

Received December 25, 2016, accepted January 27, 2017, date of publication January 31, 2017, date of current version March 13, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2661858

No-Reference Quality Assessment of Deblurred Images Based on Natural Scene Statistics

LEIDA LI^{1,4}, (Member, IEEE), YA YAN¹, ZHAOLIN LU¹, JINJIAN WU²,
KE GU³, AND SHIQI WANG⁴

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

²School of Electronic Engineering, Xidian University, Xi'an 710071, China

³School of Computer Science and Engineering, Nanyang Technological University, 639798, Singapore

⁴School of Electrical and Electronic Engineering, Nanyang Technological University, 639798, Singapore

Corresponding author: L. Li (reader1104@hotmail.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61379143 and Grant 51604217, in part by the National Key Research and Development Program under Grant 2016YFC0801800, in part by the Fundamental Research Funds for the Central Universities under Grant 2015QNA66, and in part by the Qing Lan Project.

ABSTRACT Blurring is one of the most common distortions in digital images. In the past decade, extensive image deblurring algorithms have been proposed to restore a latent clean image from its blurred version. However, very little work has been dedicated to the quality assessment of deblurred images, which may hinder further development of more advanced deblurring techniques. Motivated by this, this paper presents a no-reference quality metric for defocus deblurred images based on Natural Scene Statistics (NSS). Two categories of NSS features are extracted in both the spatial and frequency domains to account for both the global and local aspects of distortions in deblurred images. Specifically, the spatial domain NSS features are used to characterize the global naturalness, and the frequency domain NSS features are used to portray the local structural distortions. All features are combined to train a support vector regression model for quality prediction of defocus deblurred images. The performance of the proposed metric is evaluated in a subjectively rated defocus deblurred image database. The experimental results demonstrate the advantages of the proposed metric over the relevant state-of-the-arts. As an application, the proposed metric is further used for benchmarking deblurring algorithms and very encouraging results are achieved.

INDEX TERMS Image quality assessment, defocus deblurring, natural scene statistics, support vector regression.

I. INTRODUCTION

Image quality assessment (IQA) has been increasingly popular due to the prevalence of new types of image processing and communication systems [1]–[4]. The underlying principle of objective IQA is to design computational models for measuring image distortions, and meantime to maintain consistency with the human perception. Objective IQA models are pretty useful in many applications, such as image restoration [5], image/video compression [6], [7], image forensics [8], etc. According to the required amount of reference information, the current IQA metrics can be classified into full-reference (FR), reduced-reference (RR) and no-reference (NR) [9]. For FR metrics, high-quality pristine images are needed, which are usually not available in practice. Therefore, NR metrics are more useful for practical applications [5].

Image restoration is the technique to estimate a latent clean image from its corrupted version. Since images are very

easily subject to distortions in all stages of the processing chain, image restoration has been a classical problem in image processing, such as denoising, deblocking, deblurring, etc. Since the objective of image restoration is always to generate high-quality output images, the accurate quality evaluation of restored images is of great concern. Most of the current image restoration algorithms employ the peak signal-to-noise ratio (PSNR) for measuring the quality of restored images. Unfortunately, it has been widely acknowledged that PSNR does not correlate well with human perception [10]. In addition, PSNR is a FR quality metric, but a high-quality original image is not available in image restoration.

Recently, several pioneering works have been done towards the perceptual evaluation of image restoration. In [11], Yeganeh *et al.* proposed a quality model for image super resolution (SR). The statistical model of frequency energy falloff was first built using high quality

natural images. Then the quality of a SR reconstructed image was calculated by quantifying the departure from this model. Zhu *et al.* [12] proposed a quality model for automated parameter selection in image denoising. The distortions in denoised images were measured based on the singular value decomposition of local image gradient. In [13], Kong *et al.* proposed NR quality metric for image auto-denoising by maximizing the structure similarities between the input noisy image and the estimated noise as well as the denoised image. Zeng *et al.* [14] also proposed a quality model for denoised images by measuring structural fidelity and statistical naturalness. Structural distortion was evaluated in the wavelet domain following the SSIM approach [10]. For naturalness, natural scene statistics (NSS) features were extracted based on the marginal distributions of wavelet coefficients and frequency energy fall-off characteristics. In [15], Ma *et al.* conducted a subjective user study on image dehazing. Then the performances of the existing quality metrics were tested on the dehzaed images. Li *et al.* [16] proposed a quality metric for image deblocking. A sharpness module in texture areas and a blockiness module in smooth areas were combined to produce the overall quality score for deblocked images. This metric was further used for optimizing image deblocking algorithms. More recently, Liang *et al.* [17] proposed a general comparison-based framework for quality evaluation of restored images. This method was shown effective in the automatic parameter selection of image denoising algorithms. The aforementioned approaches have achieved notable success towards the quality evaluation of restored images. However, it is worth noting that image restoration problems are usually application-specific. Different restoration problems tend to produce images with different distortion characteristics. So different quality models are needed for different restoration types.

Blurring is one of the most commonly encountered distortion types in practice. As a classical image restoration problem, deblurring has been extensively investigated. However, not much effort has been dedicated to the development of perceptual evaluation models for deblurred images. Lai *et al.* [18] conducted a large-scale user study of single image motion deblurring. The performances of several FR and NR image quality metrics were then evaluated. The results show that the current quality metrics are quite limited in the perceptual evaluation of motion deblurred images. Hou *et al.* [19] proposed the transduced contrast-to-distortion ratio (TCDR) for measuring the quality of motion deblurred images, which is also a FR metric. It was calculated based on the luminance maps of two images by considering the characteristics of perceived contrast. In [20], Liu *et al.* addressed a NR quality metric for motion deblurred images based on a comprehensive user study of blurring, noise and ringing effect. A set of low-level features was extracted and Logistic Regression (LR) was employed to generate the overall quality score.

