Multimedia Tools and Applications
https://doi.org/10.1007/s11042-018-6524-1

A no-reference quality assessment for contrast-distorted image based on improved learning method

Yaojun Wu[1] · Yonghe Zhu[1] · Yang Yang[1] · Weiming Zhang[2] · Nenghai Yu[2]

Received: 31 October 2017 / Revised: 1 July 2018 / Accepted: 10 August 2018 /
Published online: 01 September 2018

Abstract
No-reference image quality assessment (NR-IQA), which aims to predict image quality without accessing to reference images, is a fundamental and challenging problem in the field of image processing. Nevertheless, there are few researches about contrast-distorted images and results of the existing NR-IQA methods which cannot be in accordance with the subjective assessment results further. Therefore, an effective NR-IQA method for contrast-distorted image is proposed in this paper. Firstly, the proposed method extracts five features of all images from the database based on the natural scene statistics (NSS) model. Then the curve fitting method is subsequently represented to calculate values of natural image features. Finally, an improved Support Vector Regression (SVR) learning method based on grid search is proposed to establish the mapping between image feature values and the quality score of a test image. Experiments proved that the proposed method is effective when compared with other related state-of-the-art image quality assessment (IQA) methods based on three standard databases.

Keywords No-reference image quality assessment · Learning method · Contrast-distorted · Support vector regression · Grid search

1 Introduction

With the rapid development of visual communication system, image quality assessment (IQA), which is employed to evaluate image quality score, is one of essential issues in

This work was supported in part by the Natural Science Foundation of China under Grants 61502007, 61572452, in part by the Natural Science Research Project of Anhui province under Grant 1608085MF125, in part by the NO.58 China Post-doctoral Science Foundation under Grant 2015M582015, in part by the Backbone Teacher Training Programof Anhui University, in part by the Doctoral Scientific Re-search Foundation of Anhui University under Grant J01001319.

✉ Yang Yang
sky_yang@ahu.edu.cn

Extended author information available on the last page of the article.

image processing. In the past, the subjective assessment, which gave a reliable score through Human Visual System (HVS), was a common method in IQA. But the subjective evaluation is invariably limited in time and economic costs, which makes it difficult to apply in practical occasion. Therefore, many objective image quality assessments are proposed to overcome the disadvantage of the subjective method [17], and the aim of these methods is to evaluate the quality score which is in accordance with the subjective assessment results.

To date, a series of IQA that work well in various image types have been adopted [32]. Existing IQA metrics can be divided into Full-reference (FR) [3, 27] Reduced-reference (RR) [24, 33] and No-reference (NR) [2, 26] three categories according to the availability of the reference images. Both of the FR-IQA and RR-IQA methods require information from undegraded source images, while NR-IQA method does not rely on reference information to obtain the objective score.

In recent years, with the rapid development of perceptual quality assessment, there is a leap in the growth of no-reference (NR) assessment method. The reason is that an original reference image does not exist in most cases. According to different assessment objects, NR assessment methods can usually be classified into NR pixels (NR-P) type and NR bitstream (NR-B) type [28]. In this paper, we focus on discussing NR-P assessment methods here. NR-P assessment methods are also divided into artifact measurement methods and features measurement methods [28]. For artifact measurement methods, Chen et al. have proposed a method which is said that can apply to any kind of blurriness [6]. Besides, Liang et al. have proposed that the quality of JPEG2000 coded images can be predicted by the combination of blur and ringing assessment methods [16]. Zhu et al. have proposed a sharpness method which is sensitive to the prevailing blur and noise in an image [37]. The mathematical formulation of the method is based on image gradients which is calculated by singular value decomposition (SVD). For feature measurement, the NSS-based quality estimator presented [19] performs in the spatial domain. Normalized luminance and empirical distribution are used to calculate the relevant features for building a spatial NSS model. An completely blind IQA model is derived in [20], which only makes use of measurable deviations from statistical regularities observed in natural images without training on human-rated distorted images or exposing distorted images. Ref [18] is developed based on local spatial and spectral entropy features on distorted images, which utilizes a support vector machine to train an image distortion and quality prediction engine. In the method of [4], the features of images, which are based on the HVS, are extracted to train a radial base function (RBF) network to evaluate the quality of images. This method adopts support vector machine to find the mapping between image features and the quality score. The authors of [9] go a step further, they evaluate the label of the image pair and calculate the result of the image pair to predict the objective score. In particular, to evaluate the quality of contrast-distorted images, the method proposed by [8] estimates the quality of contrast image through training contrast-distorted image databases. The method using the revealed free-energy-based brain theory and classical human visual system (HVS)-inspired features is proposed in [12]. Zhang et al. [36] is developed based on a multivariate Gaussian model of image patches from a collection of pristine natural images. Additionally, NSS features-based quality estimator also has been widely employed in stereoscopic images. For example, 2D- and 3D- statistical features are extracted to calculate the image quality [7]. The NSS features are used to train a support vector machine model to predict the quality of a tested stereo pair.

Generally, contrast has big implications for visual effect. Poor contrast will make the whole image gray. And contrast is very helpful for image definition, detail presentation and grayscale performance. Ref [13, 21, 38] have proposed that poor image quality often makes feature extraction, analysis, recognition, and quantitative measurements problematic and unreliable. Contrast sensitivity is one of the main indicators that describe the spatial characteristics of the human visual system. It is closely related to many human visual characteristics. For example, the commonly used contrast sensitivity function (CSF) describes the relationship between contrast sensitivity and spatial frequency [23, 29]. The human eye transfer function model based on contrast-sensitive visual experimental data is mainly divided into an exponential model or a Gaussian model of a low-pass filter, and a Barte model or a compound model of a band-pass filter. There are many reasons for producing contrast distortion, but they can be divided into two kinds. One is the interference from electromagnetic wave. Especially in the process of collecting medical images, interference from electromagnetic wave from mobile phone, medical apparatus and instruments could cause contrast distortion. Chang et al. have proposed that image contrast enhancement is important in medical applications [5]. There is the reason that visual examination of medical images is essential in the diagnosis of many diseases. The other is natural environment, for example, uneven illumination has a great impact on contrast. Both histogram stretch and histogram equalization can achieve contrast enhancement and are commonly employed for image enhancement because of their simplicity and comparatively better performance on almost all types of images. But human visual system can not distinguish which one is better, then image quality assessment for contrast distortion is employed to choose the one with less distortion.

There are large quantities of NR-IQA methods adopting the method of machine learning to evaluate image quality. The existing methods uniformly adopt feature extraction and appropriate evaluation model to obtain the objective score. Nevertheless, these methods still have two problems in the way of evaluation. First, the calculation of the image feature value is not accurate enough. Since different features have different dimensions, normalization is normally used to process the values of features. Generally, probability density function is designed to scale value of features [8], which may cause the error between the normalized feature data and true feature data. Second, the regression model is not optimized. Support vector regression (SVR), which is trained in default hyper parameters setting, is adopted to evaluate the objective quality score of the image in standard NR-IQA [9]. Although the default hyper parameters setting of SVR is invariably proposed to train the prediction model [8], it cannot reach optimal regression ability.

