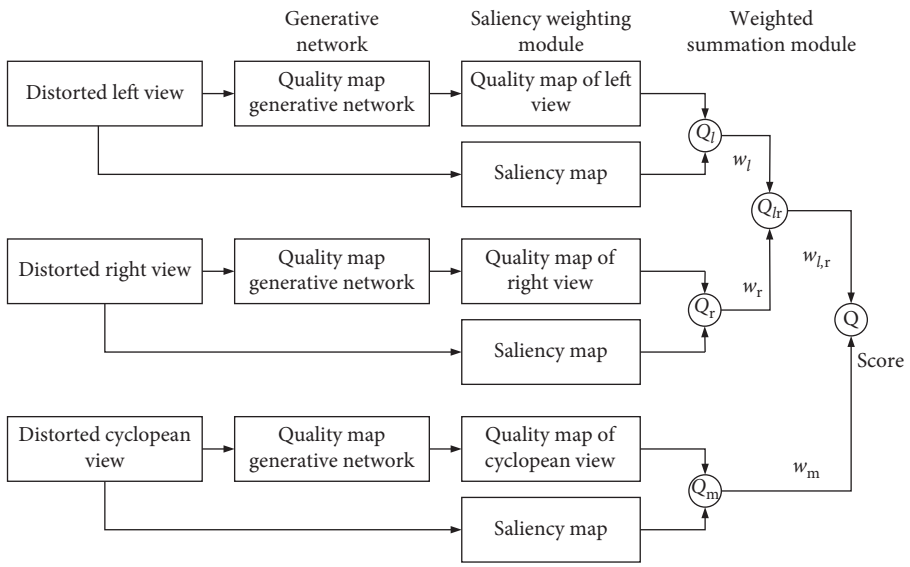

Figure 1: Overall structure diagram.

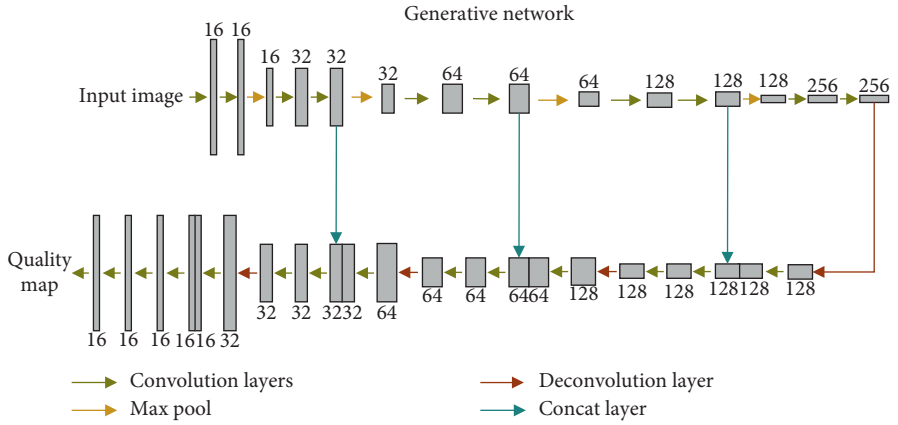

Figure 2: Quality map generation network.

2.2. Weighted Fusion Module. The weighted fusion module is one of the important components of the proposed method, and its main structure is depicted in Figure 1.

In this study, the fusion of the quality score of the left and right views is based on the widely used Bayesian theory [24], in which the binocular quality score can be obtained by

$$P\left(\vartheta \mid V_{\mathrm{l}}, \widehat{V}_{\mathrm{r}}\right) = \frac{P\left(V_{\mathrm{l}}, \widehat{V}_{\mathrm{r}} \mid \vartheta\right) \cdot P(\vartheta)}{P\left(V_{\mathrm{l}}, \widehat{V}\right)}$$

$$\approx \frac{P(V_{\mathrm{l}})}{P\left(V_{\mathrm{l}}, \widehat{V}_{\mathrm{r}}\right)} \cdot P(\vartheta \mid V_{\mathrm{l}}) + \frac{P\left(\widehat{V}_{\mathrm{r}}\right)}{P\left(I_{\mathrm{l}}, \widehat{V}_{\mathrm{r}}\right)} P\left(\vartheta \mid \widehat{V}_{\mathrm{r}}\right). \tag{3}$$

