No-Reference Quality Assessment of Super-Resolution Reconstructed Images by Incorporating Domain Knowledge

Guangcheng Wang[1], Feng Zhu[2], Zhaolin Lu[1], Xiaoping Yuan[1,*]

[1]School of Information and Control Engineering
China University of Mining and Technology
Xuzhou 221116, China

[2]Airforce Logistic College
Xuzhou 221005, China

*Corresponding author: xpyuan@cumt.edu.cn

ABSTRACT. *Super resolution (SR) has been extensively studied these years due to its wide applications in machine vision, medical imaging and remote sensing, etc. With a great number of SR image reconstruction algorithms proposed, an accompanying task is how to evaluate the quality of SR reconstructed images. Although a number of image quality metrics have been reported, they are not specifically designed for image super resolution, so the are usually limited for this task. Motivated by these, this paper presents a no-reference quality metric for SR reconstructed images by measuring structure degradations and SR-relevant domain distortions. By incorporating domain knowledge, the proposed metric is more effective than the state-of-the-art models for the quality assessment of SR reconstructed images. Experimental results based on three subjectively-rated SR image databases demonstrate the advantages of the proposed metric in terms of both prediction performance and generalization ability.*
Keywords: Quality assessment, Super resolution, Structure degradation, Ringing, blurring

1. **Introduction.** Super resolution (SR) image reconstruction is the technique to restore a high-resolution (HR) image from one or multiple low-resolution (LR) images [1]. SR has wide applications in machine vision, medical imaging and remote sensing, etc. A great number of image super-resolution algorithms have been reported [2, 3]. With these algorithms in hand, a natural question is how to evaluate their relative performances. Up to now, the performances of SR algorithms are mainly evaluated subjectively by visual comparison and objectively by popular image quality metrics, e.g., Peak Signal to Noise Ratio (PSNR) and Structural Similarity (SSIM) [4], which are often used to evaluate the related technologies of image processing [5]-[8]. PSNR and SSIM need high-quality reference images, which are usually not available in practical SR applications. Therefore the full-reference image quality metrics cannot be used for the quality assessment of SR images. While general-purpose no-reference (NR) and single distortion quality metrics have been proposed in the literature [9]-[20], they are limited in the quality assessment of SR images due to the lack of domain knowledge. As a result, we are devoted to the design of a NR quality metric for single-image super-resolution (SISR) reconstructed images.

In the past few years, a few works have been done on the quality assessment of image super-resolution. Yeganeh *et al.* [21] presented a natural scene statistics (NSS) approach to the quality assessment of super-resolved images. Yang *et al.* [22] conducted a benchmark study of the state-of-the-art SISR algorithms. Subjective study was first conducted, producing a subjectively rated SISR database with 540 images, which we call ECCV-2014 hereafter. Fang *et al.* [23] proposed a reduced-reference (RR) quality metric for image super-resolution. Markov Random Field (MRF) was first employed to obtain the pixel-wise correspondence between LR and super-resolved HR images. Ma *et al.* [24] conducted a large-scale human subjective study, producing a SISR database, which we call CVIU-2017. More recently, Wang *et al.* [25] built a Super-resolution Reconstructed Image Database (SRID), where 20 LR images were processed by six SISR algorithms and two interpolation methods.

In real-world applications, high-quality reference images are usually not available for SISR. Hence, the existing FR quality metrics are not applicable to the quality evaluation of SISR, and NR quality metrics are needed. Although a number of general-purpose NR image quality metrics have been proposed in the literature [9]-[16], they are mainly designed and tested for the common distortions in images, e.g., added noise and JPEG compression. However, super resolution image reconstruction tends to introduce additional impairments, particularly ringing and blurring effects. These combined distortions cannot be readily measured by the existing image quality models [25]. With these inspirations, this paper presents a NR image quality metric for SISR by incorporating domain knowledge. The proposed method consists of a structure measurement module and a SISR-relevant distortion measurement module.