To overcome the aforementioned limitations, a no-reference quality assessment of contrast-distorted images based on improved learning method (NR-ICDIQA) is proposed in this paper. In this method, we first extract features of all images in SUN2012 database to represent overall structure of images. Curve fitting is subsequently designed to establish the function between statistical probability and image features. Finally, an improved learning method based on grid search is used to establish the mapping between image feature values and the quality score of an image. Through experiments conducted on the three standard databases, the proposed method has proven that it is effective and demonstrates the comparable performance to the state-of-the-art quality assessment.Our contribution in this paper has two aspects: 1. A curve fitting method is represented to calculate value of the natural image features. Curve fitting is more accurate than Gaussian fitting [8]. Besides, a rough fitting method leads to the case that the fitting function may not adequately reflect the probability distribution of the image. 2. The method of Grid search is proposed to find the best set of hyper parameters in machine learning algorithm.

The rest of the paper is organized as follows. In Section 2, the framework of the proposed NR-IQA method is detailed. Section 3 shows the experimental results and analyses. The Conclusion of proposed method is given in Section 4.

2 Proposed method

The state-of-the-art IQA methods pay little attention to the contrast-distorted image, which has a great impact on the human final subjective feelings of an image [23, 29]. And the performances of existing methods of NR-IQA are limited by the model of default regression and roughly fitting of image features. To overcome this problem, we propose a NR-ICDIQA framework to evaluate objective quality scores in this section. Inspired by the method of machine learning, natural scene statistics (NSS) model [25] is used to extract picture features firstly. Since this study majors in the assessment of contrast-distorted images, features that most likely to represent the difference about contrast change are selected. Then the SUN2012 database, which contains 16873 images, is used to count the distribution of the feature value. A curve fitting method is subsequently represented to calculate the value of the natural image features. Based on the result of calculation, the algorithm of machine learning is adopted to find the mapping relationship between image features and quality scores. To optimize the performance of machine learning, the grid search method is adopted in this method. The procedure of the proposed method is shown in Fig. 1. Details will be introduced in the following section.

2.1 Feature extraction and statistics

The moment features of images included mean, variance, skewness and kurtosis have been widely used and proved promising in many studies related to contrast-distorted images [10, 22, 34, 39]. Mean of image intensity reflects concentration trend of image pixel value and can be used to represent the overall brightness of images [10]. It demonstrates the variance of image intensity can be used to calculate the expected context-free contrast for optimal contrast-tone mapping in previous studies [34]. Skewness reflects the symmetry of an image pixel value and kurtosis is adopted to measure the deviation from the normal distribution. Recently, skewness and kurtosis have been proved that they are related to human perception of image surface [22, 39]. Entropy of image intensity is a statistical form of feature and represents the average amount of information on images.

Inspired by the wide use of moment features in image contrast researches [10, 22, 34, 39] and the application of entropy in image processing, moment and entropy features from images are extracted in our study.

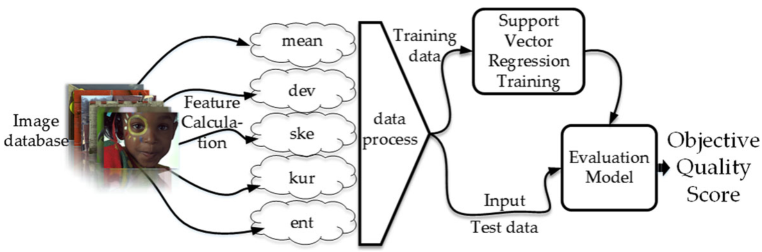


Fig. 1 The proceduce of NR-ICDIQA

Let I_i denotes pixel value of image i in image library I. The mean $mean(I)$, standard deviation $dev(I)$, skewness $ske(I)$, kurtosis $kur(I)$ and entropy $ent(I)$ of the image library I can be estimated as:

$$mean(I) = \frac{1}{n}\sum_{i=1}^{n} I_i \tag{1}$$

$$dev(I) = \left[\frac{1}{N-1}\left(\sum_{i=1}^{n} I_i^2 - mean(I)^2\right)\right]^2 \tag{2}$$

$$ske(I) = \frac{\frac{1}{n}\sum_{i=1}^{n}[I_i - mean(I)]^3}{dev(I)^3} \tag{3}$$

$$kur(I) = \frac{\frac{1}{n}\sum_{i=1}^{n}[I_i - mean(I)]^4}{dev(I)^4} - 3 \tag{4}$$

$$ent(I) = -\sum_{j} p_j(I) * \log_2 p_j(I) \tag{5}$$

where $p_j(I)$ denotes the frequency of intensity value j in image library I, I_i indicates the pixel value of the i-th image in library I.

In our method, we subsequently calculated sample mean, standard deviation, skewness, kurtosis and entropy of all images in the SUN2012 database [35], which includes 16873 images and covers a large variety of image contents. Then the frequency histogram of these features, which can roughly reflect distribution of these features in natural images, is calculated. The histogram of the frequency is used to curve fitting in Section 2.2.

2.2 Curve fitting of feature

In the proposed method, features of an image are adopted as the inputs of SVR to predict the quality score. In Section 2.1, the dimensions of features are different, which may reduce the performance of the training model in SVR. Therefore, the curve fitting of features is adopted in this section. Since SVR needs scaled data as the input to improve the performance of its evaluation, the data normalization of feature values becomes an essential step in most image assessment methods. In this section, the feature values of SUN2012 database are calculated. Frequency histograms of the mean, standard deviation, skewness, kurtosis and entropy are subsequently plotted to analyze the data structure of features after the calculation. Then the probability density functions of the image features can be calculated based on frequency histograms of the features to scale image feature values. In the literature [8], the frequency curve is roughly fitted by Gaussian probability density function in their methods, which leads to the fact that the fitting function may not adequately reflect the probability distribution of the image. To solve this problem, polynomial fitting is designed to fit the frequency curves of these features in this method. Nevertheless, the number of intervals in the group number n and the polynomial order of polynomial fitting m can affect the final prediction accuracy. Consequently, optimization model is used to solve the optimal combination mathematics model of m and n. The optimal model is as follows:

$$\begin{aligned} &min\ \left[\textstyle\sum_{i=1}^{n}(f_i - f_i')^2\right] \\ &s.t.\ \ \textstyle\sum_{i=1}^{N} f_i = 1 \\ &\qquad\qquad f_i' = a_0 + \textstyle\sum_{i=1}^{m} a_i m_i^i \end{aligned} \tag{6}$$

where f_i denotes the frequency of ith interval, m_i denotes the median of the interval and f_i' denotes the fitting value of the ith value. The range of the i is 1 to n. Since f_i is a frequency, the sum of all the f should be 1. The core of the model is to acquire the solution that gives the closest fitting value of the actual frequency. As shown in the (6), there are two constraints on this optimization model. The first constraint means that the sum of the probability densities is one, for the frequency plot is based on the statistics of SUN2012. And the second constraint means that the feature value is calculated by fitting polynomials which has been calculated by frequency histograms. Figure 2 shows histograms and the corresponding fitting curves of different features based on images in SUN2012 database. From which we can see that curve fitting method fits more compact than other fitting methods, such as gaussian fitting, inverse gaussian fitting and extreme value fitting.

