

Blind Quality Index for Camera Images with Natural Scene Statistics and Patch-based Sharpness Assessment

Lijuan Tang^{†§}, Leida Li[†], Ke Gu[‡], Xingming Sun[#], and Jianying Zhang^{†*}

[†]*School of Info. and Elec. Eng., China Univ. of Mining and Technology, China*

[§]*School of Info. and Elec. Eng., Jiangsu Vocational College of Business, China*

[‡]*School of Computer Science and Engineering, Nanyang Technological University, Singapore*

[#]*School of Compu. and Soft., Nanjing Univ. of Information Science and Technology, China*

*Corresponding author: zhangjianyingqia@163.com

Abstract

The current image quality metrics work on the assumption that an image contains single and simulated distortions which are not representative of real camera images. In this paper we address the problem of quality assessment of camera images from two respects, natural scene statistics (NSS) and local sharpness, and associated three types of features. The first type of four features measures the naturalness of an image, inspired by a recent finding that there exists high correlation between structural degradation information and free energy entropy on natural scene images and this regulation will be gradually devastated as more distortions are introduced. The second type of four features originates from an observation concerning the NSS that a broad spectrum of statistics of distorted images can be caught by the generalized Gaussian distribution (GGD). Both the two types of features above belong to the NSS-based models, but they come from the considerations of local auto-regression (AR) and global histogram, respectively. The third type of three features focuses on estimating the local sharpness, which works by computing log-energies in discrete wavelet transform domain. Finally our quality

metric is achieved via a SVR-based machine learning tool, and its performance is proved to be statistically better than state-of-the-art competitors on the CID2013 database dedicated to the quality assessment of camera images.

Keywords: Image quality assessment (IQA), camera images, blind/no-reference (NR), natural scene statistics (NSS), local sharpness, free energy theory, structural degradation model

1. Introduction

With the soaring development of mobile devices and network, an enormous amount of images are being presented to users every moment. It is challenging to evaluate and control the quality of digital photographs. At the same time, a supreme effort is still made by camera manufacturers to improve the quality of photography. As thus, it is in an urgent pursuit of finding ways to automatically predict the perceptual quality of camera images.

In the past few years, as for the issue of image quality assessment (IQA), many objective metrics of the ability to faithfully evaluate the quality of distorted images have been developed with applications to compression [1, 2, 3], transmission [4], enhancement [5, 6, 7], tone mapping [8] and image forensics [9, 10, 11]. If the distortion-free image which distorted image can be compared with is available, the metric is called full-reference (FR) IQA [12]. But in most cases only the distorted image is available, this type of IQA models are called no-reference (NR) IQA. Furthermore, according to the requirement of prior knowledge of the images or their distortions, current NR IQA algorithms also can be further classified into

two categories, namely general-purpose metrics and distortion-specific metrics. Typical distortion-specific blind quality measures are devoted to ringing effect [13], blockiness [14], sharpness/blurriness [15, 16, 17, 18, 19, 20, 21, 22, 23], etc. Ferzli *et al.* proposed a blur metric by integrating just noticeable blur into a probability summation model to evaluate the amount of blurriness in distorted images, dubbed as just-noticeable blur metric (JNB) [16]. Inspired by JNB, Narvekar *et al.* pooled the localized probability of blur detection by means of a cumulative probability of blur detection to measure distorted images [17]. In [18], the slope of the magnitude spectrum and the total spatial variation is used to create sharpness map to be used to predict image blurriness. Thereafter, Vu *et al.* [19] used the log-energies in high frequency wavelet subbands to predict the global and local sharpness of distorted image. Very recently, few attempts to estimate camera images estimations have been made. Nuutinen *et al.* [24, 25] tried to search for efficient feature sets for predicting visual quality of real photographs. In [26], the authors proposed an approach by utilizing NSS modeling as well as the consumer-centric, quality aware interpretable features for real consumer-type images quality prediction. They also presented a framework [27] for blind quality consumer content images evaluation.