2. **Methods.** The flowchart of the proposed quality metric is shown in Fig. 1. The idea of the proposed metric is that we believe the quality of super-resolved images can be evaluated from two aspects, namely structure degradation measurement and SR-induced distortion measurement. It has been demonstrated that human eyes perceive image quality mainly based on structure changes [4]. Therefore, structure degradation measurement is the first module of the proposed metric. Moreover, as a special image restoration technique, super-resolution tends to introduce additional impairments, typically ringing and blurring. So these SR-induced distortion measurement constitutes the second module of the proposed metric.
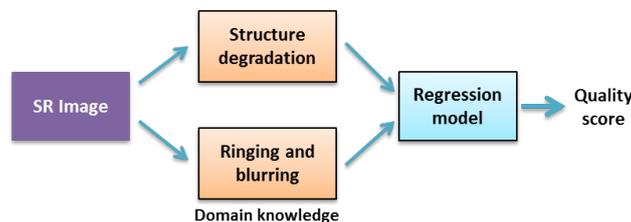

FIGURE 1. Flowchart of the proposed quality metric for image super-resolution.

2.1. **Measurement of Structure Degradation.** In the literature, a number of general-purpose NR image quality metrics have been proposed, which are mainly based on the measurement of structure degradations. As a special case of image quality assessment, structural distortions also occur in super-resolved images. Therefore, we include structure degradation measurement as the first module of the proposed method. In the existing metrics, Blind/Referenceless Image Spatial QUality Evaluator (BRISQUE) [11] has been

shown very effective for the evaluation of general structure distortions in common images. Therefore, we employ the BRISQUE features in this paper.

The BRISQUE features are extracted in the spatial domain based on natural scene statistics (NSS). Given an image $I(i,j), i \in [1, M], j \in [1, N]$, the Mean Subtracted Contrast Normalized (MSCN) coefficients are first computed as follows:

$$\hat{I}(i,j) = \frac{I(i,j) - \mu(i,j)}{\sigma(i,j) + C}, \tag{1}$$

with

$$\mu(i,j) = \sum_{k=-K}^{K} \sum_{l=-L}^{L} w(k,l) I_{k,l}(i,j), \tag{2}$$

$$\sigma(i,j) = \sqrt{\sum_{k=-K}^{K} \sum_{l=-L}^{L} w(k,l)(I_{k,l}(i,j) - \mu(i,j))^2}, \tag{3}$$

where $i \in [1, M], j \in [1, N]$, $\mu(i,j)$ and $\sigma(i,j)$ are the local mean and standard deviation of the surrounding local patch, and $\{w(k,l)|k \in [-K, K], j \in [-L, L]\}$ is the Gaussian weighting window of size $K \times L$. In this work, the window size is 3×3. C is a small constant for preventing the denominator to be zero.

The MSCN-based normalization can highlight texture information in images by reducing the dependencies between neighboring pixels, which is helpful for the subsequent structure feature extraction. Typically, the histogram of the MSCN coefficients follows the Generalized Gaussian Distribution (GGD) [11]. So GGD is employed to portray the distribution of MSCN coefficients:

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

where

$$\beta = \sigma\sqrt{\frac{\Gamma(1/\alpha)}{\Gamma(3/\alpha)}}, \tag{5}$$

and the gamma function $\gamma(a)$ is defined as:

$$\Gamma(a) = \int_0^{\infty} t^{a-1} e^{-t} dt, a > 0. \tag{6}$$

In GGD function, α denotes the shape of the distribution, and σ^2 denotes the variance. The two parameters (α, σ^2) constitute the first set of features.

In additional to the direct modeling of MSCN coefficients, the statistical relationships between neighboring MSCN coefficients are further modelled along four directions, namely horizontal (H), vertical (V), main-diagonal (D1) and secondary diagonal (D2):

$$H(i,j) = \hat{I}(i,j)\hat{I}(i, j+1), \tag{7}$$

$$V(i,j) = \hat{I}(i,j)\hat{I}(i+1, j), \tag{8}$$

$$D1(i,j) = \hat{I}(i,j)\hat{I}(i+1, j+1), \tag{9}$$

$$D2(i,j) = \hat{I}(i,j)\hat{I}(i+1, j-1), \tag{10}$$

where $i \in [1, M], j \in [1, N]$. Then the Asymmetric Generalized Gaussian Distribution (AGGD) with zero mode [11] is adopted to model the neighboring MSCN coefficients:

$$f(x; v, \sigma_l^2, \sigma_r^2) = \begin{cases} \frac{v}{(\beta_l + \beta_r)\Gamma\left(\frac{1}{v}\right)} \exp\left(-\left(\frac{-x}{\beta_l}\right)^v\right), & x < 0 \\ \frac{v}{(\beta_l + \beta_r)\Gamma\left(\frac{1}{v}\right)} \exp\left(-\left(\frac{-x}{\beta_r}\right)^v\right), & x \geq 0 \end{cases} \tag{11}$$