2.3 Machine learning algorithm

Since no-reference image quality assessment is a quality metric without reference as the input, a great deal of images are needed to train the predict model. Therefore, machine learning algorithm, which needs abundant images to train model, is adopted to establish the model of NR-IQA. In our method, an improved learning algorithm is adopted to get an objective perceptual quality score according to the image features. The SVR learning method is designed to find the mapping function between feature set and quality score firstly, for this is a regression problem. There are multiple SVR methods in machine learning algorithm, such as v-SVR and epsilon-SVR [1]. Since epsilon-SVR has a better image assessment ability in [30], it is selected in this method. The procedure of epsilon-SVR is shown as follows.

Firstly, let set x_i denote the set of the ith image feature, the x_i is given by

$$x_i = \{mean(I), dev(I), ske(I), kur(I), ent(I)\} \tag{7}$$

The definition of $mean(I)$, $dev(I)$, $ske(I)$, $kur(I)$, $ent(I)$ is given in Section 2.1. Then let y_i denote the mean opinion score given by subjective experiment. SVR is designed to find a function $y_i' = f(x_i)$ that gives reasonable objective quality score. The error between y_i and y_i' is in an acceptable threshold through SVR algorithm. The function is determined as follows:

$$y_i' = f(x_i) = w^T \psi(x_i) + \gamma \tag{8}$$

In this formula, $\psi(x_i)$ is a kernel function of the feature vector, w is a weighting vector and γ means the bias term. In this method, RBF is adopted as kernel, for it has outstanding performances in vast majority of occasions. The RBF kernel is determined as follows:

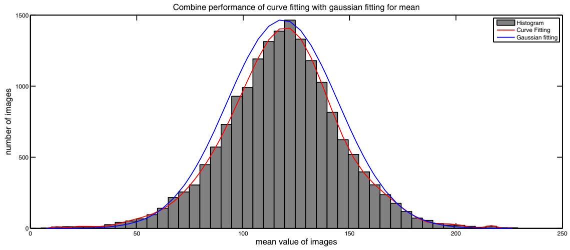
$$\psi(x_i) = e^{-g*|x_i - v|^2} \tag{9}$$

where g is the width parameter of the Gaussian function, v is the center of the RBF, x is the input training data. The image database included CID2013, TID2013 and CSIQ will be used in the training process. The feature vector of the database is the input as training data in SVR, while the subjective score is used as training label. Then the SVR system is employed to predict ψ w and γ in (8).

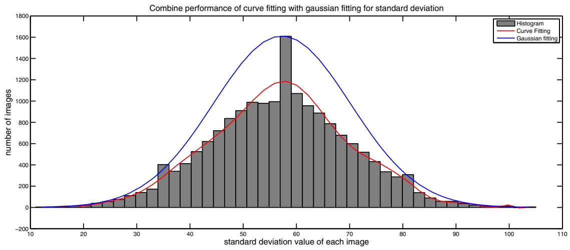
2.4 Grid search method

During the model training process, hyper parameter optimization is an essential problem of choosing a set of hyper parameters, for it can optimize the algorithm performance. Nevertheless, there is no concern about the hyper parameter optimization in state-of-the-art

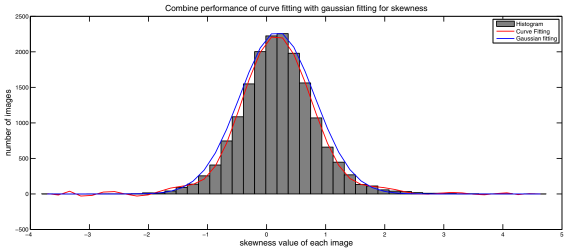
Fig. 2 Histograms and the corresponding tting curves of different features based on images in SUN2012 database **a** Mean **b** Standard Deviation **c** Skewness **d** Kurtosis **e** Entropy



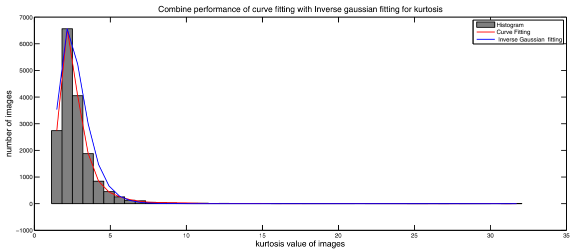
(a)



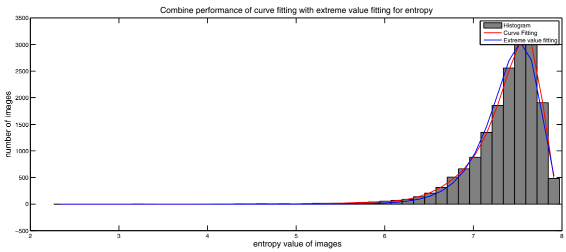
(b)



(c)



(d)



(e)

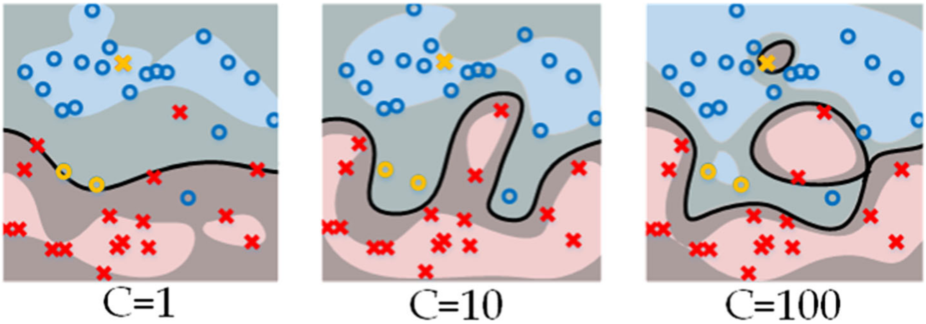


Fig. 3 The influence of hyperparameter C

methods. In the literature [8], default setting is always used in SVR application, for it has a good performance in most occasions. However, the default setting cannot reach the optimized accuracy to evaluate the image quality. So we need to consider the optimized hyper parameters set to improve the models' performance. Consequently, the grid search method is proposed to find the best set of hyper parameters in machine learning algorithm.

Since our learning method is a typical SVR algorithm equipped with RBF kernel, there are two hyper parameters that need to be tuned for better performance on IQA: regularization constant C and kernel function hyper parameter. Since there is an optimal value in the selection of C and g, analyses of the hyper parameters are made. The result is shown in Figs. 3 and 4.

Since the effect of C in SVR is difficult to explain in figure, support vector classification (SVC), which is similar to SVR, is adopted to explain the effect on C. In Fig. 3, there are three types of data. The blue dot belongs to one class and the red fork belongs to the other class, while the yellow dot and fork belongs to the noise. The SVC is employed to define the boundary of the data. In fact, the hyper parameter C is a tradeoff of the appropriate relaxation of the margin size and the tolerance of some regression error. Nevertheless, the increased C also reduces the tolerance of noise and other disturbances, which may cause overfitting in training.

Figure 4 illustrates that if the value of g is high, it will make the function distribution high and thin, which causes the kernel function to act only near the support vector sample.