In recent years, general-purpose NR-IQA metrics have been an active research field [28, 29, 30, 31, 32, 33, 34, 36, 37]. In [28], the authors proposed a two-step framework to evaluate a distorted image based on natural scene statistics (NSS), in which the first step is to estimate the existence of distortion types in the image and the second one is to evaluate the distorted image through each of these distortions. Natural image quality evaluator (NIQE) [29] was devised to predict

the deviations from statistical regularities observed in natural images without any prior knowledge of the images or their distortion types. Also inspired by some underlying statistics, Saad *et al.* [30] used discrete cosine transform coefficients to extract features, and then predicted image quality scores with a simple Bayesian inference approach. Scene statistics of locally normalized luminance coefficients was used by blind/referenceless image spatial quality evaluator (BRISQUE) [31] to measure possible losses of “naturalness” in the image referable to the presence of distortions and provide an overall quality measure of the distorted images.

Although aforementioned metrics perform well on the popular databases such as the LIVE database [38], they do not perform as well on real photographs which are subjected to many different distortion sources and types. Because these image quality metrics are based on the assumption that an image contains single or simulated distortions which are not representative of what one encounters in practical real scenarios [26]. Camera images contain more practical distortions unlike most distortions present in the popular databases.

Compared with the previous works, to the best of our knowledge, this paper is the first to propose to a modular framework for IQA of camera images based on the NSS regulation and local sharpness assessment. And furthermore, the proposed blind quality index for camera images (BQIC) has acquired a substantially high performance, it is the only metric with the correlation coefficient of beyond 0.8 in both linear and monotonic performance indices.

The paper is structured as follows. In section II, we present the details of the BQIC algorithm. Section III provides performance measures and comparisons of our BQIC and state-of-the-art blind quality metrics on the CID2013 database [39]

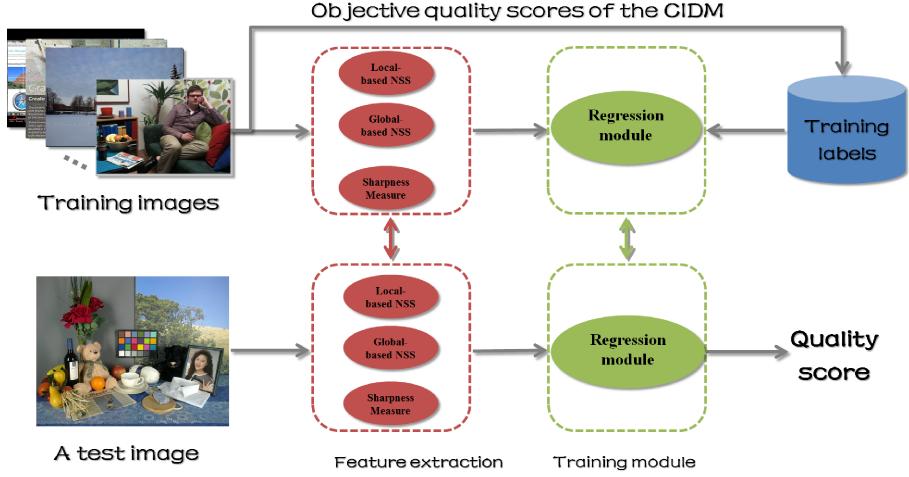


Figure 1: The framework of our proposed blind IQA algorithm.

dedicated to the IQA of camera images. General conclusions and future works are given in Section IV.

2. Proposed Blind Quality Measure

Selecting appropriate features plays an important part in IQA. The features of the proposed metric consists of three parts. The flowchart of the proposed NR IQA metric is outlined in Fig. 1.

The first group of features is extracted based on the free-energy principle, which is recently developed in brain theory and neuroscience [35], and structure degradation measurement [40]. The free-energy principle operates under the assumption that there always exists a difference between an input genuine visual signal and its processed one by human brain. Human perceptual process is manipulated by an internal generative model, which can infer predictions of the input visual signal and avoid the residual uncertainty information. On this basis, the

psychovisual quality of a scene is defined by both the scene itself and the output of the internal generative model. It differs from most traditional methods which are based on signal decomposition.