with

$$\beta_l = \sigma_l \sqrt{\frac{\Gamma(1/v)}{\Gamma(3/v)}}, \beta_r = \sigma_r \sqrt{\frac{\Gamma(1/v)}{\Gamma(3/v)}}, \tag{12}$$

where v controls the shape of the distribution, while σ_l^2 and σ_r^2 control the spread on each side of the mode. The parameters $(\eta, v, \sigma_l^2, \sigma_r^2)$ of the best AGGD fitting are produced, where η is determined by:

$$\eta = (\beta_r - \beta_l) \frac{\Gamma(2/v)}{\Gamma(1/v)}. \tag{13}$$

Since four parameters $(\eta, v, \sigma_l^2, \sigma_r^2)$ can be extracted for each direction, 16 features can be obtained in total. With the consideration that the HVS exhibits obvious multi-scale nature when perceiving the visual world, the above two groups of features are extracted in two scales, i.e., the original scale and a downsampled version by a factor of 2. Finally, a set of 36 features are extracted, which are used to characterize the structural degradations in super-resolved images.

2.2. **Measurement of SR-induced Artifacts.**

2.2.1. *Ringing Effect.* Ringing effect is a dominated distortion in super-resolved images, especially for high scaling ratios. Fig. 2 shows some example images after super-resolution reconstruction with ringing distortions around edges. In order to effectively measure the quality of super-resolved images, the ringing effect should be properly measured.

In this paper, we propose a no-reference approach to measure the ringing effects in the Log-Gabor transform domain [16]. It has been shown that the Log-Gabor filters resemble the human visual system (HVS) in terms of multi-scale and multi orientation representations of images. The Log-Gabor filter is defined as:

$$G_{s,o}(\omega, \theta) = \exp\left\{-\frac{[\log(\omega/\omega_s)]^2}{2[\log(\sigma_s/\omega_s)]^2}\right\} \times \exp\left[-\frac{(\theta - \mu_0)^2}{2\sigma_0^2}\right], \tag{14}$$

where $G_{s,o}$ represents the Log-Gabor filter with scale s and orientation o, ω is the normalized radial frequency and θ denotes the orientation. ω_s is the centre frequency of filter and σ_s/ω_s is used to determine the radial bandwidth, where ω_s and σ_s/ω_s are set to 2/3 and 0.65 in this work. $\sigma_0 = \frac{\pi}{1.5 \cdot \max(o)}$ determines the pattern bandwidth.

In this work, we adopt the Log-Gabor filter to decompose the image with two scales ($s = 1, 2$) and four orientations ($o = 1, 2, 3, 4$). The two scales are the original scale ($s = 1$) and the downsampled version ($s = 2$) by a factor of 2. The four orientations are calculated based on $\mu_0 = (o-1)\frac{\pi}{4}$, which correspond to 0°, 45°, 90° and 135°, respectively. Then the ringing effect is measured based on the decomposed images. Fig. 3 shows an example of the Log-Gabor decomposition in the four directions. In this paper, ringing effect is measured based on the decomposed subband images.

In the proposed method, the ringing effect is evaluated using the Log-Gabor subbands in four directions simultaneously. Let us take Fig. 3(b) as an example. The ringing is evaluated row by row. Fig. 4 shows the projection of Log-Gabor coefficients in one row of Fig. 3(b). Since ringing effects typically occur around edges, regions that may have ringing effects are marked by A, B, C, ..., G. Fig. 4(b) further shows the detailed distribution of coefficients of region B in Fig. 4(a). Since real edge regions in images usually have large oscillations, it is intuitive that local extreme points H8-H10 correspond to real edge. Oscillations around them are very likely to be ringings.

FIGURE 2. Super resolution reconstructed images with ringing effects (best viewing after zooming in).
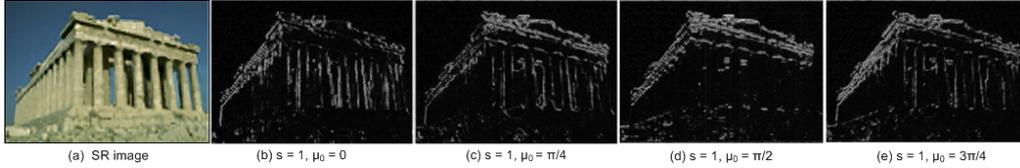