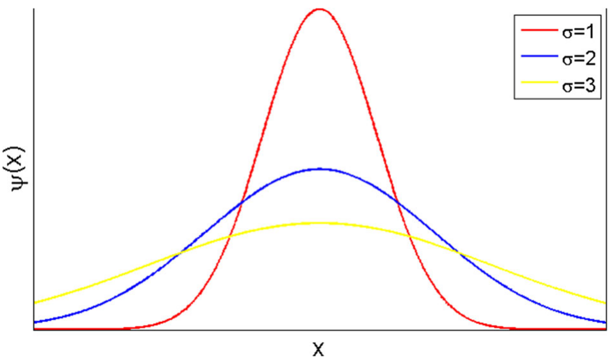


Fig. 4 The influence of hyperparameter g

If the value of g is small, it will cause the smooth affect, which will also affect the test set and training set accuracy. Consequently, there is also a tradeoff in the selection of hyper parameter g.

Both hyper parameters are continuous, so to perform grid search, we need to select a finite set of reasonable values for each hyper parameter. The common finite set of each hyper parameter is shown as follows:

$$\begin{aligned} &C \in \{2^{-a}, 2^{-a+b}, 2^{-a+2b}, \cdots, 2^{-a+(n-1)b}, 2^{a}\} \\ &g \in \{2^{-a}, 2^{-a+b}, 2^{-a+2b}, \cdots, 2^{-a+(n-1)b}, 2^{a}\} \\ &a = nb > 0 \end{aligned} \tag{10}$$

where a is the threshold of the hyper parameters and b denotes the span value of the grid.

In the grid search method, we then train SVR model with each pair (C, g) in the Cartesian product of these two sets and evaluate their performance to find the best regression pair. The optimal pair value of hyper parameters will be obtained in experiments.

In this paper, we choose to optimize C and g only because these two parameters have the most significant effect on the performance of the model. In fact, we have also tried to optimize other parameters, such as polynomial kernel function parameter d and termination condition e, but their performance improvement is not obvious, and the increase of parameters will cause the exponential increase of the algorithm complexity. Therefore, we finally choose only these two parameters to increase the credibility of experimental results.

3 Experimental evaluation

In this study, four databases are used to conduct a series of experiments, i.e. the SUN2012 [35], CID2013 [11], TID2013 [25] and CSIQ [15] database. Details about four databases are shown as below.

The SUN2012 database covers a large variety of image content, which has been introduced in Section 2. There are 400 contrast-distorted images in the CID2013 which are generated from 15 original images [11]. Twenty-two inexperienced viewers were involved in the subjective experiments to provide their overall perception of quality on a continuous quality scale from 1 to 5 for each contrast-distorted image, then the mean opinion score (MOS) of each image is computed. TID2013 is composed of 1700 distorted images of 25 original natural images. It includes 200 contrast-distorted images whose MOS is in the scale of 0 to 9. As the result, larger MOS relates to a better visual quality [25]. CSIQ is composed of different types of distortions of 35 reference images and contains 116 contrast-distorted images. 35 subjects are involved in the experiments to provide 5000 subjective ratings for distorted images. Final subjective results are the Difference of MOS (DMOS) between the reference and distorted images [15]. In this study, we select all contrast-distorted images of the CID2013, TID2013 and CSIQ databases to test the performance of the proposed NR-ICDIQA method.

Three experiments of this paper, which are conducted on four commonly used databases, are presented in Part 3.1, 3.2 and 3.3, respectively. Firstly, through the experiment in Part 3.1, we find out the best polynomial parameters over three databases and show the performance of the curve fitting varies with the parameters included in it. After that, the experiment in Part 3.2 is carried out to exhibit the effectiveness of grid search in our NR-ICDIQA model and find out the best hyper parameters which influence the performance of prediction. Finally, the best hyper parameters, which are found out in the first two experiments, are employed in the experiment in Part 3.3 to achieve the best performance of the proposed NR-ICDIQA method and compare it with other related methods, including PSNR, SSIM [31], MAD [14], NIQE [20], NR-CDIQA [8], NFERM [12], and IL-NIQE [36] through three image databases.

Since the subjective score is an important standard of IQA, in all experiments, three commonly used performance metrics are employed to compare the subjective and objective quality evaluation results' Pearson Linear Correlation Coefficient (PLCC), Spearman Rank-Order Correlation Coefficient (SRCC) and Root Mean-Square Error (RMSE). High values of PLCC and SRCC is expected in our study, while RMSE is low.

PLCC is used to evaluate the linear correlation of two sets of data. It can be estimated as:

$$\rho_p = \frac{Cov(\Upsilon, \Upsilon')}{\sqrt{Var(\Upsilon)Var(\Upsilon')}} \tag{11}$$

where ρ_p denotes the linear correlation coefficient of Υ and Υ' which are two sets of data,$Cov(\Upsilon, \Upsilon')$ indicates the covariance Υ and Υ' and , whose variances are denoted as $Var(\Upsilon)$ and $Var(\Upsilon')$,respectively.

SRCC is adopted to evaluate the rank-order correlation of two sets of data and can be calculated as:

$$\rho_s = 1 - \frac{6\sum_{i=1}^{N}(r_i - r_i')}{N(N^2 - 1)} \tag{12}$$

where ρ_s indicates the rank-order correlation coefficient of Υ and $\Upsilon'.r_i$ and r_i' denote the rank-order of the i-th data in the set of Υ and Υ', respectively. The number of data in the set of Υ and Υ' can be indicated as N.

RMSE is used to calculate the deviation of two sets of data and can be estimated as:

$$\rho_r = \sqrt{\frac{\sum_{i=1}^{N}(y_i - y_i')^2}{N}} \tag{13}$$

where ρ_r indicates the deviation of Υ and Υ' which are two sets of data. y_i and y_i' denote the i-th data of the set of Υ and Υ', respectively. The number of data in set Υ or Υ' can be indicated as N.

In addition, tenfold leave-one-out cross-validation is designed to test proposed metric in all experiments. We randomly divide each image database into 10 subsets. Specifically, nine of the subsets are used for training and the rest of images are used for testing at each time.This procedure is repeated 10 times and the average result of evaluation for each database are calculated.