To facilitate operation, we assume that the internal generative model G for visual perception is parametric, which infers perceived scenes by adjusting the parameter vector \mathbf{x} . Given an input visual signal I , the joint distribution $p(I, \mathbf{x})$ over the space of model parameters \mathbf{x} can compute the “surprise”¹ information (measured by entropy) of the given image. The joint distribution function can be computed as follows:

$$-\log p(I) = -\log \int p(I, \mathbf{x}) d\mathbf{x}. \quad (1)$$

However, the joint distribution $p(I, \mathbf{x})$ is difficult to be directly measured according to current knowledge. As thus, a dummy term $q(\mathbf{x}|I)$ that is an auxiliary posterior distribution of the model parameters given the image is brought into both the numerator and the denominator. So we can rewrite Eq. (1) to be

$$-\log p(I) = -\log \int q(\mathbf{x}|I) \frac{p(I, \mathbf{x})}{q(\mathbf{x}|I)} d\mathbf{x}. \quad (2)$$

Next, Jensen’s inequality is used to apply to Eq. (2), and we have

$$-\log p(I) \leq - \int q(\mathbf{x}|I) \log \frac{p(I, \mathbf{x})}{q(\mathbf{x}|I)} d\mathbf{x} \quad (3)$$

¹The free energy principle works on the assumption that all biological agents resist the natural tendency to disorder in an ever-changing environment. Therefore, it suggests that biological agents can somehow violate the second law of thermodynamic by keeping their internal states at low entropy level to maintain themselves within some physiological bounds. This process is to avoid encountering “surprise” under different environment [35].

and the right side of Eq. (3) is defined as the free energy:

$$J(\mathbf{x}) = - \int q(\mathbf{x}|I) \log \frac{p(I, \mathbf{x})}{q(\mathbf{x}|I)} d\mathbf{x}. \quad (4)$$

Eq. (4) expresses the free energy $J(\mathbf{x})$ to be energy minus entropy. And the free energy estimation of the image I can be expressed by

$$F(I) = J(\hat{\mathbf{x}}) \quad \text{with} \quad \hat{\mathbf{x}} = \arg \min_{\mathbf{x}} J(\mathbf{x}). \quad (5)$$

Any quantitative application of Eq. (5) operates under the assumption that the brain generative model excites. A model with higher expressive power approximates the brain better but incurs higher computational complexity. In this paper we choose the linear AR model as the generative model for its effectiveness and simplicity to describe natural scenes [22, 36]. The AR model is defined as

$$y_n = \chi^k(y_n)\boldsymbol{\theta} + \varepsilon_n \quad (6)$$

where y_n is a pixel of the distorted image, $\chi^k(y_n)$ is a row-vector consisting of k nearest neighbors of y_n , $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_k)^T$ is a vector of AR coefficients, and ε_n is the error term. Then, the predicted version of the input distorted visual signal I can be estimated by $\chi^k(y_n) \cdot \boldsymbol{\theta}_{opt}$, where $\boldsymbol{\theta}_{opt}$ is the optimal estimate of AR parameters for y_n based on the least square method. Consequently, the estimated AR parameters can be used to represent the distribution of the model parameters $q(\mathbf{x}|I)$. In order to demonstrate the distribution of model parameters exhibits a center-peaked appearance, a natural image and its posterior distribution of the model parameters $q(\mathbf{x}|I)$ are shown in Fig.2. According to [41], the total description length of I with the k th-order AR model can be expressed by

$$L(\hat{\mathbf{x}}) = -\log p(I|\hat{\mathbf{x}}) + \frac{k}{2} \log M \quad (7)$$

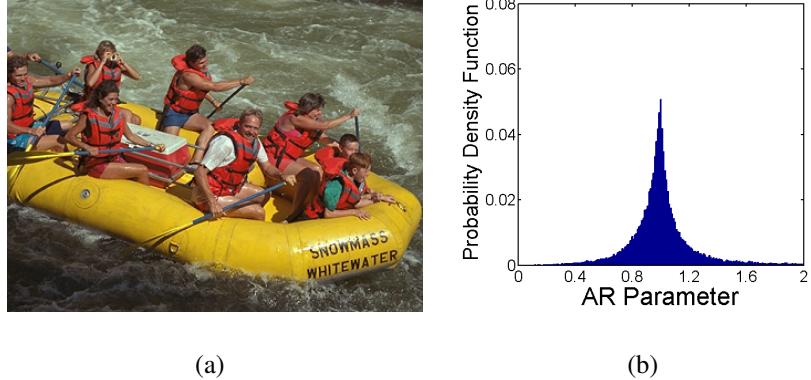


Figure 2: Illustration of the posterior distribution of the model parameters $q(\mathbf{x}|I)$ by: (a) a natural image; (b) the associated distribution of $q(\mathbf{x}|I)$ computed using the AR model.