FIGURE 3. An example of Log-Gabor decomposition in four directions.
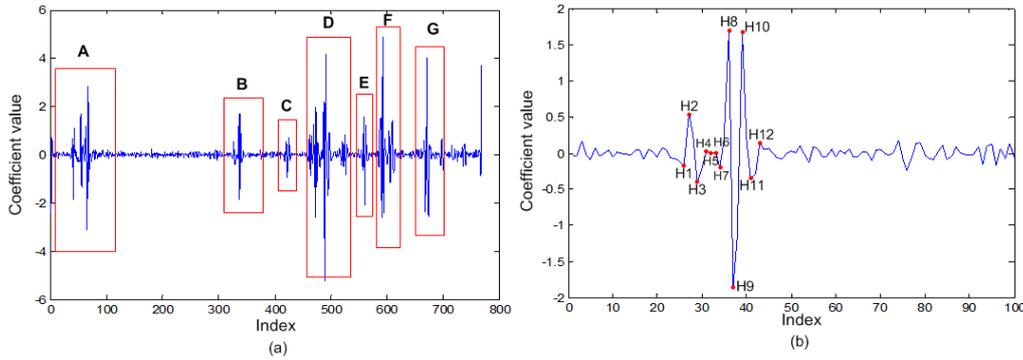

FIGURE 4. (a) Log-Gabor coefficients in one row of subband image Fig.3(b), (b) Magnified view of region B in (a).

In this work, the local extreme values are first determined and denoted by $H_i, i = 1, 2, \cdots, n$. Then the differences between adjacent extreme points are calculated as

$$T_i = |H_{i+1} - H_i|, i = 1, 2, 3 \cdots n - 1. \tag{15}$$

Then the ringing score in a row is computed by:

$$Q_{\text{row}} = \sum_{i=1}^{M} \mathbf{T}_1(i) - \sum_{j=1}^{N} \mathbf{T}_2(j), \tag{16}$$

with

$$\mathbf{T}_1 = \{T_i | T_i > \alpha \cdot \max(T_1, T_2, \cdots, T_{n-1}), i = 1, 2, \cdots, M\}, \tag{17}$$

$$\mathbf{T}_2 = \{T_j | T_j > \beta \cdot \max(T_1, T_2, \cdots, T_{n-1}), j = 1, 2, \cdots, N\}, \tag{18}$$

where $\alpha = 0.45$ and $\beta = 0.6$ are set empirically, which are used to alleviate the impact of smooth regions on the accuracy of ringing evaluation.

The ringing score of a subband image is defined as:

$$Q_1 = \sum_{k=1}^{K} Q_{\text{row}}(k), \tag{19}$$

where K denotes the number of rows in the subband image, $Q_{\text{row}}(k)$ denotes the ringing score in row k.

Similarly, the ringing scores in the other three subband images can be calculated, which are denoted by Q_2, Q_3 and Q_4, respectively. Finally, the overall ringing score of an image is calculated as follows:

$$Q_r = \sum_{s=1,2} \left(Q_{s,1} + Q_{s,2} + Q_{s,3} + Q_{s,4} \right), \tag{20}$$

where s denotes the scale, scale 1 represents the original scale, scale 2 is obtained by downsampling the original image by a factor of 2.

Fig. 5 shows an example of ringing effect evaluation using the proposed method on three images with different extents of ringing distortions. From Fig. 5(a) to Fig. 5(c), the ringing effects increase. Meantime, the predicted ringing scores increase accordingly. This indicates that the proposed method can evaluate the ringing distortions consistently with the perceived quality.
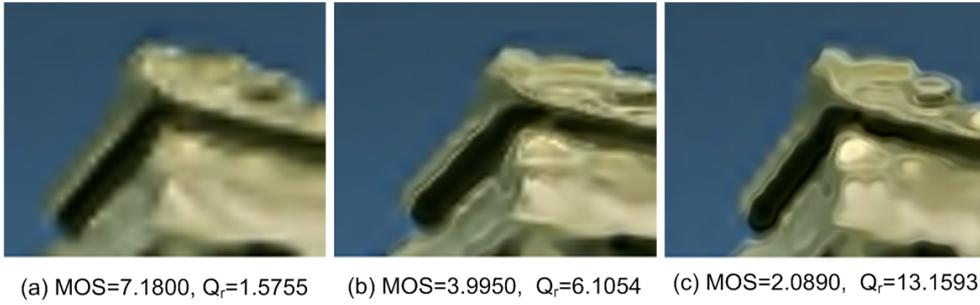

(a) MOS=7.1800, Q_r=1.5755   (b) MOS=3.9950, Q_r=6.1054   (c) MOS=2.0890, Q_r=13.1593

FIGURE 5. Ringing scores predicted by the the proposed method.