3.1 Variation with algorithm parameters

In this subsection, we focus on the impact of choosing different algorithm's parameters. 1) group number n of images in the SUN2012 database, 2) the polynomial order m of curve fitting. We calculate the sample mean, standard deviation, skewness, kurtosis and entropy of all images from SUN2012 database, according to parameters calculated, the images contained in the SUN2012 database are divided into n, $n = 30, 35, \ldots, 70$, groups to calculate the frequency histograms of these features. Then we calculate m, $m \in \{15, 16, 17, 18, 19, 20\}$, order curve fitting base on these histograms to scale the features of CID2013, TID2013 and CSIQ. It has been proved that the value of n at the range from 30 to 75 or the value of m at the range from 15 to 20 has a big bad impact on the effectiveness of the algorithm through our pretreatment experiment. After that, the SVR learning method is adopted to find the mapping function between feature values set and quality score. Finally, to compare the subjective and objective quality evaluation, three commonly used

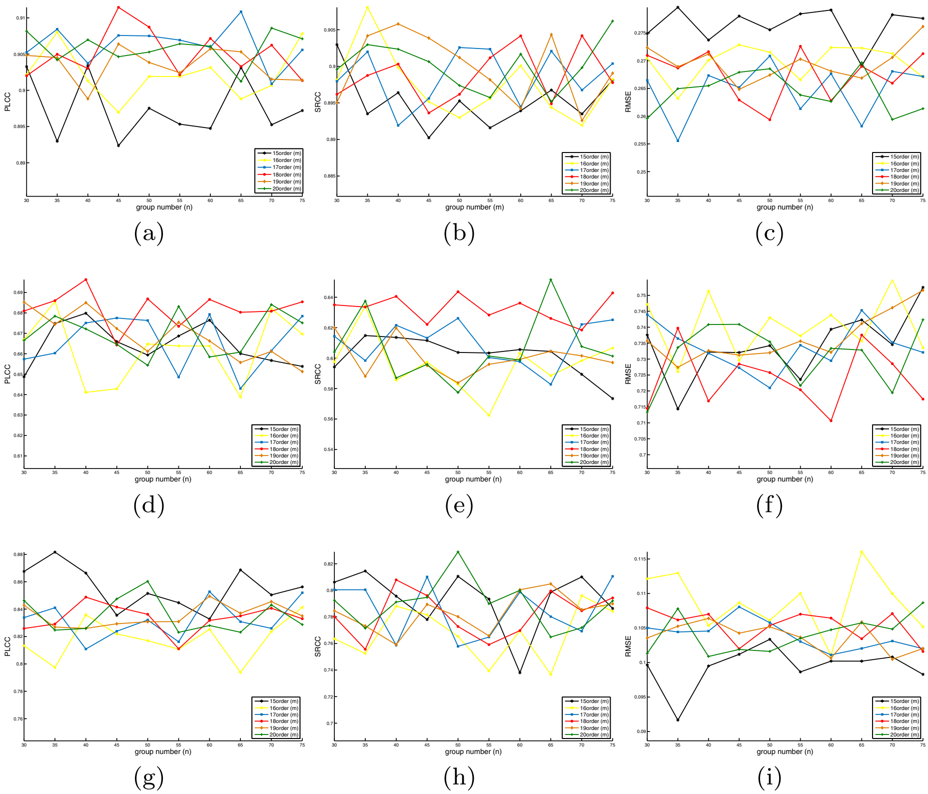


Fig. 5 The performance on CID2013, TID2013 and CSIQ varies with the settings of m and n **a b c** CID2013 **d e f** TID2013 **g h i** CSIQ

metrics are employed in this experiment: PLCC, SRCC and RMSE. Specifically, tenfold leave-one-out cross validation is designed in this experiment to calculate the average value of each metric. The performances of CID2013, TID2013 and CSIQ varied with the settings of m and n are shown in Fig. 5.

In our study, high values of PLCC and SRCC represent for great performance of the proposed NR-ICDIQA method, while the low value of RMSE means unsatisfactory result. Figure 5a b, c, d, e, f, g, h, i show how the performances on CID2013, TID2013 and CSIQ varied with the combinations of m and n, respectively.

In Fig. 5, for each highest value of PLCC or SRCC and lowest value of RMSE, there is a corresponding combination of m and n. Therefore, in accordance with Fig. 5, the best polynomial parameters m and n over three databases to achieve the best values of PLCC, SRCC and RMSE, which are tabulated in Table 1. PLCC, SRCC and RMSE are adopted to measure the linear correlation, rank-order correlation and deviation of two variables, respectively. For different evaluation metrics and databases, the proposed NR-ICDIQA method can achieve the best performance by selecting the most appropriate polynomial parameters according to the Table 1. In addition, the polynomial parameters m and n, which are tabulated in Table 1, are designed in the last experiment to achieve the best performance in different evaluation metrics for different databases. Details are presented in Part 3.3.

Table 1 Best polynomial parameters over three databases to achieve best value of three evaluation metrics

Parameters	CID2013			TID2013			CSIQ		
	PLCC	SRCC	RMSE	PLCC	SRCC	RMSE	PLCC	SRCC	RMSE
Order m	18	16	17	18	18	18	15	15	15
Group n	45	35	35	40	75	60	35	35	35

3.2 Grid search

Previous discussion in Section 2.2 has implied that the performance of prediction is determined mainly by two parameters-parameter C and parameter g. In this section, we determine on the different performance of parameter C and g. For each of the CID2013, TID2013 and CSIQ database, we all use the grid search method to find their different optimal parameter value. Contour chart is used to show the actual effects of parameters' changes. By experience, the search ranges a is 8 and the steps of b is 0.8. To evaluate the performance of a certain set selection, PLCC, SRCC and RMSE are measured by cross validation on the training set as a metric to decide the best performance set. The actual effect on image database is shown in Fig. 6.

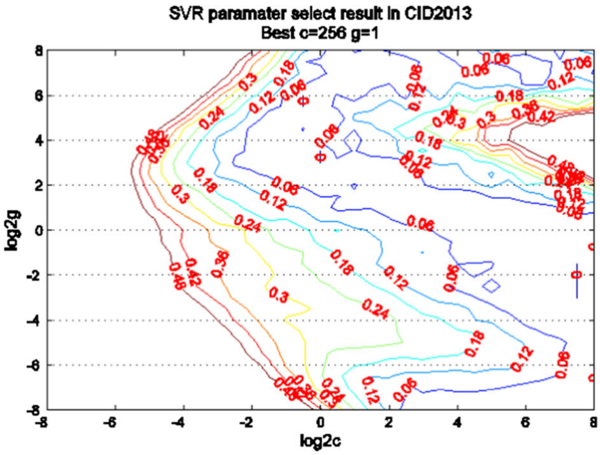
In Fig. 6, the abscissa is the logarithm of the parameter C, the ordinate is the logarithm of the parameter g, the error between the subjective score and the objective score is calculated and displayed in the map. And the coordinates corresponding to the same predicted error value are connected by using the contour map. Consequently, the best parameter selection should be at the bottom of the contour map. From the Fig. 6 we can see that the settings of different parameters on prediction model has great difference. This is mainly caused by two reasons: 1) the parameter C is penalty coefficient, the training model has proven to cause overfitting if the value of c is too small. And if the value of C is too large, there will be a huge gap between fitting result and actual result, 2) the parameter g is the coefficient of RBF, the value of gamma directly affects the ability to find an optimal hyperplane that separates multiple classes of data.

From the Fig. 6 we can also see that the optimal parameter values are different when the image database is changed. The optimal parameter C is 256 and parameter g is 1 in CID2013 database, while the best value is 45.25 and 2.82 in TID2013 database. In CSIQ database, the optimal parameter values are 2 and 1.41. C and g represent penalty coefficient and kernel function parameter in SVM, respectively. Both C and g depend on the number and quality of training samples. Different training samples may lead to different C and g. CID2013, TID2013 and CSIQ database include 400, 200 and 116 contrast-distorted images, respectively. The phenomenon of large changes in C and g of CSIQ database is due to the fact that the number of contrast-distorted images in CSIQ database is much less than that in CID2013 and TID2013 database.