where M is the number of pixels. And in the large sample limit $M \rightarrow \infty$, the free energy is the total description length:

$$J(\hat{\mathbf{x}}) = -\log p(I|\hat{\mathbf{x}}) + \frac{k}{2} \log M \quad \text{with} \quad M \rightarrow \infty. \quad (8)$$

We choose a fixed-model order, and thus the second term $\frac{k}{2} \log M$ is constant and can be ignored in the quality evaluation.

It is known that low-pass filtered versions of the distorted images have different degrees of spatial frequency decrease. The reduced-reference structural degradation model (SDM) [40] measures the similarity information between original and distorted images. According to the definition of the SSIM [42], μ_I and σ_I as the local mean and variance of I_r with a 2D circularly-symmetric Gaussian weighting function $\mathbf{w} = \{w(k, l)|k = -K, \dots, K, l = -L, \dots, L\}$, which satisfies $\text{sum}(\mathbf{w}) = 1$ and $\text{var}(\mathbf{w}) = 1.5$ (the function of $\text{sum}(\cdot)$ and $\text{var}(\cdot)$ are used to compute sum and variance values respectively). The structural degradation infor-

mation is thus given by

$$S(I) = E\left(\frac{\sigma_{(\mu_I \bar{\mu}_I)} + C}{\sigma_{(\mu_I)} \sigma_{(\bar{\mu}_I)} + C}\right) \quad (9)$$

where $\bar{\mu}_I$ and $\bar{\sigma}_I$ represent the mean intensity and the standard deviation; $\sigma_{(\mu_I \bar{\mu}_I)}$ and $\sigma_{(\sigma_I \bar{\sigma}_I)}$ represent the local covariance the same as the definition in SSIM [42]; $E(\cdot)$ is a direct average pooling; C is a small constant to avoid instability when denominator is very close to zero. Because different sizes of Gaussian weighting functions introduce different amounts information, this paper picks three pairs of (K, L) as $(1, 1)$, $(3, 3)$ and $(5, 5)$. And for noise images of poorer quality, SDM's predictions are quite distinct for subjective scores. We thereby modify $S(I)$ to keep different types of distortions consistent:

$$\tilde{S}(I) = \begin{cases} -S(I) & \text{if } F(I) > T \\ S(I) & \text{otherwise} \end{cases} \quad (10)$$

where the threshold $T = 5$ is chosen empirically to avoid SDM's predicted scores being not consistent with the quality of image for specific distortion types, e.g. noise and blur. We redefine \tilde{S} for different values of (K, L) with $K = L = 1$, $K = L = 3$ and $K = L = 5$. An approximate linear relationship between the structural degradation information and the free energy feature of original image in the LIVE database [38] has been revealed in [43]. In order to further validate this, we randomly selected thirty images from Berkeley database (as shown in Fig. 3) [44] and draw the scatter plots in Fig. 4. The linear dependence feature provides possibility to describe distorted images without primitive image. According to [40], we define the linear regression model as follows:

$$F(I_r) = \alpha \cdot \tilde{S}(I_r) + \beta \quad (11)$$



Figure 3: The selected 30 images from the Berkeley database [44].

where I_r indicates the original image; α and β are gained by the least square algorithm. Their values are tabulated in Table I. We then reduce the dependence of original references by using $FS_I = F(I_d) - (\alpha \cdot \tilde{S}_{I_d} + \beta)$ to be defined as the features, where I_d indicates the distorted image. Note that FS_I values of high-quality images are quite nearly to zero, whereas they will be far from zero when distortions become larger. Finally, we supplement the free energy entropy as the final feature of group one, because it presents good performance on noisy and blurred images.