In (3), V_{l} and $\widehat{V}_{\mathrm{r}}$ denote the left retinal view and the disparity/depth-compensated right retinal view, respectively; ϑ denotes the quality; $P(V_{\mathrm{l}})/P(V_{\mathrm{l}}, \widehat{V}_{\mathrm{r}})$ and $P(\widehat{V}_{\mathrm{r}})/P(V_{\mathrm{l}}, \widehat{V}_{\mathrm{r}})$ denote feature distributions that are utilized to balance the roles of binocular visual mechanism in determining the overall visual quality; and the likelihoods $P(\vartheta \mid V_{\mathrm{l}})$ and $P(\vartheta \mid \widehat{V}_{\mathrm{r}})$ denote quality scores of the left and right views.

The left and right views weights of the distorted stereo image are referred to the left and right views weight assignment methods of the paper [24], and the formula is as follows:

$$w_{\mathrm{l}} = \frac{g_{\mathrm{l}}^2}{g_{\mathrm{l}}^2 + g_{\mathrm{r}}^2}, \tag{4}$$

$$w_{\mathrm{r}} = \frac{g_{\mathrm{r}}^2}{g_{\mathrm{l}}^2 + g_{\mathrm{r}}^2}, \tag{5}$$

where g_{l} and g_{r} are the significant level estimates of the left and right views, respectively, and equations (4) and (5), respectively, represent the weights of the left and right views, thereby obtaining a weighted quality score of the left and right views quality scores:

Table 1: Weights of the left and right views quality scores and the cyclopean view quality scores assigned to the PLCC, SROCC, and RMSE performance coefficients.

Databases	Indicator	$\alpha = 0.1$	$\alpha = 0.2$	$A = 0.3$	$\alpha = 0.4$	$\alpha = 0.5$	$\alpha = 0.6$	$\alpha = 0.7$	$\alpha = 0.8$	$\alpha = 0.9$
LIVE 3D Phase I	PLCC	0.913	**0.913**	0.913	0.912	0.910	0.908	0.907	0.904	0.902
	SROCC	0.885	0.887	0.889	**0.890**	0.889	0.888	0.887	0.885	0.883
	RMSE	6.347	**6.337**	6.354	6.389	6.438	6.496	6.560	6.629	6.700
LIVE 3D Phase II	PLCC	0.858	**0.863**	0.861	0.856	0.849	0.841	0.833	0.824	0.816
	SROCC	0.849	**0.863**	0.849	0.842	0.833	0.825	0.818	0.812	0.802
	RMSE	5.739	**5.651**	5.690	5.787	5.910	6.045	6.186	6.329	6.470

$$Q_{lr} = w_l Q_l + w_r Q_r, \tag{6}$$

where $Q_l = (\sum S_{map,l} \times Q_{map,l}) / (\sum S_{map,l})$, $Q_r = (\sum S_{map,r} \times Q_{map,r}) / (\sum S_{map,r})$. $S_{map,l}$ and $S_{map,r}$ denote the saliency maps of the left and right views (we use the method of reference [25] to derive saliency maps from distorted views), respectively. $Q_{map,l}$ and $Q_{map,r}$ denote the predict quality maps of the left and right views, respectively.

In addition to the weights used in the combination of the left and right views quality scores, there is a set of weights that left and right views weighted score Q_{lr} and cyclopean view quality score $Q_m = (\sum S_{map,m} \times Q_{map,m}) / (\sum S_{map,m})$. $S_{map,m}$ denotes the saliency maps of the cyclopean views. $Q_{map,m}$ denotes the predict quality maps of the cyclopean views. For the distorted stereo images, the over quality score can be obtained by

$$Q = \alpha Q_{lr} + (1 - \alpha) Q_m, \tag{7}$$

where α is the weight. The selection of weight α is discussed in the experimental section.