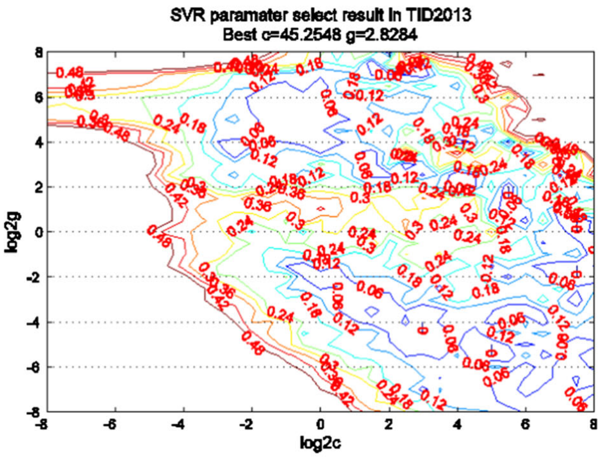
3.3 Comparison between Proposed Method and Other Methods

In this experiment, three databases are used to test the performance of the proposed NR-ICDIQA method and compare it with PSNR, SSIM [31], MAD [14], NIQE [20], NR-CDIQA [8], NFERM [12] and IL-NIQE [36]: CID2013 [11], TID2013 [25] and CSIQ [15]. PSNR, SSIM and MAD are widely used FR-IQA, while NIQE, NFERM, IL-NIQE and

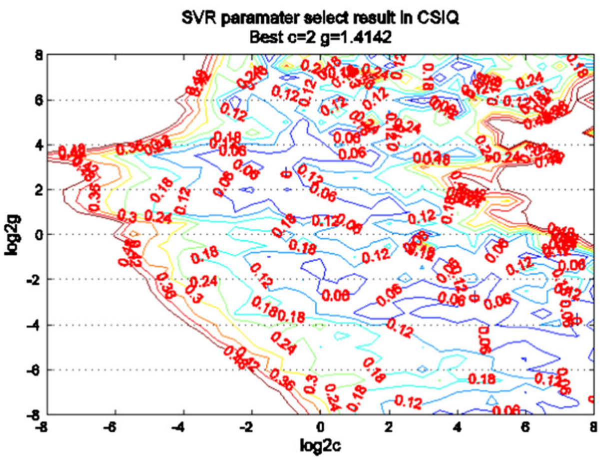
Fig. 6 Hyperparameter selects result in different image database



(a)



(b)



(c)

NR-CDIQA are widely used NR-IQA. Specially, NR-CDIQA is recent quality metric for contrast-distorted images.

Previous experiment in Part 3.1 has implied that the performance of the proposed NR-ICDIQA method is influenced by group number n of images in the SUN2012 database and the polynomial order m of curve fitting. Therefore, in this experiment, the best polynomial parameters m and n over three databases, which are shown in Table 1, are employed to achieve great performance of the proposed NR-ICDIQA method in different evaluation metrics. Then the curve fitting, which is employed to calculate five features of CID2013, TID2013 and CSIQ database, is calculated according to m and n. In addition, the experiment in Part 3.2 demonstrates that the performance of prediction is determined mainly by two parameters-parameter C and parameter g. Inspire by that, the optimal hyper parameters, which has been discussed in Part 3.2, is adopted in this experiment. After that, the SVR learning method is adopted to find the mapping function between feature set and quality score. To compare the subjective and objective quality evaluation, PLCC, SRCC and RMSE are also employed in this experiment. Similar to previous experiments, each image database is divided into 10 subsets. Tenfold leave-one-out cross validation is also used to test the proposed method.

The performances of the proposed and other related IQA methods are tabulated in Table 2. In addition, the best image quality assessment for each index is highlighted in boldface. High values of PLCC and SRCC, while low value of RMSE, is expected. Our method achieves better performance than other related method tabulated in Table 2 on CID2013 and TID2013 database, while it has poor performance on CSIQ database. Since our method is designed for contrast-distorted images, it’s normal that the proposed method achieves great performance on CID2013, a database designed for contrast distortion. The effect of SVR model is influenced by the number of training samples. As mentioned in Section 3, TID2013 database has enough contrast-distorted images for training. Therefore, the proposed NR-ICDIQA method also has great performance on TID2013 database. In fact, the hyper parameter C of SVR is a tradeoff of the appropriate relaxation of the margin size and the tolerance of some regression error. C is too big or too small may trigger overfitting or underfitting, which both lead to poor performance of the proposed method. Penalty coefficient C in CSIQ database is much smaller than that in CID2013 and TID2013 database, which causes underfitting and then leads to poor performance. As shown in Table 2, over CSIQ database, the performance of the proposed method ranks only second to MAD and

Table 2 Performance evaluation based on three image databases

Models		CID2013			TID2013			CSIQ		
		PLCC	SRCC	RMSE	PLCC	SRCC	RMSE	PLCC	SRCC	RMSE
FR-IQA	PSNR	0.6504	0.6649	0.4733	0.5071	0.5434	0.8453	0.8987	0.8621	0.0739
	SSIM	0.8072	0.8132	0.3678	0.6870	0.5510	0.7127	0.7437	0.7397	0.1126
	MAD	0.8151	0.8079	0.3610	0.6028	0.5515	0.8738	**0.9320**	**0.9207**	**0.0611**
NR-IQA	NIQE	0.4778	0.3824	0.8193	0.0234	0.0661	1.7063	0.3572	0.2292	1.0498
	NFERM	0.2341	0.2083	0.2221	0.0150	0.1344	0.6296	0.4982	0.4281	0.4968
	IL-NIQE	0.5386	0.5276	0.4032	0.0298	0.0644	0.5066	0.5324	0.5182	0.3536
	NR-CDIQA	0.8491	0.8550	0.3364	0.5263	0.4904	0.8405	0.8508	0.8044	0.0823
	Proposed	**0.9129**	**0.9081**	**0.2555**	**0.6963**	**0.6429**	**0.7107**	0.8817	0.8145	0.0917