The second type of features comes from a classical NSS-based model [29]. We

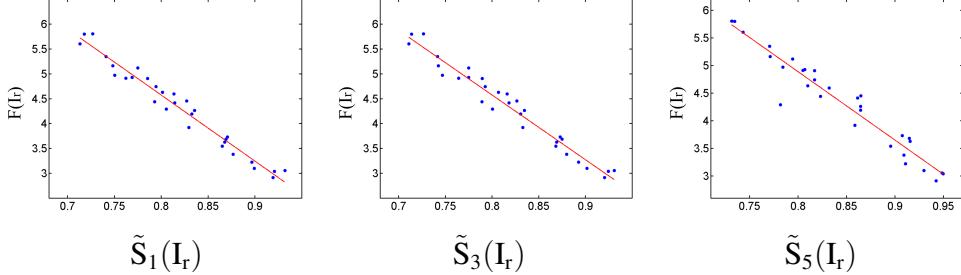


Figure 4: Scatter plots of the structural degradation information $\tilde{S}_s(I_r)$ ($s = \{1, 3, 5\}$) vs. the free energy feature $F(I_r)$ on thirty images in the Berkeley database [44]. The straight lines are fitted with the least square method.

Table 1: The estimates of α, β for \tilde{S} ($s = \{1, 3, 5\}$).

	α	β
\tilde{S}_1	-12.3989	14.8080
\tilde{S}_3	-13.0193	14.9884
\tilde{S}_5	-13.2793	15.1943

can estimate decorrelating effect by exerting a local non-linear operation on log-contrast luminance to remove local mean displacements from zero log-contrast and to normalize the local variance of the log-contrast, as used in some popular blind IQA metrics [29, 30, 31, 36, 37]. It was found that the normalized luminance values of natural images without distortions appears Gaussian characteristic [45], and moreover, the distribution will be broken when the images suffer distortions, as illustrated in Fig. 5.

However, a wider spectrum of statistics of distorted images can be effectively caught by the generalized Gaussian distribution (GGD). The probability density

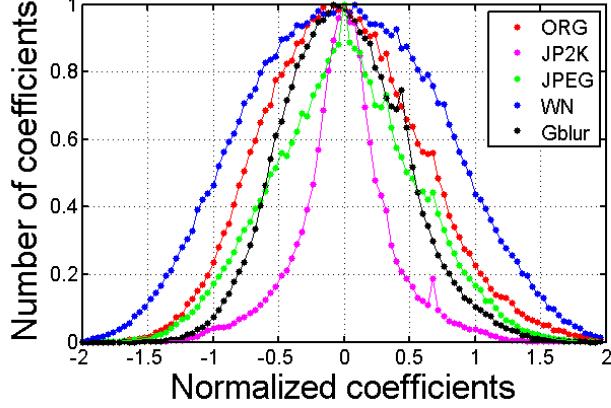


Figure 5: Histogram of normalized coefficients for a natural undistorted image. ORG denotes the original image. JP2K denotes JPEG2000 compression. JPEG denotes JPEG compression. WN denotes additive white Gaussian noise. Gblur denotes Gaussian blur.

function of GGD defined as following :

$$f(x; \mu, \alpha, \beta) = \frac{\alpha}{2\beta\Gamma(\frac{1}{\alpha})} \exp\left(-\left(\frac{|x-\mu|}{\beta}\right)^\alpha\right) \quad \alpha > 0 \quad (12)$$

where μ is the mean, α is the shape parameter that controls the “shape” of the distribution and

$$\beta = \sigma \sqrt{\frac{\Gamma(\frac{1}{\alpha})}{\Gamma(\frac{3}{\alpha})}} \quad (13)$$

and the gamma function $\Gamma(\cdot)$ is given by:

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt \quad (14)$$

where the parameter σ is defined as standard deviation. In this paper, we deploy the GGD with zero mean to fit the mean subtracted contrast normalized (MSC-N), because MSCN is global-based NSS and MSCN coefficient distributions are symmetric [45]. The GGD with zero mean is defined as following:

$$f(x; \alpha, \beta) = \frac{\alpha}{2\beta\Gamma(\frac{1}{\alpha})} \exp\left(-\left(\frac{|x|}{\beta}\right)^\alpha\right). \quad (15)$$

For every image, two pairs of parameters (α, σ^2) from a GGD fit of the MSCN coefficients. We used the method in [46] to estimate parameters (α, σ^2) . One pair is from the original scale, another is at the reduced resolution via a low-pass filtering by a downsampling with the factor of 2. These form the second group of four features which will be used to capture image distortion.