In parallel with motion deblurring, defocus blurring is also very common in practice. However, little work has been

reported for the perceptual evaluation for image defocus deblurring. Although similar, motion deblurring quality metrics are not readily applicable to defocus deblurred image, which will be demonstrated in the experiment section of this paper. The only one related work is [21], where the authors proposed a defocus deblurred image quality metric based on the concept of basic edges. However, this method is a FR one, so the ground truth clean image is needed and thus the application scope is quite limited. Motivated by this, this paper presents a NR quality metric for defocus deblurred images based on natural scene statistics. Two categories of NSS features are extracted for characterizing the loss of global and local naturalness in defocus deblurred images, which are achieved in the spatial domain and frequency domain, respectively. Support vector regression (SVR) is employed to build the quality model. The experiments in a defocus deblurred image database (DDID) demonstrate that the proposed metric is advantageous over the relevant state-of-the-arts.

II. PROPOSED NR QUALITY METRIC FOR DEFOCUS DEBLURRED IMAGES

Image deblurring tries to estimate the latent sharp image from a blurred version. While the blurring effect is alleviated after deblurring, the deblurred image tends to be unnatural, especially when the original blurring is heavy. Fig. 1 shows an example of image defocus deblurring using two popular image deblurring algorithms TVMM [22] and SADCT [23]. It is observed from the figure that after deblurring, the blurring effect is significantly alleviated. Meantime, it is also noticed that the deblurred images exhibit obvious unnaturalness, particularly in the highlighted regions. Since image deblurring is a typical ill-posed problem, naturalness loss is inevitable, which also sheds light on how the quality of deblurred images should be measured.

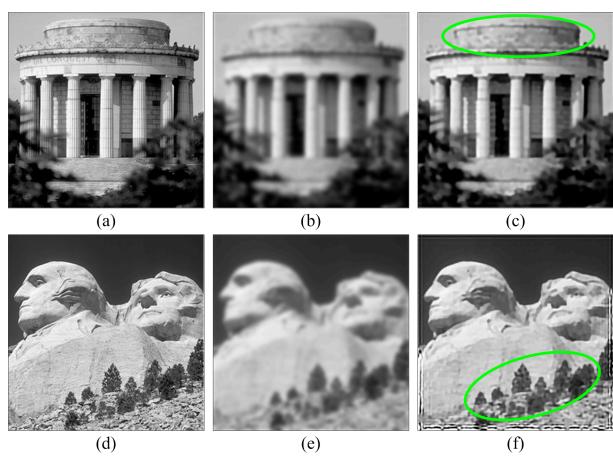


FIGURE 1. An example of image defocus deblurring. (a) and (d) are original images; (b) and (e) are blurred images; (c) is the deblurred image of (b) using TVMM [22]; (f) is the deblurred image of (e) using SADCT [23].

In this paper, we propose to evaluate the quality of defocus deblurred images by learning both global and local natural scene statistics features. This is consistent with the

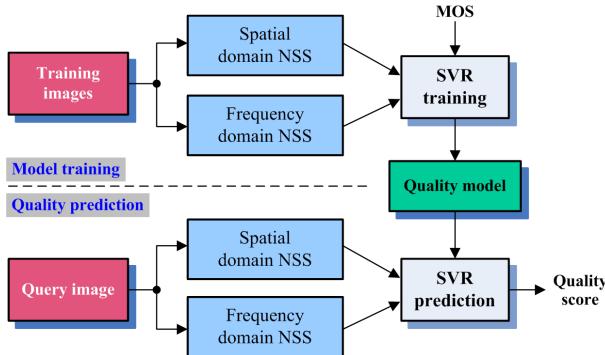


FIGURE 2. Flowchart of the proposed quality metric for defocus deblurred images.

working mechanism of the human visual system (HVS), because human eyes employ both global-to-local and local-to-global strategies for judging the quality of images with different extents of distortions [24]. Fig. 2 illustrates the flowchart of the proposed metric, which consists of a model training phase and a quality prediction phase. The basic idea behind the proposed method is to measure the loss of naturalness in deblurred images by exploiting NSS features in both the spatial domain and the frequency domain. The spatial domain NSS features are used to characterize the loss of global naturalness, while the frequency domain NSS features are used to portray the local structural distortions in different scales and orientations. The two groups of NSS features are combined to learn a SVR model, which is used for the subsequent quality prediction of defocus deblurred images. In the following subsections, we shall detail the extraction of the NSS features.

A. SPATIAL DOMAIN NSS FEATURES

Image deblurring is an ill-posed inverse problem. In order to achieve better deblurring performance, prior information is commonly adopted. Gradient distribution prior (GDP) is such a kind of image prior that has been widely used in image restoration [25]. It has been proved that GDP has two key properties [26]. (1) GDP is closely related to image quality. Psychophysics studies have shown that the human visual system mainly detects gradient information for processing. Furthermore, the neurons have been evolved to be adapted to the environment based on gradient distribution. (2) GDP is pretty stable. Different people have almost the same visual perception, so the gradient distribution of natural scene images is stable. These properties naturally fit into the requirements of image quality assessment. In [25], the authors proposed to use the gradient and Laplace distribution priors for image enhancement. In this work, we employ them to measure the global naturalness of deblurred images.