2.2.2. *Blurring Effect.* In image super-resolution, blurring is one of the dominated side effects, especially for high scaling factors. Generally, the higher the scaling factors, the more severe the blurring effects. In this paper, we employ the Fast Image SHarpness (FISH) model [26] to evaluate the blurring effect in super-resolved images, due to its computational efficiency and consistent good performances across images.

The FISH model is achieved in the Discrete Wavelet Transform (DWT) domain by measuring the log-energies of the subbands. For an image, it is first transformed into the DWT domain with three levels, and the high-frequency subbands are denoted by S_{LH_n}, S_{HL_n} and S_{HH_n}, n=1, 2, 3. Then the log-energy of each subband at level n is computed as:

$$E_{XY_n} = \log_{10} \left(1 + \frac{1}{N_n} \sum_{i,j} S^2_{XY_n}(i,j) \right), \tag{21}$$

where XY is LH, HL or HH, and N_n denotes the number of coefficients in the subband at level n. Then the total energy of the high-frequency subbands is obtained by:

$$E_n = (1 - \alpha) \frac{E_{LH_n} + E_{HL_n}}{2} + \alpha E_{HH_n}, \tag{22}$$

where α is set to 0.8 to give more weights to the HH subband. Finally, the sharpness score is defined as:

$$\text{FISH} = \sum_{i=1}^{3} 2^{3-n} E_n. \tag{23}$$

In this paper, we use the raw FISH score as a feature for measuring the blurring effect in super-resolved images.

2.3. Model Training and Quality Prediction. After obtaining the aforementioned features, they are further concatenated to form a feature vector. Then a quality regression model can be trained. In this paper, we employ two of the most commonly used regression approaches to train the quality model, i.e., Support Vector Regression (SVR) [27] and Random Forest (RF) [28].

3. Experimental Results and Discussion.

3.1. Evaluation Protocols.

3.1.1. *Databases.* In this work, we employ three SISR databases to evaluate the performance of the proposed metric and compare the results with those of other relevant state-of-the-art quality metrics. The databases are briefly reviewed here.

- CVIU-2017 database [24]. This database consists of 1620 super-resolved color images from 30 source images. Specifically, each source image is first processed by six different combinations of downsampling and blurring to generate six LR images. Then nine super-resolution image reconstruction algorithms are adopted to generate the HR images.
- ECCV-2014 database [22]. This database consists of 540 super-resolved monochrome images from 10 source images. Specifically, each source image is processed by nine different combinations of downsampling and blurring to generate nine LR images. Then six super-resolution image reconstruction algorithms are adopted to generate the HR images.
- SRID database [25]. In CVIU-2017 and ECCV-2014 databases, the LR images are generated by conducting downsampling and blurring based on HR original images. However, this process is not natural. In order to simulate the real application scenario, the authors built another super-resolution image database in [25].

3.1.2. *Performance Criteria.* Four commonly used criteria are utilized to measure the performance of the proposed quality metric. Specifically, Spearman Rank order Correlation Coefficient (SRCC) and Kendalls Rank Correlation Coefficient (KRCC) are used to measure prediction monotonicity, while Pearson Linear Correlation Coefficient (PLCC) and Root Mean Square Error (RMSE) are used to measure prediction accuracy. Before computing these values, a logistic fitting is first conducted between the predicted objective scores and the subjective scores. In this work, the five-parameter logistic fitting function is adopted [29]:

$$f(x) = \tau_1 \left(\frac{1}{2} - \frac{1}{1 + e^{\tau_2(x-\tau_3)}} \right) + \tau_4 x + \tau_5, \tag{24}$$

where x denotes the predicted score, $f(x)$ denotes the corresponding subjective score, and $\tau_i, i = 1, 2, 3, 4, 5$, are the parameters to be fitted. A good quality metric produces high SRCC, KRCC, PLCC values, as well as low RMSE value.

To test the performance of the proposed metric on a specific database, 80% of the images are randomly selected for training, and the rest 20% images are used for testing, based on which the above performance values are obtained. In order to avoid bias, this operation is repeated 1, 000 times and the median values are reported.