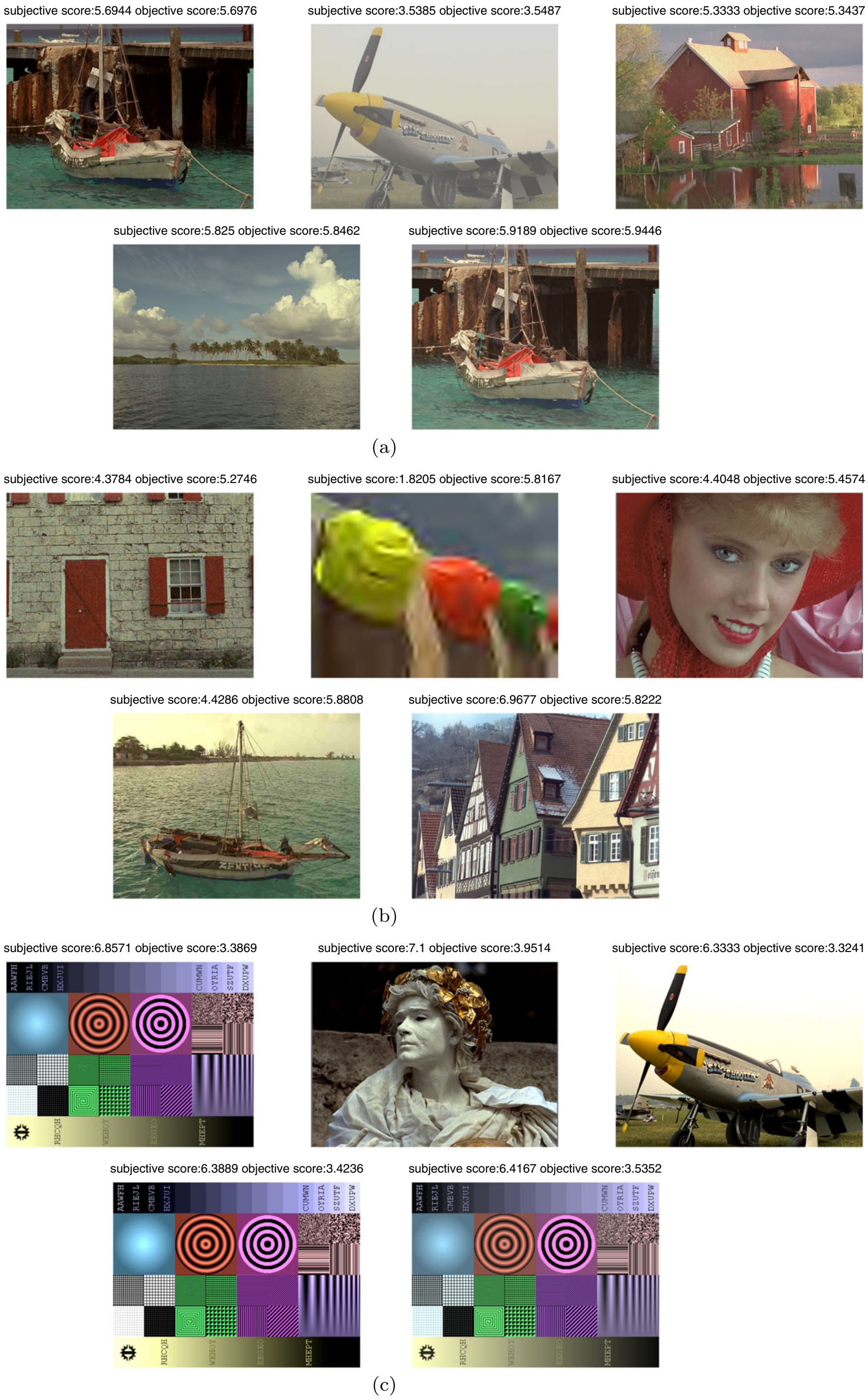


Fig. 7 The result of where model work well and where not

PSNR. The reason is that MAD and PSNR are FR-IQA, while our method is NR-IQA. Specially, FR-IQA is relied on original images and not influenced by the scale of image databases.

3.4 Strength and weakness of our model

In order to analyze the strength and Weakness of the proposed method further, we used cross-validation to test the objective quality scores of each image in the TID2013 image database through the experiment. The reason why we use TID2013 is that in addition to having a lot of contrast-distorted images, there are other types of distorted images in the image library itself, which can allow us to do better contrast experiments. By comparing the objective quality score obtained by training under the optimal parameters with the subjective quality, we obtain the most consistent images of the objective score and subjective score and the most inconsistent images. They are shown in Fig. 7a and c. We also randomly selected other types of distorted pictures to make predictions in our model. The results are shown in Fig. 7b.

In this experiment, two control experiments were performed by using group (a) as a control group, which fully analyzed where the model worked well and where it did not work well. First of all, we find that our model is better for assessing quality of contrast image, for it is designed for contrast-distortion images. Comparing Fig. 7a and b, we find that the subjective and objective scores of Fig. 7a are more consistent because images of Fig. 7a are all images with contrast distortion, and images of Fig. 7b are other distortion types. This paper does have a better evaluation performance on the quality of contrast images than other types of distortion.

Then we analyze the model's work on contrast-distorted images. From Fig. 7a and c we can see that our model also don't work well under some contrast distortion. There are two main reasons for this problem. The first reason is that some of the images we tested are not natural images. Since our model first performs feature extraction based on natural scene statistics, for Figs. 1, 4 and 5 in Fig. 7c, since it is not a natural image itself, Its predictions have been greatly biased. The second reason is that this kind of images have less contrast distortion, it is closer to the original image, which means it is much different from the general contrast images. So our model does not have particularly good performance in this regard, such as Fig. 7c.

4 Conclusion

In this paper, an improved NR-IQA of contrast-distorted image based on learning method is proposed by improving the accuracy of the image feature value and optimizing the regression model of learning method. Experiments on three public databases demonstrate that the proposed method has better performance for contrast-distorted image when compared with other related NR-IQA methods. Specially, for CID2013 and TID2013, the proposed method is superior to classical and state-of-the-art FR-IQA and NR-IQA methods. In CSIQ database, the proposed method ranks second to FR-IQA method due to lack original image and exist a small scale of image databases. In addition, we can optimize learning method and improve the accuracy of hyper parameters of SVR in future research based on this paper. Further research is also needed to explore better curve fitting and extend to other types of distortion.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