1) GRADIENT DISTRIBUTION

For a gray-scale image $I(x, y)$, the gradient G is defined as follows:

$$G = (G^x, G^y) = (\nabla_x I(x, y), \nabla_y I(x, y)), \quad (1)$$

where ∇_x and ∇_y denote the finite-difference approximations in x and y directions respectively. In this work, $\nabla_x I(x, y) = I(x + 1, y) - I(x, y)$ and $\nabla_y I(x, y) = I(x, y + 1) - I(x, y)$. It is intuitive that for gray-scale images, the gradient values are in the range $[-255, 255]$. Based on the gradient maps G^x and G^y , the normalized gradient histograms can be easily calculated, which are denoted by H^x and H^y , respectively.

The gradient histograms H^x and H^y can be effectively modeled based on the cumulative distribution function (CDF), which is defined as:

$$C(G) = \int_{-255}^{G^x} \int_{-255}^{G^y} P(x, y) dx dy, \quad (2)$$

where $P(x, y)$ denotes the probability of gradient value. In [25], the above CDF is approximated by the following model:

$$\tilde{C}(G) = \left(\frac{\text{atan}(T_1 G^x)}{\pi} + \frac{1}{2} \right) \left(\frac{\text{atan}(T_1 G^y)}{\pi} + \frac{1}{2} \right), \quad (3)$$

where T_1 is the model parameter to be fitted.

2) LAPLACE DISTRIBUTION

The Laplace field of image $I(x, y)$ is defined as:

$$L(x, y) = \Delta I(x, y), \quad (4)$$

where Δ is the Laplace operator.

Similar to the gradient CDF, the Laplace CDF is defined as:

$$L(t) = \int_{-\infty}^t P(\Delta I(x)) d\Delta I(x). \quad (5)$$

The following parametric model is employed to approximate the Laplace CDF [25]:

$$\tilde{L}(t) = \frac{\text{atan}(T_2 t)}{\pi} + \frac{1}{2}, \quad (6)$$

where T_2 is the model parameter to be fitted.

3) NATURALNESS FACTOR

The gradient and Laplace distributions denote the first-order and second-order statistics of images, which are both effective for representing natural images. The model parameters T_1 and T_2 can be combined to define the image naturalness factor N_f . For a high quality image, the values are expected to be similar to the priors. To be specific, the gradient distribution and Laplace distribution priors can be first aggregated from a large number of natural-scene images. Then the distance between the gradient/Laplace distributions of a query image and the prior distributions indicates the naturalness level of the query image. In accord with this, the naturalness factor is defined as:

$$N_f = (1 - \theta) \frac{T_1}{T_1^{pr}} + \theta \frac{T_2}{T_2^{pr}}, \quad (7)$$

where $\theta \in [0, 1]$ is a weighting factor, T_1^{pr} and T_2^{pr} are the aforementioned priors, which are 0.3754 and 0.1446 based on the aggregation results on 23613 natural-scene images [25].

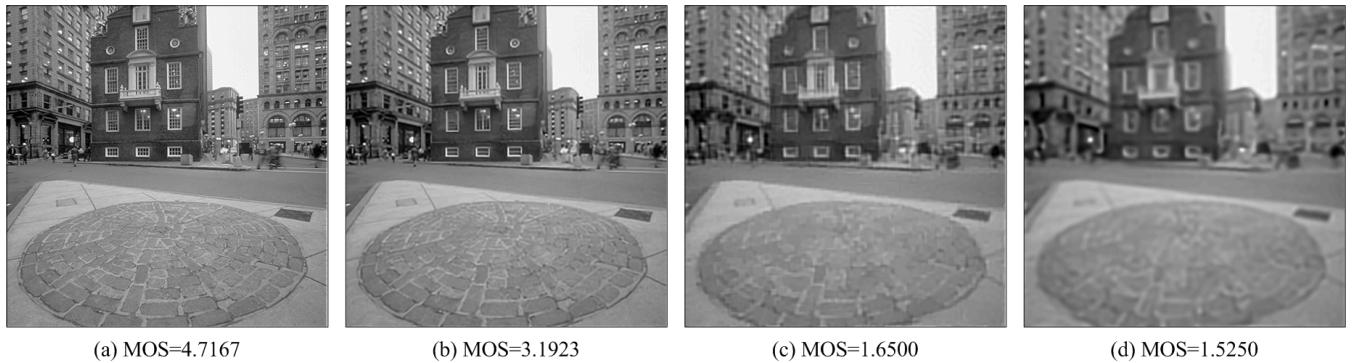


FIGURE 3. Four deblurred images and their subjective scores (MOS).

TABLE 1. Estimated naturalness parameters for the deblurred images shown in Fig. 3.

Parameter	Fig. 3(a)	Fig. 3(b)	Fig. 3(c)	Fig. 3(d)
MOS	4.7167	3.1923	1.6500	1.5250
T_1	0.3885	0.5805	0.6663	0.8311
T_2	0.1522	0.3410	0.6425	0.9158
N_f	1.0440	1.9523	3.1091	4.2737

For high-quality natural images, the N_f value should be close to 1. In this work, we set $\theta = [0.5, 0.55, 0.6, \dots, 1]$, so 11 naturalness factors can be obtained. These values are used as the first group of NSS features.

Fig. 3 shows four defocus deblurred images and their subjective ratings indicated by the mean opinion score (MOS) values. Table 1 lists the model parameters T_1 , T_2 and the corresponding naturalness factor N_f (θ is set to 0.5). It is observed from the table that with the decreased qualities from Fig. 3(a) to Fig. 3(d), T_1 and T_2 increase monotonically. Furthermore, the N_f values are getting more and more bigger than 1, which indicate that the images are becoming more and more unnatural, namely worse quality.

B. FREQUENCY DOMAIN NSS FEATURES

It has been shown that the neurons in the HVS exhibit orientation and frequency selection mechanism in perceiving visual scenes [27]. Therefore, the deployment of multi-scale and multi-resolution features is expected to facilitate more advanced quality evaluation. The frequency and orientation characteristics of the Gabor filters behave similarly to the simple cortical cells in the HVS [28]. Inspired by this, we further extract multi-scale and multi-resolution NSS features in the frequency domain to portray the local structural distortions in deblurred images. In this paper, we adopt the Log-Gabor filters to achieve this goal [29].