3. Experimental Results and Analysis

3.1. Experimental Database. In the experiment, the performance of the proposed method was verified using two published three-dimensional image quality evaluation databases: LIVE 3D Phase I [26] and LIVE 3D Phase II [27]. Basic information of the two databases is provided below.

LIVE 3D Phase I: This database includes 20 pairs of original undistorted stereo images and 365 pairs of distorted stereo images. The size of each view is 640×360. All distorted images contain five different levels of distortion, including JPEG distortion, JPEG 2000 distortion, additive white Gaussian noise (WN) distortion, fast decay channel (FF) distortion, and Gaussian blur (GB) distortion. In addition, each pair of distorted images has a differential mean opinion score (DMOS), which is a human subjective image quality score obtained through a large number of experiments.

LIVE 3D Phase II (LIVE 3D Phase II-Symmetric and LIVE 3D Phase II-Asymmetric) database includes 8 pairs of original undistorted stereo images and 360 pairs of distorted stereo images. The size of each view is 640×360. Similar to LIVE 3D Phase I, all the distorted images contain 5 different levels of distortion. For each type of distortion, three sets of symmetrically distorted stereo image pairs and six sets of asymmetrically distorted stereo image pairs are generated for each pair of original stereo images. "Asymmetric" means that the left and right views of the stereo image have different types or different levels of distortion. Each pair of stereo images has a DMOS value.

3.2. Experimental Training Step. The experiment was implemented on a 64 bit memory computer using an NVIDIA GTX 1080 TI. The deep network framework was implemented using the Keras depth framework with TensorFlow as the backend. When training deep neural networks, the optimizer used is Adam, which, also known as "adaptive momentum estimation," is a parameter updating method that is used to calculate the adaptive learning rate for each parameter. The learning rate of some parameter updating methods is adjusted in the global scope, and the adjustment of all network parameters is equivalent, which makes the adjustment of the learning rate difficult, and requires very good initial network parameters, compared to these parameter updating methods. Adam has a bigger advantage. We use the Adam [28] optimization method with a learning rate of 1×10^{-4} with a decay of 0.5 every 50 epochs and a weight decay of 1×10^{-5} for regularization to optimize the network. In the iterative process, the batch size used is 16, and the loss function uses the mean square error (MSE).

3.3. Analysis of Experimental Results. The three classical performance evaluation indexes are used to evaluate the performance of the proposed SIQE methods, namely, Pearson linear correlation coefficient (PLCC), Spearman correlation coefficient (SROCC), and root mean square error (RMSE). The PLCC index can be used to evaluate the prediction accuracy of the NR image quality evaluation algorithm. The higher the correlation between the objective prediction score of the image and the subjective score is, the closer the PLCC correlation coefficient is to 1, indicating better performance of the image quality evaluation algorithm.

Since the planar database we created contains only four types of distortion, namely, JPEG, JPEG 2000, WN, and GB, only four types of distortions are tested and analyzed in the experiment. The data shown in Table 1 are the data for all four distortions.

The results of the weight distribution experiment when the left and right views quality scores are combined with the cyclopean view quality score are shown in Table 1. From the experiment, it can be concluded that, in the stereo distortion database, the best performance is obtained under this condition, that is, the value is $\alpha = 0.2$.

Table 2: Results of the proposed method in terms of the performance coefficients of PLCC, SROCC, and RMSE.

Databases	Indicator	JPEG	JPEG 2000	WN	GB	All
LIVE 3D Phase I	PLCC	0.621	0.908	0.941	0.915	0.913
	SROCC	0.592	0.881	0.931	0.857	0.887
	RMSE	5.124	5.439	5.632	5.845	6.337
LIVE 3D Phase II	PLCC	0.827	0.727	0.973	0.960	0.863
	SROCC	0.790	0.715	0.962	0.852	0.863
	RMSE	4.120	6.742	2.484	3.882	5.651

Table 3: Comparison of the results of the proposed method and several other methods in terms of the PLCC coefficient of performance.