3.2. **Performance Evaluation.** In order to demonstrate the effectiveness and advantage of the proposed metric, we compare it with the state-of-the-art general-purpose no-reference image quality metrics, including BIQI [9], DIIVINE [10], BRISQUE [11], BLIINDS-II [12], NIQE [14], NFERM [15] and DESIQUE [16]. The performance of a recent SISR quality metric [24] is also included. For the proposed metric, we provide the results based on both SVR and RF regression models. Tables 1 to 3 summarize the experimental results on CVIU-2017, ECCV-2014 and SRID databases, where the best results are marked in boldface.

TABLE 1. Performances of different quality metrics on CVIU-2017 database.

Criterion	BIQI [9]	DIIVINE [10]	BRISQUE [11]	BLIINDS-II [12]	NIQE [14]
SRCC	0.7429	0.7699	0.8728	0.8180	0.6257
KRCC	0.5483	0.5791	0.6927	0.6223	0.4575
PLCC	0.7461	0.7889	0.8887	0.8292	0.6440
RMSE	1.5856	1.4775	1.1018	1.3442	1.8391
Criterion	NFERM [15]	DESIQUE [16]	Ref. [24]	Proposed (RF)	Proposed (SVR)
SRCC	0.8718	0.8310	0.8923	**0.9215**	0.8907
KRCC	0.6845	0.6426	0.7213	**0.7624**	0.7099
PLCC	0.8829	0.8569	0.9095	**0.9344**	0.9024
RMSE	1.1071	1.2474	0.9944	**0.8554**	1.0494

It is observed from the results that the proposed metric achieves the best performances in three databases. The state-of-the-art general-purpose NR image quality metrics only perform moderately in these databases.

With respect to the regression models, the RF-based approach achieves slightly better results than SVR-based approach in all databases. Meantime, it should be noted that the proposed metric with SVR achieves the second best results in both ECCV-2014 and SRID databases. In CVIU-2017 database, the proposed metric ranks the third and it is only slightly worse than Ref. [24], which is also designed specifically for SISR images. From these results, we know that the proposed method with RF achieves the best performances in all three databases.

3.3. **Generalization Ability.** In this subsection, the quality model is first trained using the whole CVIU-2017 database, which contains the most images. Then the trained model is used for quality evaluation on the other two databases. In implementation, we test both RF-based and SVR-based approaches. The experimental results are summarized in Tables 4 and 5, where the best results are marked in boldface.

It is observed from Tables 4 and 5 that the proposed metric (SVR-based) achieves the best performance in both cross-database tests. This indicates that the generalization ability of the proposed metric is better than those of the state-of-the-art metrics. For the RF-based approach, the performance ranks the second in the first cross-database test. However, it does not perform very well in the second test. The reason may be that the distortion characteristics of CVIU-2017 and SRID databases are more different. Especially, the SRID database consists of super-resolved images with obvious ringing effects when the scaling factor is 8. By comparison, CVIU-2017 does not have SR images with obvious ringing effects. As a result, the model trained based on CVIU-2017 cannot effectively capture the characteristics of ringing effects.

4. **Conclusion.** In this paper, we have been devoted to the objective quality evaluation of super-resolution reconstructed images. Distortions in super-resolved images can be classified into general structure distortions and domain-specific distortions. Based on these observations, we have proposed a learning-based quality metric to quantify both aspects of the distortions. We have evaluated the performance of the proposed metric on three

TABLE 2. Performances of different quality metrics on ECCV-2014 database.

Criterion	BIQI [9]	DIIVINE [10]	BRISQUE [11]	BLIINDS-II [12]	NIQE [14]
SRCC	0.4165	0.6040	0.8045	0.5940	0.4925
KRCC	0.2904	0.4351	0.6280	0.4371	0.3387
PLCC	0.5508	0.7172	0.8818	0.7412	0.6295
RMSE	1.6025	1.3063	0.9221	1.3047	1.4968
Criterion	NFERM [15]	DESIQUE [16]	Ref. [24]	Proposed (RF)	Proposed (SVR)
SRCC	0.7593	0.7433	0.7210	**0.8637**	0.8520
KRCC	0.5630	0.5600	0.5474	**0.6919**	0.6762
PLCC	0.8206	0.8219	0.8423	**0.9193**	0.9177
RMSE	1.0813	1.0739	1.0331	**0.7528**	0.7693

TABLE 3. Performances of different quality metrics on SRID database.