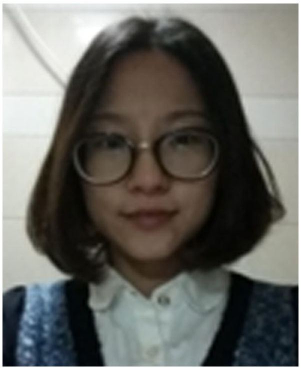
References

1. [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm/
2. Amiri SA, Hassanpour H, Marouzi OR (2017) No-reference image quality assessment based on localized discrete cosine transform for JPEG compressed images. Multimedia Tools and Applications, pp 1–17
3. Attar A, Shahbahrami A, Rad RM (2016) Image quality assessment using edge based features. Multimed Tools Appl 75(12):7407–7422
4. Babu RV, Suresh S, Perkis A (2007) No-reference JPEG-image quality assessment using GAP-RBF. Signal Process 87(6):1493–1503
5. Chang DC, Wu WR (1998) Image contrast enhancement based on a histogram transformation of local standard deviation. IEEE Trans Med Imaging 17(4):518–531
6. Chen C, Bloom JA (2011) A universal reference-free blurriness measure. Proc SPIE - The Int Soc Opt Eng 7867(13):1216–1217
7. Chen MJ, Cormack LK, Bovik AC (2013) No-Reference Quality assessment of natural stereopairs. IEEE Trans Image Process 22(9):3379–3391
8. Fang Y, Ma K, Wang Z, Lin W, Fang Z, Zhai G (2014) No-Reference Quality assessment of Contrast-Distorted images based on natural scene statistics. IEEE Trans Signal Process Lett 22(7):838–842
9. Gao F, Tao D, Gao X, Li X (2015) Learning to rank for blind image quality assessment. IEEE Trans Neural Netw Learn Syst 26(10):2275–2290
10. Gonzalez RC, Woods RE (2008) Image enhancement in the spatial domain, in digital image processing, 3rd edn. Addison-Wesley, Reading
11. Gu K, Zhai G, Yang X et al (2013) Subjective and objective quality assessment for images with contrast change. In: 2013 20th IEEE international conference on image processing (ICIP). IEEE, pp 383–387
12. Gu K, Zhai G, Yang X, Zhang W (2015) Using free energy principle for blind image quality assessment. IEEE Trans Multimedia 17(1):50–63
13. Ibrahim H, Kong NSP (2007) Brightness preserving dynamic histogram equalization for image contrast enhancement. IEEE Trans Consum Electron 53(4):1752–1758
14. Larson EC, Chandler DM (2010) Most apparent distortion: full-reference image quality assessment and the role of strategy[J]. J Electron Imag 19(1):011006.01–011006.24
15. Larson EC, Chandler DM Categorical image quality (CSIQ) database [Online]. Available: http://vision.okstate.edu/csiq
16. Liang L, Wang S, Chen J, Chen J, Ma S, Zhao D, Gao W (2010) No-reference perceptual image quality metric using gradient profiles for JPEG2000. Signal Process Image Commun 25(7):502–516
17. Lin W, Kuo CJ (2011) Perceptual visual quality metrics: A survey. J Vis Commun Image Represent 22(4):297–312
18. Liu L, Liu B, Huang H, Bovik AC (2014) No-reference image quality assessment based on spatial and spectral entropies. Signal Process Image Commun 29(8):856–863
19. Mittal A, Moorthy AK, Bovik AC (2012) No-reference image quality assessment in the spatial domain. IEEE Trans Image Process A: Publ IEEE Signal Process Soc 21(12):4695
20. Mittal A, Soundararajan R, Bovik AC (2013) Making a completely blind image quality analyzer. IEEE Signal Process Lett 20(3):209–212
21. Morrow WM, Paranjape RB, Rangayyan RM, Desautels JL (1992) Region-based contrast enhancement of mammograms. IEEE Trans Med Imaging 11(3):392
22. Motoyoshi I, Nishida S, Sharan L, Adelson EH (2007) Image statistics and the perception of surface qualities. Nature 447:206–209
23. Nadenau MJ, Reichel J, Kunt M (2003) Wavelet-based color image compression: exploiting the contrast sensitivity function. IEEE Trans Image Process A: Publ IEEE Signal Process Soc 12(1):58
24. Omari M, Hassouni ME, Abdelouahad AA, Cherifi H (2015) A statistical reduced-reference method for color image quality assessment. Multimed Tools Appl 74(19):1–17

25. Ponomarenko N, Ieremeiev O, Lukin V, Egiazarian K, Jin L, Astola J, Vozel B, Chehdi K, Carli M, Battisti F, Jay Kuo C-C (2013) Color image database TID2013: peculiarities, and preliminary results. In: Proceedings of the 4th European workshop on visual information processing, pp 106–111
26. Rezaie F, Helfroush MS, Danyali H (2017) No-reference image quality assessment using local binary pattern in the wavelet domain. Multimedia Tools and Applications, pp 1–13
27. Oszust M (2017) Image quality assessment with lasso regression and pairwise score differences. Multimedia Tools and Applications, pp 1–16
28. Shahid M, Rossholm A, Lovstrom B, Zepernick HJ (2014) No-reference image and video quality assessment: a classification and review of recent approaches. Eurasip J Image Video Process 2014(1): 1–32
29. Shi J, Yao J, Yu H, Yun L (2007) Measurement of luminance contrast sensitivity function of human visual system on cathode ray tube display. Acta Optica Sinica 27(4):744–748
30. Tolambiya A, Kalra PK, Zhai Contrast sensitive epsilon-SVR and its application in image compression. In: IEEE international conference on systems, man and cybernetics, vol 66. IEEE, pp 359–364
31. Wang Z, Bovik A, Sheikh H, Simoncelli E (2004) Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 13(4):600–612
32. Wang Z, Bovik AC (2006) Modern image quality assessment, video and multimedia processing. Morgan and Claypool, San Mateo
33. Wu J, Lin W, Shi G, Liu A (2013) Reduced-reference image quality assessment with visual information delity. IEEE Trans Multimedia 15(7):1700–1705
34. Wu X (2011) A linear programming approach for optimal contrast-tone mapping. IEEE Trans Image Process 20(5):1262–1272
35. Xiao J, Hays J, Ehinger K, Oliva A, Torralba A (2010) SUN database: Large-scale scene recognition from abbey to zoo. IEEE Conf Comput Vis Pattern Recognit 23(3):3485–3492
36. Zhang L, Zhang L, Bovik AC (2015) A feature-enriched completely blind image quality evaluator. IEEE Trans Image Process 24(8):2579–2591
37. Zhu X, Milanfar P (2009) A no-reference sharpness metric sensitive to blur and noise. In: International workshop on quality of multimedia experience. IEEE, pp 64–69
38. Zong X, Laine AF, Geiser EA (1998) Speckle reduction and contrast enhancement of echocardiograms via multiscale nonlinear processing. IEEE Trans Med Imaging 17(4):532–40
39. Zoran D, Weiss Y (2009) Scalein variance and noise in natural images. In: Proceedings of the IEEE international conference on computer vision, vol 30, no 2, pp 2209–2216

Yaojun Wu is a student with the school of Anhui University. He will pursue a graduate degree at 2018 in University of Science and Technology of China. His research interests include image quality assessment and deep learning.

Yonghe Zhu is a student with the school of Anhui University. He will pursue a graduate degree at 2018 in University of Science and Technology of China. His research interests include informantion hiding and image quality evaluation.



Yang Yang received her M.S. degree and PH.D. degree in 2007 and 2013 respectively from Anhui University and University of Science and Technology of China. She is an associate professor with the Anhui University, and she is also a postdoctoral researcher with the University of Science and Technology of China. Her research interests include reversible information hiding and image quality assessment.

Weiming Zhang received his M.S. degree and PH.D. degree in 2002 and 2005 respectively from the Zhengzhou Information Science and Technology Institute, Zhengzhou, China. Currently, he is a professor with the School of Information Science and Technology, University of Science and Technology of China. His research interests include information hiding and cryptography.

Nenghai Yu received his B.S. degree in 1987 from Nanjing University of Posts and Telecommunications, M.E. degree in 1992 from Tsinghua University and Ph.D. degree in 2004 from University of Science and Technology of China, where he is currently a professor. His research interests include multimedia security, multimedia information retrieval, video processing and information hiding.

Affiliations

Yaojun Wu[1] · Yonghe Zhu[1] · Yang Yang[1] · Weiming Zhang[2] · Nenghai Yu[2]

Yaojun Wu
yaojun_wu@yeah.net

Yonghe Zhu
zyh_ahu@126.com

Weiming Zhang
zhangwm@ustc.edu.cn

Nenghai Yu
ynh@ustc.edu.cn

[1] School of Electronics and Information Engineering, Anhui University, Hefei 230601, China

[2] School of Information Science and Technology, University of Science and Technology of China, Hefei 230027, China