Databases	Distortion	FR			Blind			
		SSIM	FSIM	GMSD	Chen	Bensalma	Shao	Proposed
LIVE 3D Phase I	JP2K	0.868	**0.937**	0.928	0.855	0.848	0.872	0.908
	JPEG	0.496	0.601	**0.652**	0.476	0.376	0.597	0.622
	WN	0.938	0.931	**0.947**	0.9533	0.914	0.916	0.941
	Gblur	0.912	0.933	**0.938**	0.939	0.916	0.923	0.915
	ALL	0.899	0.936	**0.943**	0.929	0.895	0.899	0.913
LIVE 3D Phase II-symmetric	JP2K	0.8162	0.8183	0.875	0.670	0.690	0.903	**0.898**
	JPEG	0.6770	0.8456	0.844	0.601	0.551	0.873	**0.890**
	WN	0.9749	0.9630	0.961	0.946	0.936	0.917	**0.982**
	GBLUR	0.8325	0.8638	0.928	0.918	0.953	**0.977**	0.953
	ALL	0.7326	0.8301	**0.925**	0.814	0.823	0.912	0.918
LIVE 3D Phase II-asymmetric	JP2K	0.676	0.785	0.868	0.722	0.619	0.789	**0.876**
	JPEG	0.685	0.796	**0.869**	0.564	0.631	0.705	0.699
	WN	0.823	0.941	0.916	0.945	0.933	0.924	**0.965**
	GBLUR	0.840	0.888	0.741	0.692	0.862	0.855	**0.951**
	ALL	0.750	0.678	0.653	0.634	0.743	0.565	**0.766**

Table 4: Comparison of the results of the proposed method and several other methods in terms of the SROCC coefficient of performance.

Databases	Distortion	FR			Blind			
		SSIM	FSIM	GMAD	Chen	Bensalm	Shao	Proposed
LIVE 3D Phase I	JP2K	0.867	0.901	**0.905**	0.871	0.817	0.900	0.881
	JPEG	0.456	0.563	**0.610**	0.435	0.328	0.607	0.594
	WN	0.938	0.930	**0.947**	0.939	0.906	0.927	0.931
	Gblur	0.899	0.925	**0.937**	0.921	0.918	0.924	0.858
	ALL	0.882	0.913	**0.922**	0.882	0.840	0.894	0.887
LIVE 3D Phase II-Symmetric	JP2K	0.726	0.824	0.867	0.662	0.608	**0.904**	0.872
	JPEG	0.718	0.841	0.838	0.630	0.548	**0.910**	0.884
	WN	**0.945**	0.937	0.927	0.907	0.924	0.937	**0.942**
	GBLUR	0.770	0.850	0.836	0.845	0.846	**0.911**	0.652
	ALL	0.700	0.909	**0.910**	0.837	0.805	0.897	0.893
LIVE 3D Phase II-Asymmetric	JP2K	0.724	0.805	0.854	0.722	0.619	0.789	**0.872**
	JPEG	0.714	0.805	**0.876**	0.636	0.678	0.696	0.686
	WN	0.882	0.952	0.937	0.929	0.941	0.924	**0.955**
	GBLUR	0.807	0.850	**0.888**	0.691	0.840	0.803	0.838
	ALL	**0.719**	0.661	0.642	0.611	0.697	0.524	**0.711**

The overall test results are shown in Table 2. Here, the final result obtained with a value of 0.2 was used. It can be seen from the relevant performance index that the image quality evaluation method proposed in this paper has a good generalization ability.

In Tables 3 and 4, two quality evaluation performance indicators PLCC and SROCC are used to compare the SIQE method proposed in this paper with the methods proposed by the predecessors, including three 2D full-reference image quality evaluation methods (SSIM [1], FSIM [2], and GMSD [3]) and three SIQE methods (Chen [4], Bensalma and Larabi [20] and Zhou [21]). It can be seen from Table 3 that the stereo image evaluation method proposed in this paper is better than the other NR image evaluation methods. As can be seen from Table 4, the method proposed in this paper is more advantageous for evaluating asymmetric stereo images. Since the proposed method does not require original undistorted images and human subjective scores, its

evaluation performance may be slightly lower than that of the FR evaluation method, but it achieves the desired effect of no-reference SIQE.