The Log-Gabor filter is defined as:

$$G_{s,o}(\omega, \theta) = \exp \left\{ -\frac{[\log(\omega/\omega_s)]^2}{2[\log(\sigma_s/\omega_s)]^2} \right\} \times \exp \left\{ -\frac{(\theta - \mu_0)^2}{2\sigma_0^2} \right\}, \quad (8)$$

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. In implementation, the input deblurred image is decomposed at the highest scale ($s = 1$) and two orientations ($o = 1, 2$). Similar to [29], the central frequency is set to be $1/3$, and accordingly $\omega_1 = 2/3$. The bandwidth parameter σ_1/ω_1 is set to be 0.975. The two orientations are calculated based on $\mu_0 = (o - 1)\pi/2$, which correspond to 0 and 90 degree.

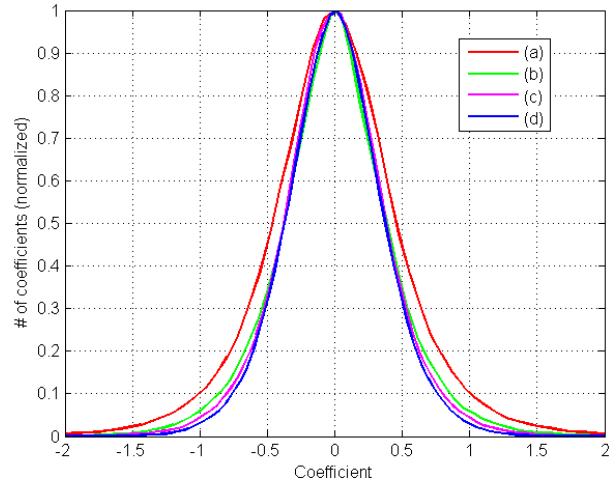


FIGURE 4. Histogram distribution of Log-Gabor coefficients in scale 1 and orientation 0 degree for images shown in Fig. 3.

The distribution of Log-Gabor coefficients is sensitive to image distortions, so it can be employed for quality evaluation. Fig. 4 shows the histogram distributions of the Log-Gabor coefficients in scale 1 and orientation 0 degree for the four images shown in Fig. 3. It is observed from the figure that from Fig. 3(a) to Fig. 3(d), the distributions become more heavy-tailed and center-peaked. As a result, after modeling these distributions, the model parameters can be used as perceptual features for quality assessment of deblurred images.

In [29], several types of derivatives are defined to account for orientation selectivity mechanism of the HVS in both the spatial and frequency domains. For each subband, the Generalized Gaussian Distribution (GGD) is employed to model

TABLE 2. Performance comparison of the proposed metric and the existing NR image blur metrics in the DDID database.

Criterion	PBRM [38]	JNB [39]	CPBD [40]	S3 [41]	FISH [42]	LPC [43]	SVC [44]	ARISM [45]	BIBLE [46]	SPARISH [47]	Proposed
PLCC	0.7479	0.6900	0.8151	0.7700	0.8453	0.8991	0.7791	0.7808	0.8119	0.8061	0.9459
SRCC	0.7280	0.6806	0.8092	0.7049	0.8142	0.8325	0.7813	0.7786	0.7317	0.7538	0.9016
RMSE	0.9477	1.0332	0.8272	0.9108	0.7628	0.6249	0.8949	0.8919	0.8333	0.8449	0.4546

the histogram of log-derivative statistics of the frequency coefficients:

$$f(x; \alpha, \sigma^2) = \frac{\alpha}{2\beta\Gamma\left(\frac{1}{\alpha}\right)} \exp\left[-\left(\frac{|x|}{\beta}\right)^\alpha\right], \quad (9)$$

where $\beta = \sigma\sqrt{\frac{\Gamma(1/\alpha)}{\Gamma(3/\alpha)}}$, and $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$, $x > 0$ is the Gamma function. The model parameters (α, σ) are used as frequency domain NSS features.

In implementation, similar to [29], we also extract 24 NSS features from the 0 deg and 90 deg subbands based on six derivative types. Besides, the original image is also downsampled by two times, and additional four features are extracted based on one derivative type from the two subbands. Finally, 28 Log-Gabor domain NSS features are obtained. More details of the feature extraction can be found in [29].

C. MODEL TRAINING AND QUALITY PREDICTION

In this work, we employ the support vector regression (SVR) [30] to learn the quality model for defocus deblurred images. To be specific, given a set of deblurred images, the spatial and frequency NSS features are extracted using the approaches described above. Then the NSS features, together with the subjective ratings, are fed into a SVR for training. Finally, the trained SVR model is used for quality prediction of query deblurred images. In this work, the Radial Basis Function (RBF) is used as the SVR kernel. Although linear and polynomial kernels can also be used in SVR, in implementation we find that the RBF kernel delivers the best performances. So we use the RBF kernel in this paper.

III. EXPERIMENTAL RESULTS AND ANALYSIS

A. EVALUATION PROTOCOLS

The performance of the proposed metric is evaluated based on a recently developed subjectively-rated defocus deblurred image database (DDID) [31]. The DDID database contains 240 defocus deblurred images with different distortion levels, which are generated by eight popular image defocus deblurring algorithms including TVMM [22], SADCT [23], FISTA [32], ASDS-AR [33], CSR [34], BM3D [35], NCSR [36] and JSM [37]. Mean opinion score (MOS) values are provided as the ground truth of image quality. Some sample images from the DDID database are shown in Fig. 5.