Criterion	BIQI [9]	DIIVINE [10]	BRISQUE [11]	BLIINDS-II [12]	NIQE [14]
SRCC	0.6281	0.6724	0.8287	0.6995	0.4606
KRCC	0.4623	0.4956	0.6556	0.5216	0.3296
PLCC	0.6806	0.6993	0.8521	0.7443	0.4639
RMSE	1.1645	1.1535	0.8431	1.0620	1.4328
Criterion	NFERM [15]	DESIQUE [16]	Ref. [24]	Proposed (RF)	Proposed (SVR)
SRCC	0.8164	0.7639	0.6597	**0.8793**	0.8720
KRCC	0.6347	0.5823	0.4765	**0.7025**	0.6942
PLCC	0.8479	0.8062	0.7413	**0.9052**	0.8929
RMSE	0.8498	0.9348	1.0899	**0.6780**	0.7216

TABLE 4. Performances of different quality metrics when trained on CVIU-2017 database and tested on ECCV-2014 database.

Criterion	BIQI [9]	DIIVINE [10]	BRISQUE [11]	BLIINDS-II [12]	NFERM [15]
SRCC	0.4105	0.4981	0.6346	0.5841	0.5770
KRCC	0.2821	0.3455	0.4545	0.4100	0.4008
PLCC	0.5638	0.6502	0.7729	0.6445	0.7244
RMSE	1.5911	1.4638	1.2225	1.4729	1.3282
Criterion	DESIQUE [16]	Ref. [24]	Proposed (RF)	Proposed (SVR)	
SRCC	0.5771	0.6040	0.7114	**0.7506**	
KRCC	0.4127	0.4200	0.5310	**0.5578**	
PLCC	0.7255	0.7714	0.8123	**0.8503**	
RMSE	1.3259	1.2260	1.1236	**1.0139**	

TABLE 5. Performances of different quality metrics when trained on CVIU-2017 database and tested on SRID database.

Criterion	BIQI [9]	DIIVINE [10]	BRISQUE [11]	BLIINDS-II [12]	NFERM [15]
SRCC	0.3043	0.4912	0.7151	0.5821	0.6585
KRCC	0.2085	0.3433	0.5205	0.4242	0.4737
PLCC	0.2539	0.4480	0.7496	0.5923	0.6610
RMSE	1.5644	1.4461	1.0705	1.3033	1.2137
Criterion	DESIQUE [16]	Ref. [24]	Proposed (RF)	Proposed (SVR)	
SRCC	0.4599	0.2338	0.5495	**0.7822**	
KRCC	0.3164	0.1445	0.3781	**0.5740**	
PLCC	0.4396	0.2500	0.5364	**0.8260**	
RMSE	1.4528	1.5661	1.3650	**0.9118**	

subjectively rated public super-resolution image databases. The experimental results and comparisons with the state-of-the-art metrics have demonstrated the advantages of the proposed method. The proposed metric also features good generalization ability, which is a highly desired property in practical applications.

REFERENCES

[1] S.C. Park, M.K. Park, M.G. Kang, Super-resolution image reconstruction: a technical overview, *IEEE Signal Process Magazine*, vol. 20, no. 3, pp. 21-36, 2003.

[2] J. Yang, J. Wright, T.S. Huang, Y. Ma, Image super-resolution via sparse representation, *IEEE Trans Image Process,* vol. 19, no. 11, pp. 2861-2873, 2010.

[3] Y. Romano, J. Isidoro, P. Milanfar, RAISR: rapid and accurate image super resolution, vol. 3, no. 1, pp. 110-125, 2017.

[4] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, *IEEE Trans Image Process,* vol. 13, no. 4, 2004, pp. 600-612.

[5] S. Weng, Y. Zhao, J. Pan, R. Ni, Reversible Watermarking Based on Invariability and Adjustment on Pixel Pairs, *IEEE Signal Process Lett* , vol. 15, pp. 721-724, 2008.

[6] S. Weng, J. Pan, L. Li, Reversible data hiding based on an adaptive pixel-embedding strategy and two-layer embedding, *Information Sciences*, vol. 369, pp. 144-159, 2016.

[7] T. Yin, G. Yang, L. Li, D. Zhang, X. Sun, Detecting seam carving based image resizing using local binary patterns, *Computers Security*, vol. 55, pp. 130-141, 2015.

[8] S. Weng, G. Zhang, J. Pan, Z. Zhou, Optimal PPVO-based reversible data hiding, *J. Visual Communication and Image Representation* , vol. 48, pp. 317-328,2017.