A good image quality evaluation method not only yields high performance, but also provides good computational efficiency. Predictive speed is also an important component of image quality evaluation performance indicators, because we need to consider the practicality of the algorithm. Here, we present the calculation time required for a pair of stereo images. In the LIVE 3D Phase I database, the prediction time for a pair of stereo images by using the NVIDIA 1080Ti GPU is approximately 0.034 s (more than 25 frames per second), while testing the pair in the LIVE 3D Phase II database, the prediction time for the stereo image is 0.039 s (more than 25 frames per second). In terms of speed, the method proposed in this paper can achieve real-time prediction (the real-time system run at the speed of 25 frames per second).

4. Conclusions

In this paper, we propose an effective SIQE method, which is different from the existing image evaluation method based on deep learning. First, the proposed method training dataset is a self-made distortion image dataset and the corresponding quality maps obtained by the high-performance FR image quality evaluation method as labels. Second, the main network framework used by the algorithm is the quality map generation network, which is used to train the distortion image dataset and quality map label to obtain an optimal network framework. The physiological function of the neurons ensured that the predicted results are highly consistent with the original quality map. Finally, the left view, the right view, and the cyclopean view of the stereo image are, respectively, input into the quality map generation network framework, and the corresponding quality maps are predicted, the three quality prediction scores are obtained using saliency weighting, and the three quality values are obtained. Weighted fusion is sequentially performed to obtain the final stereo image quality score. Experiment was performed on two 3D LIVE databases to verify the effectiveness of the proposed algorithm. The experimental results show that the algorithm can improve the accuracy of image quality prediction and prediction scores and has a good generalization ability.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant no. 61502429), the Zhejiang Provincial Natural Science Foundation of China (grant no. LY18F020012), the Zhejiang Open Foundation of the MOST Important Subjects, and the China Postdoctoral Science Foundation (grant no. 2015M581932).

References

[1] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[2] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: a feature similarity index for image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, 2011.

[3] W. Zhou, L. Yu, W. Qiu, Y. Zhou, and M. Wu, "Local gradient patterns (LGP): an effective local-statistical-feature extraction scheme for no-reference image quality assessment," *Information Sciences*, vol. 397-398, pp. 1–14, 2017.

[4] M.-J. Chen, C.-C. Su, D.-K. Kwon, L. K. Cormack, and A. C. Bovik, "Full-reference quality assessment of stereopairs accounting for rivalry," *Signal Processing: Image Communication*, vol. 28, no. 9, pp. 1143–1155, 2013.

[5] Y. Zhang, D. M., and Chandler, "3D-MAD: a full reference stereo image quality estimator based on binocular lightness and contrast perception," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3810–3825, 2015.

[6] W. Zhou, G. Jiang, M. Yu, F. Shao, and Z. Peng, "PMFS: a perceptual modulated feature similarity metric for stereoscopic image quality assessment," *IEEE Signal Processing Letters*, vol. 21, no. 8, pp. 1003–1006, 2014.

[7] X. Geng, L. Shen, K. Li, and P. An, "A stereoscopic image quality assessment model based on independent component analysis and binocular fusion property," *Signal Processing: Image Communication*, vol. 52, pp. 54–63, 2017.

[8] F. Gao, Y. Wang, P. Li, M. Tan, J. Yu, and Y. Zhu, "DeepSim: deep similarity for image quality assessment," *Neurocomputing*, vol. 257, pp. 104–114, 2017.

[9] W. Zhou, G. Jiang, M. Yu, F. Shao, and Z. Peng, "Reduced-reference stereoscopic image quality assessment based on view and disparity zero-watermarks," *Signal Processing: Image Communication*, vol. 29, no. 1, pp. 167–176, 2014.

[10] W. Zhou, L. Yu, Y. Zhou, W. Qiu, M.-W. Wu, and T. Luo, "Local and global feature learning for blind quality evaluation of screen content and natural scene images," *IEEE Transactions on Image Processing*, vol. 27, no. 5, pp. 2086–2095, 2018.

[11] W. Zhou, S. Zhang, T. Pan et al., "Blind 3D image quality assessment based on self-similarity of binocular features," *Neurocomputing*, vol. 224, pp. 128–134, 2017.