Prediction accuracy and monotonicity are the two commonly used criteria for evaluating the performance of image quality metrics. In this paper, Pearson Linear Correlation

Coefficient (PLCC) and Root Mean Squared Error (RMSE) are used to measure the prediction accuracy, and Spearman Rank order Correlation Coefficient is employed to measure the prediction monotonicity [51]. In order to compute these values, a five-parameter logistic mapping is first performed between the predicted and subjective scores, which is used to bring the predicted scores to the same scale with the subjective scores:

$$f(x) = \tau_1 \left(\frac{1}{2} - \frac{1}{1 + e^{\tau_2(x - \tau_3)}} \right) + \tau_4 x + \tau_5, \quad (10)$$

where τ_i , $i = 1, 2, \dots, 5$, are the fitting parameters.

In our experiment, 80% of the images in DDID database are randomly selected for training, and the remaining 20% images are used for test. In order to obtain reliable results, this process is repeated 1000 times and the median performance values are reported. This is a standard approach for the performance evaluation of learning-based image quality metrics [51].

B. PERFORMANCE EVALUATION

1) COMPARISON WITH NR IMAGE BLUR METRICS

For all deblurring algorithms, it is unlikely that the blurring can be completely removed. So the residual blurring still has great impact on the overall quality of deblurred images. Therefore, we first compare the performance of the proposed metric with the existing image blur/sharpness metrics, ranging from the classical to the state-of-the-art. Specifically, the following ten blur metrics are compared: PBRM [38], JNB [39], CPBD [40], S3 [41], FISH [42], LPC [43], SVC [44], ARISM [45], BIBLE [46] and SPARISH [47]. These blur metrics are all NR approaches, and the codes are either downloaded from the authors' homepages or provided by the original authors. Table 2 summarizes the experimental results.

It is easily known from Table 2 that the existing blur metrics do not perform very well on defocus deblurred images. Many of the performance values are below 0.8, especially the monotonicity criterion SRCC. By comparison, the proposed metric produces much better results, and both PLCC and SRCC are higher than 0.9. This indicates that the predicted scores are more consistent with the subjective scores. As a result, only measuring blur is not sufficient for the quality assessment of deblurred images. This is, in fact, intuitive because during deblurring other kinds of distortions are very likely to be introduced, causing loss of overall naturalness.

TABLE 3. Performance comparison of the proposed metric and the general-purpose NR image quality metrics in the DDID database.

Criterion	BIQI [48]	DIIVINE [49]	BLIINDS-II [50]	BRISQUE [51]	NIQE [52]	DESIQUE [29]	CORNIA [53]	NFERM [54]	Proposed
PLCC	0.8497	0.7827	0.7469	0.8606	0.8826	0.8077	0.8863	0.8812	0.9459
SRCC	0.8026	0.7867	0.7172	0.8411	0.8611	0.8038	0.8767	0.8544	0.9016
RMSE	0.7528	0.8885	0.9493	0.7271	0.6712	0.8417	0.6613	0.6749	0.4546

**FIGURE 5.** Sample images of the DDID database (best viewed after zooming in).

2) COMPARISON WITH GENERAL-PURPOSE NR IMAGE QUALITY METRICS

General-purpose NR image quality metrics have been proposed for evaluating image quality without knowing the distortion types in advance. In principle, this kind of quality metrics can be used for quality evaluation of all distortion types. In this part, we compare the proposed metric with the state-of-the-art general-purpose NR image quality metrics, including BIQI [48], DIIVINE [49], BLIINDS-II [50], BRISQUE [51], NIQE [52], DESIQUE [29], CORNIA [53] and NFERM [54]. Again, the source codes of all the above metrics are downloaded from the authors' websites. Table 3 summarizes the experimental results.

It is observed from Table 3 that the proposed metric outperforms all the compared metrics. Some of the general-purpose metrics can produce quality scores that moderately correlate with subjective ratings. In addition, compared with the blur metrics in Table 2, these general-purpose NR quality metrics generally perform better.

3) COMPARISON WITH RELEVANT IMAGE RESTORATION QUALITY METRICS

We further compare the proposed model with two relevant quality models that are designed for image restoration. The first one is the Logistic Regression (LR) based quality metric for motion deblurred images [20]. The second one is the comparison-based framework for general image

TABLE 4. Comparison of the proposed metric with a NR motion deblurring quality metric [20] and two general-purpose image restoration quality metrics [17]. CQ: Comparison-based IQA; CTQ: Comparison-Texture-based IQA.

Criterion	LR [20]	CQ [17]	CTQ [17]	Proposed
PLCC	0.7891	0.4160	0.3048	0.9459
SRCC	0.7598	0.4600	0.3125	0.9016
RMSE	0.8769	1.2982	1.3596	0.4546

TABLE 5. Individual contributions of the two categories of NSS features.

Criterion	N_f	Log-Gabor	All
PLCC	0.8956	0.9035	0.9459
SRCC	0.8704	0.8745	0.9016
RMSE	0.6304	0.6182	0.4546

restoration [17]. In [17], there are two versions of the model, i.e., Comparison-based IQA (CQ) and Comparison-Texture-based IQA (CTQ), and we include both versions for comparison. Table 4 summarizes the experimental results in the DDID database.

It is observed from Table 4 that all three models do not perform well on defocus deblurred images. Although [20]

TABLE 6. Performance rankings of eight deblurring algorithms according to the MOS values and predicted scores by different image quality metrics. Number “1” represents the best performance and “8” represents the worse performance. “Statistics” denotes the number of objective rankings that are consistent with the subjective rankings. Consistent rankings are marked in boldface.