[9] A.K. Moorthy, A.C. Bovik, A two-step framework for constructing blind image quality indices, *IEEE Signal Process Lett,* vol. 17, no. 5, pp. 513-516, 2010.

[10] A.K. Moorthy, A.C. Bovik, Blind image quality assessment: from natural scene statistics to perceptual quality, *IEEE Trans Image Process,* vol. 20, no. 12, pp. 3350-3364, 2011.

[11] A. Mittal, A.K. Moorthy, A.C. Bovik, No-reference image quality assessment in the spatial domain, *IEEE Trans Image Process,,* vol. 21, no. 12, pp. 4695-4708, 2012.

[12] M.A. Saad, A.C. Bovik, C. Charrier, Blind image quality assessment: a natural scene statistics approach in the DCT domain, *IEEE Trans Image Process,* vol. 21, no. 8, pp. 3339-3352, 2012.

[13] P. Ye, J. Kumar, L. Kang, D. Doermann, Unsupervised feature learning framework for no-reference image quality assessment, *IEEE Int Conf Computer Vision Pattern Recognition,* vol. 157, pp. 1098-1105, 2012.

[14] A. Mittal, R. Soundararajan, A.C. Bovik, Making a "completely blind" image quality analyzer, *IEEE Signal Process Lett,* vol. 20, no. 3, pp. 209-212, 2013.

[15] K. Gu, G.T. Zhai, X.K. Yang, W.J. Zhang, Using free energy principle for blind image quality assessment, *IEEE Trans Multimed,* vol. 17, no. 1, pp. 50-63, 2015.

[16] Y. Zhang, D.M. Chandler, An algorithm for no-reference image quality assessment based on log-derivative statistics of natural scenes, *J Electron Imag,* vol. 8653, no. 4, pp. 1-11, 2013.

[17] J. Wu, W. Lin, G. Shi, L. Li, Y. Fang, Orientation selectivity based visual pattern for reduced-reference image quality assessment, *Information Sciences,* vol. 351, pp. 18-29, 2016.

[18] L. Li, W. Lin, X. Wang, G. Yang, K. Bahrami, A.C. Kot, No-reference image blur assessment based on discrete orthogonal moments, *IEEE transactions on cybernetics,* vol. 46, no. 1, pp. 39-50, 2016.

[19] L. Li, Y. Zhou, W. Lin, J. Wu, X. Zhang, B. Chen, No-reference quality assessment of deblocked images, *Neurocomputing,* vol. 177, pp. 572-584,2016.

[20] L. Li, W. Lin, H. Zhu, Learning Structural Regularity for Evaluating Blocking Artifacts in JPEG Images, *IEEE Signal Processing Letters,* vol. 21, no. 8, 2014, pp. 918-922.

[21] H. Yeganeh, M. Rostami, Z. Wang, Objective quality assessment for image super-resolution: A natural scene statistics approach, *IEEE Int Conf Image Proc,,* pp. 1481-1484, 2013.

[22] C.Y. Yang, C. Ma, M.H. Yang, Single image super-resolution: a benchmark, *European Conference on Computer Vision* 8692, pp. 372-386,2014.

[23] Y. Fang, J. Liu, Y. Zhang, W. Lin, G. Guo, Quality assessment for image super-resolution based on energy change and texture variation, *IEEE Int Conf Image Proc,* pp. 2057-2061, 2016.

[24] C. Ma, C.Y. Yang, X. Yang, M.H Yang, Learning a no-reference quality metric for single-image super-resolution, *Computer Vision Image Understanding,* vol.158, pp. 1-16, 2017.

[25] G. Wang, L. Li, Q. Li, K. Gu, Z. Lu, J. Qian, Perceptual evaluation of signle-image super-resolution reconstruction, *IEEE Int Conf Image Proc*, 2017.

[26] P.V. Vu, D.M. Chandler, A fast wavelet-based algorithm for global and local image sharpness estimation, *IEEE Signal Process Lett,* vol. 19, no. 7, pp. 423-426, 2012.

[27] C.C. Chang, C.J. Lin, LIBSVM: a library for support vector machines, *ACM Trans Intelligent Symp Technol,* vol. 2, no. 3, 2011.

[28] L. Breiman, Random forests, *Machine Learning,* vol.45, no. 1, pp. 5-32, 2001.

[29] L. Li, D. Wu, J. Wu, H. Li, W. Lin, A.C. Kot, Image sharpness assessment by sparse representation, *IEEE Trans Multimed,* vol. 18, no. 6, pp. 1085-1097, 2016.