[12] W. Zhang, C. Qu, L. Ma, J. Guan, and R. Huang, "Learning structure of stereoscopic image for no-reference quality assessment with convolutional neural network," *Pattern Recognition*, vol. 59, pp. 176–187, 2016.

[13] G. Jiang, H. Xu, M. Yu, T. Luo, and Y. Zhang, "Stereoscopic image quality assessment by learning non-negative matrix factorization-based color visual characteristics and considering binocular interactions," *Journal of Visual Communication and Image Representation*, vol. 46, pp. 269–279, 2017.

[14] W. Hachicha, M. Kaaniche, A. Beghdadi, and F. A. Cheikh, "No-reference stereo image quality assessment based on joint wavelet decomposition and statistical models," *Signal Processing: Image Communication*, vol. 54, pp. 107–117, 2017.

[15] W. Zhou, W. Qiu, and M.-W. Wu, "Utilizing dictionary learning and machine learning for blind quality assessment of

3-D images," *IEEE Transactions on Broadcasting*, vol. 63, no. 2, pp. 404–415, 2017.

[16] J. Yang, P. An, J. Ma, K., L. Li, and Shen, "No-reference stereo image quality assessment by learning gradient dictionary-based color visual characteristics," in *Proceedings of the 2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1–5, IEEE, Florence, Italy, May 2018.

[17] J. Wu, J. Zeng, W. Dong, G. Shi, and W. Lin, "Blind image quality assessment with hierarchy: degradation from local structure to deep semantics," *Journal of Visual Communication and Image Representation*, vol. 58, pp. 353–362, 2019.

[18] T.-J. Liu, C.-T. Lin, H.-H. Liu, and S.-C. Pei, "Blind stereoscopic image quality assessment based on hierarchical learning," *IEEE Access*, vol. 7, pp. 8058–8069, 2019.

[19] Q. Jiang, F. Shao, W. Gao, Z. Chen, G. Jiang, and Y.-S. Ho, "Unified no-reference quality assessment of singly and multiply distorted stereoscopic images," *IEEE Transactions on Image Processing*, vol. 28, no. 4, pp. 1866–1881, 2019.

[20] R. Bensalma and M.-C. Larabi, "A perceptual metric for stereoscopic image quality assessment based on the binocular energy," *Multidimensional Systems and Signal Processing*, vol. 24, no. 2, pp. 281–316, 2013.

[21] W. Zhou, L. Yu, Y. Zhou, W. Qiu, M.-W. Wu, and T. Luo, "Blind quality estimator for 3D images based on binocular combination and extreme learning machine," *Pattern Recognition*, vol. 71, pp. 207–217, 2017.

[22] W. Zhou and L. Yu, "Binocular responses for no-reference 3D image quality assessment," *IEEE Transactions on Multimedia*, vol. 18, no. 6, pp. 1077–1084, 2016.

[23] K. Ma, Z. Duanmu, Q Wu et al., "Waterloo exploration database: new challenges for image quality assessment models," *IEEE Transactions on Image Processing*, vol. 26, no. 2, pp. 1004–1016, 2017.

[24] J. Wang, A. Rehman, K. Zeng, S. Wang, and Z. Wang, "Quality prediction of asymmetrically distorted stereoscopic 3D images," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3400–3414, 2015.

[25] L. Zhang, Y. Gu, and H. Li, "SDSP: a novel saliency detection method by combining simple priors," in *Proceedings of the 2013 IEEE International Conference on Image Processing ICIP*, pp. 171–175, IEEE, Melbourne, VIC, Australia, September 2013.

[26] A. K. Moorthy, C.-C. Su, A. Mittal, and A. C. Bovik, "Subjective evaluation of stereoscopic image quality," *Signal Processing: Image Communication*, vol. 28, no. 8, pp. 870–883, 2013.

[27] M.-J. Chen, L. K. Cormack, and A. C. Bovik, "No-reference quality assessment of natural stereopairs," *IEEE Transactions on Image Processing*, vol. 22, no. 9, pp. 3379–3391, 2013.

[28] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, 2014, http://arxiv.org/abs/412.6980.