Algorithm	MOS	CPBD	FISH	LPC	CORNIA	BRISQUE	DESIQUE	NFERM	DIIVINE	NIQE	Proposed
BM3D [35]	1	6	2	1	3	1	4	1	3	1	1
JSM [37]	2	8	7	3	1	6	7	7	7	5	2
SADCT [23]	3	3	6	8	6	3	5	2	2	2	5
ASDS_AR [33]	4	7	1	7	8	7	8	5	8	6	3
TVMM [22]	5	5	3	5	5	8	6	6	6	8	4
CSR [34]	6	4	4	4	4	2	2	4	4	4	6
NCSR [36]	7	1	5	6	7	5	3	3	5	3	7
FISTA [32]	8	2	8	2	2	4	1	8	1	7	8
Statistics	/	2	1	2	2	2	0	2	0	1	5

is also a metric for image deblurring, it does not perform well on defocus deblurred images. This indicates that motion deblurring and defocus deblurring, which both are deblurring, are still quite different in terms of the distortion characteristics. So motion deblurring quality metrics are not readily applicable to defocus deblurring. Another finding is that the comparison-based IQA model [17] is also not efficient for defocus deblurred images. This further confirms that different image restoration problems need different quality models.

4) CONTRIBUTIONS OF COMPONENTS

In the proposed metric, two categories of NSS features are employed, i.e., the gradient distribution based spatial domain NSS and the Log-Gabor based frequency domain NSS. In order to know the relative contributions of the two components, we train two quality models using each group of NSS features separately. Then the performances of the new models are tested in the DDID database. Table 5 lists the experimental results.

It is observed from Table 5 that both kinds of NSS features deliver very good performances, which are already better than those of the compared metrics in Tables 2 and 3. When they are used together, much better results are obtained. This indicates that both kinds of NSS features are needed in the proposed model, and they have complementary contributions to the overall performance.

C. APPLICATION IN BENCHMARKING

DEBLURRING ALGORITHMS

In the literature, extensive image deblurring algorithms have been proposed. However, how to objectively evaluate the performances of these algorithms is still an open problem. An objective image quality model is thus highly desirable for this purpose. In this part, we apply the proposed quality model for benchmarking image defocus deblurring algorithms. To this end, we first rank the deblurring algorithms according to the mean MOS values, which are used as ground truth

of image quality. Specifically, if the mean MOS value of a deblurring algorithm is high, then we consider this algorithm has good deblurring performance. This MOS-based ranking is then regarded as the ground truth of the relative performances. Then we do the same rankings using the predicted scores of different quality metrics. Finally, the metric score-based rankings are compared with MOS-based ranking. For a good quality metric, the metric score-based ranking should be consistent with the MOS-based ranking. Therefore, the performances of different quality metrics, when used for benchmarking image deblurring algorithms, can be easily determined by checking the number of consistent rankings. Table 6 summarizes the experimental results on the eight image defocus deblurring algorithms that are formerly used to build the DDID database. In the table, we include three best-performing image blur metrics and six general-purpose NR image quality metrics for comparison.

It is observed from Table 6 that the proposed quality metric produces the most consistent rankings. Specifically, for the eight considered deblurring algorithms, the proposed metric produces five consistent rankings. By comparison, the nine compared metrics do not perform very well because at most two consistent rankings are achieved, which are significantly fewer than that of the proposed metric. This further demonstrates the superiority of the proposed metric in benchmarking image deblurring algorithms. This is very useful in real-world applications.

IV. Conclusion

In this paper, we have presented a NR metric for measuring the quality of defocus deblurred images based on natural scene statistics. Two categories of NSS features have been extracted in both the spatial and frequency domains to characterize the loss of naturalness in the deblurred images. The performance of the proposed metric has been tested in a subjectively rated defocus deblurred image database. Compared to the existing relevant image quality metrics, the proposed

metric has been shown to be able to produce quality scores that are more consistent with human perception. It has also been used to benchmark image defocus deblurring algorithms with very promising results.

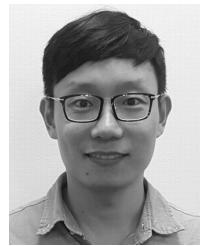
ACKNOWLEDGEMENT

The authors would like to thank Dr. Haoyi Liang from University of Virginia for providing their codes in [17] to conduct a comparison in Table 4.

REFERENCES

- [1] W. Lin and C.-C. J. Kuo, "Perceptual visual quality metrics: A survey," *J. Vis. Commun. Image Represent.*, vol. 22, no. 4, pp. 297–312, 2011.
- [2] A. Immonen, P. Pääkkönen, and E. Ovaska, "Evaluating the quality of social media data in big data architecture," *IEEE Access*, vol. 3, pp. 2028–2043, 2015.
- [3] D. L. Liu, F. Z. Li, and H. B. Song, "Image quality assessment using regularity of color distribution," *IEEE Access*, vol. 4, pp. 4478–4483, 2016.
- [4] H. Z. Nafchi, A. Shahkolaei, R. Hedjam, and M. Cheriet, "Mean deviation similarity index: Efficient and reliable full-reference image quality evaluator," *IEEE Access*, vol. 4, pp. 5579–5590, 2016.
- [5] Z. Wang, "Objective image quality assessment: Facing the real-world challenges," *Proc. IS&T Int. Sym. Electronic Imaging, Image Quality and System Performance XIII*, pp. 1–6, 2016.
- [6] Z. Pan, Y. Zhang, and S. Kwong, "Efficient motion and disparity estimation optimization for low complexity multiview video coding," *IEEE Trans. Broadcast.*, vol. 61, no. 2, pp. 166–176, Jun. 2015.
- [7] Z. Pan, J. Lei, Y. Zhang, X. Sun, and S. Kwong, "Fast motion estimation based on content property for low-complexity H.265/HEVC encoder," *IEEE Trans. Broadcast.*, vol. 62, no. 3, pp. 675–684, Sep. 2016.
- [8] J. Li, X. Li, B. Yang, and X. Sun, "Segmentation-based image copy-move forgery detection scheme," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 3, pp. 507–518, Mar. 2015.
- [9] D. M. Chandler, "Seven challenges in image quality assessment: Past, present, and future research," *ISRN Signal Process.*, vol. 2013, Art. no. 905685, pp. 1–53, Jan. 2013.
- [10] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [11] H. Yeganeh, M. Rostami, and Z. Wang, "Objective quality assessment for image super-resolution: A natural scene statistics approach," in *Proc. IEEE Int. Conf. Image Process.*, Sep./Oct. 2012, pp. 1481–1484.
- [12] X. Zhu and P. Milanfar, "Automatic parameter selection for denoising algorithms using a no-reference measure of image content," *IEEE Trans. Image Process.*, vol. 19, no. 12, pp. 3116–3132, Dec. 2010.
- [13] X. Kong, K. Li, Q. Yang, W. Liu, and M.-H. Yang, "A new image quality metric for image auto-denoising," in *Proc. Int. Conf. Comput. Vis.*, 2013, pp. 2888–2895.
- [14] K. Zeng and Z. Wang, "Perceptual quality assessment of denoised images," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2015, pp. 1–5.
- [15] K. Ma, W. Liu, and Z. Wang, "Perceptual evaluation of single image dehazing algorithms," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2015, pp. 3600–3604.
- [16] L. Li, Y. Zhou, W. Lin, J. Wu, X. Zhang, and B. Chen, "No-reference quality assessment of deblocked images," *Neurocomputing*, vol. 177, pp. 572–584, Feb. 2016.
- [17] H. Liang and D. S. Weller, "Comparison-based image quality assessment for selecting image restoration parameters," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5118–5130, Nov. 2016.
- [18] W.-S. Lai, J.-B. Huang, Z. Hu, N. Ahuja, and M.-H. Yang, "A comparative study for single image blind deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1701–1709.
- [19] T. Hou, S. Wang, and H. Qin, "Image deconvolution with multi-stage convex relaxation and its perceptual evaluation," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3383–3392, Dec. 2011.
- [20] Y. Liu, J. Wang, S. Cho, A. Finkelstein, and S. Rusinkiewicz, "A no-reference metric for evaluating the quality of motion deblurring," *ACM Trans. Graph.*, vol. 32, no. 6, 2013, Art. no. 175.
- [21] A. V. Nasarov and A. S. Krylov, "Basic edges metrics for image deblurring," in *Proc. 10th Conf. Pattern Recognit. Image Anal., New Inf. Technol.*, 2010, pp. 243–246.
- [22] J. M. Bioucas-Dias, M. A. T. Figueiredo, and J. P. Oliveira, "Total variation-based image deconvolution: A majorization-minimization approach," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2006, pp. 861–864.
- [23] A. Foi, K. Dabov, V. Katkovnik, and K. Egiazarian, "Shape-adaptive DCT for denoising and image reconstruction," *Proc. SPIE*, vol. 6064, p. 60640N, Feb. 2006.
- [24] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *J. Electron. Imag.*, vol. 19, no. 1, 2010, Art. no. 011006.
- [25] Y. Gong and I. F. Sbalzarini, "Image enhancement by gradient distribution specification," in *Proc. Asian Conf. Comput. Vis.*, 2014, pp. 47–62.
- [26] Y. Gong and I. F. Sbalzarini, "Gradient distribution priors for biomedical image processing," *arXiv:1408.3300v2*, pp. 1–28, Sep. 2014.
- [27] J. Wu, W. Lin, G. Shi, Y. Zhang, W. Dong, and Z. Chen, "Visual orientation selectivity based structure description," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4602–4613, Nov. 2015.
- [28] S. Marčelja, "Mathematical description of the responses of simple cortical cells," *J. Opt. Soc. Amer.*, vol. 70, no. 11, pp. 1297–1300, 1980.
- [29] Y. Zhang and D. M. Chandler, "No-reference image quality assessment based on log-derivative statistics of natural scenes," *J. Electron. Imag.*, vol. 22, no. 4, 2013, Art. no. 043025.
- [30] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, 2011.
- [31] L. Li, Y. Yan, Y. Fang, S. Wang, L. Tang, and J. Qian, "Perceptual quality evaluation for image defocus deblurring," *Sig. Proc.: Image Comm.*, vol. 48, pp. 81–91, Oct. 2016.
- [32] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2419–2434, Nov. 2009.
- [33] W. Dong, L. Zhang, G. Shi, and X. Wu, "Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization," *IEEE Trans. Image Process.*, vol. 20, no. 7, pp. 1838–1857, Jul. 2011.
- [34] W. Dong, L. Zhang, and G. Shi, "Centralized sparse representation for image restoration," in *Proc. IEEE Int. Conf. Comput. Vis.*, Nov. 2011, pp. 1259–1266.
- [35] A. Danielyan, V. Katkovnik, and K. Egiazarian, "BM3D frames and variational image deblurring," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1715–1727, Apr. 2012.
- [36] W. Dong, L. Zhang, G. Shi, and X. Li, "Nonlocally centralized sparse representation for image restoration," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1620–1630, Apr. 2013.
- [37] J. Zhang, D. Zhao, R. Xiong, S. Ma, and W. Gao, "Image restoration using joint statistical modeling in a space-transform domain," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 6, pp. 915–928, Jun. 2014.
- [38] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "Perceptual blur and ringing metrics: Application to JPEG2000," *Signal Process., Image Commun.*, vol. 19, no. 2, pp. 163–172, Feb. 2004.
- [39] R. Ferzli and L. J. Karam, "A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB)," *IEEE Trans. Image Process.*, vol. 18, no. 4, pp. 717–728, Apr. 2009.
- [40] N. D. Narvekar and L. J. Karam, "A no-reference image blur metric based on the cumulative probability of blur detection (CPBD)," *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2678–2683, Sep. 2011.
- [41] C. T. Vu, T. D. Phan, and D. M. Chandler, "S₃: A spectral and spatial measure of local perceived sharpness in natural images," *IEEE Trans. Image Process.*, vol. 21, no. 3, pp. 934–945, Mar. 2012.
- [42] P. V. Vu and 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, Jul. 2012.
- [43] R. Hassenz, Z. Wang, and M. M. A. Salama, "Image sharpness assessment based on local phase coherence," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2798–2810, Jul. 2013.
- [44] Q. Sang, H. Qi, X. Wu, C. Li, and A. C. Bovik, "No-reference image blur index based on singular value curve," *J. Vis. Commun. Image Represent.*, vol. 25, no. 7, pp. 1625–1630, 2014.
- [45] K. Gu, G. Zhai, W. Lin, X. Yang, and W. Zhang, "No-reference image sharpness assessment in autoregressive parameter space," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 3218–3231, Oct. 2015.
- [46] L. Li, W. Lin, X. Wang, G. Yang, K. Bahrami, and A. C. Kot, "No-reference image blur assessment based on discrete orthogonal moments," *IEEE Trans. Cybern.*, vol. 46, no. 1, pp. 39–50, Jan. 2016.

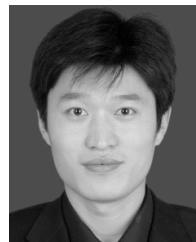
- [47] L. Li, D. Wu, J. Wu, H. Li, W. Lin, and A. C. Kot, "Image sharpness assessment by sparse representation," *IEEE Trans. Multimedia*, vol. 18, no. 6, pp. 1085–1097, Jun. 2016.
- [48] A. K. Moorthy and A. C. Bovik, "A two-step framework for constructing blind image quality indices," *IEEE Signal Process. Lett.*, vol. 17, no. 5, pp. 513–516, May 2010.
- [49] A. K. Moorthy and 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, Dec. 2011.
- [50] M. A. Saad, A. C. Bovik, and 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, Aug. 2012.
- [51] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.
- [52] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, 2013.
- [53] P. Ye, J. Kumar, L. Kang, and D. Doermann, "Unsupervised feature learning framework for no-reference image quality assessment," in *Proc. Int. Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 1098–1105.
- [54] K. Gu, G. Zhai, X. Yang, and W. Zhang, "Using free energy principle for blind image quality assessment," *IEEE Trans. Multimedia*, vol. 17, no. 1, pp. 50–63, Jan. 2015.



LEIDA LI received the B.S. and Ph.D. degrees from Xidian University, Xi'an, China, in 2004 and 2009, respectively. In 2008, he was a Visiting Ph.D. Student with the Department of Electronic Engineering, National Kaohsiung University of Applied Sciences, Taiwan. From 2014 to 2015, he was a Visiting Research Fellow with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, where he is currently a Senior Research Fellow with the School of Electrical and Electronic Engineering. He is also a Full Professor with the School of Information and Control Engineering, China University of Mining and Technology, China. His research interests include multimedia quality assessment, information hiding, and image forensics.



YA YAN received the B.S. degree from Hubei Engineering University, China, in 2014. She is currently pursuing the M.Eng. degree with the School of Information and Control Engineering, China University of Mining and Technology, China. Her research interests include image deblurring, natural scene statistics, and image quality assessment.



assessment.

ZHAOLIN LU received the M.S. degree from Xidian University in 2006 and the Ph.D. degree from the China University of Mining and Technology, China, in 2012. From 2006 to 2009, he was an Engineer with the Research Institute of China North Industries Group Corporation. He is currently an Associate Professor with the Advanced Analysis and Computation Center, China University of Mining and Technology. His research interests include image restoration and image quality



JINJIAN WU received the B.S. and Ph.D. degrees from Xidian University, Xi'an, China, in 2008 and 2013, respectively. From 2011 to 2013, he was a Research Assistant with Nanyang Technological University, Singapore. From 2013 to 2014, he was a Post-Doctoral Research Fellow with Nanyang Technological University. From 2013 to 2015, he was a Lecturer with Xidian University. Since 2015, he has been an Associate Professor with the School of Electronic Engineering, Xidian University. His research interests include visual perceptual modeling, saliency estimation, quality evaluation, and just noticeable difference estimation. He has served as a TPC member of ICME2014, ICME2015, PCM2015, and ICIP2015. He received the best student paper of ISCAS 2013.



KE GU received the B.S. and Ph.D. degrees in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2009 and 2015, respectively. In 2014, he was a Visiting Student with the Department of Electrical and Computer Engineering, University of Waterloo, Canada. He is currently a Research Fellow with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. His research interests include quality assessment and contrast enhancement.



SHIQI WANG received the B.S. degree in computer science from the Harbin Institute of Technology in 2008 and the Ph.D. degree in computer application technology from the Peking University in 2014. He was a Post-Doctoral Fellow with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Canada. He is currently with the Rapid-Rich Object Search Laboratory, Nanyang Technological University, Singapore, as a Research Fellow. He has authored and co-authored over 80 technical articles in refereed journals and proceedings in the areas of image and image/video coding, processing, and quality assessment and analysis. He has also proposed over 30 technical proposals to ISO/MPEG, ITU-T, and AVS video coding standards.