The third group of features is the modified patch-based image sharpness measure [19]. First, using Cohen-Daubechies-Feauveau filters [47] to decompose the input image signal into discrete wavelet transform (DWT) subbands only with one level. Then computing the Log-energy of each subband of discrete wavelet transform (DWT) as follows:

$$E_{XY} = \log_{10}(1 + \frac{1}{M} \sum_{i,j} S_{XY}^2(i, j)) \quad (16)$$

where S_{XY} is either S_{HH} , S_{HL} or S_{LH} , and S_{LL} is not used, and M is the number of DWT coefficients in the subband. The addition of one is used to avoid negative values of E_{XY} . In [19], the authors measured the total log-energy at each level via

$$E_n = (1 - \alpha) \frac{E_{LHn} + E_{HLn}}{2} + \alpha E_{HHn} \quad (17)$$

where the parameter α is 0.8. But according to [48], the authors used predictable sinusoidal, triangular target motions and randomized step-ramp stimuli to compare smooth pursuit in the horizontal and vertical planes. It is confirmed that most normal subjects show higher gain values during horizontal than during vertical tracking. And Grönqvist [49] also observed that vertical tracking inferior to horizontal tracking and the proportion of smooth pursuit increased with age. So horizontal tracking and vertical tracking are asymmetry. Thorough experiments

are performed to give the result that LH, HL and HH appears auto-adaptive non-linear relationship. So we separately consider using E_{LHn} , E_{HLn} and E_{HHn} as features, instead of directly combining them via Eq. (16). Finally, we use the above algorithm in a block-based way to obtain the sharpness index across the entire image. The composition procedure is similar to that in [19]. A collection of local sharpness values are computed using the DWT coefficients corresponding to each 16×16 block of the image, and the index is computed by taking the root mean square of the 1% largest values of the block sharpness indices.

After feature extraction, we need to find a proper way that can map the feature space to subjective MOS, then utilize it to produce objective quality scores. In order to make a fair comparison with other NR IQA methods, we use a support vector regression module SVR [50] to generate a proper mapping that is learnt from the feature score to human visual system. SVR has been widely used in the IQA field [30, 31, 34, 36]. Here the SVR with a radial basis function (RBF) is adopted by using the LIBSVM package [51].

3. Experimental Results

3.1. Experimental Settings

In this section, the CID2013 database [39] is used as testing bed for performance evaluation and comparison. The CID2013 database consists of 6 image sets with 36 scenarios and associated 474 distorted images that are captured by 79 different digital cameras. The images in the CID2013 database don't include so-called reference images since they were taken by real cameras, which makes it impossible to use FR- and RR-IQA methods. As for our training-based BQIC

metric, we use a similar method to that used in [52]. To be specific, the predicted rating for each image was determined by training an SVR on other 473 images via a leave-one-out cross-validation methodology [52]. We test the performance of the proposed blind BQIC metric from three aspects. The first and second are to demonstrate the effectiveness of our BQIC approach compared to state-of-the-art general-purpose and specific-distortion NR-IQA metrics². The last aspects is to analyze and compare the performances of three groups of features used in the BQIC model with each other.

In this paper we follow the video quality experts group (VQEG)'s suggestion and employ a five-parameter nonlinear fitting function to map objective quality scores to subjective human ratings [53]:

$$f(x) = \beta_1 \left(\frac{1}{2} - \frac{1}{1 + e^{\beta_2(x-\beta_3)}} \right) + \beta_4 x + \beta_5 \quad (18)$$

where x denotes the predicted score; $f(x)$ denotes the corresponding subjective score; $\beta_i \{i = 1, 2, 3, 4, 5\}$ are the parameters to be fitted. Next, four commonly used criterions are chosen for performance measure.

The first Spearman's rank ordered correlation coefficient (SRCC) is adopted to evaluate the prediction monotonicity. The second Kendall's rank correlation coefficient (KRCC) is another metric used to evaluate the prediction monotonicity. The third Pearson's (linear) correlation coefficient (PLCC) is adopted to evaluate the prediction accuracy. The last root mean square error (RMSE) is another metric to evaluate the prediction accuracy. A good IQA method is expected to produce

²Only the metrics whose codes are publicly available are used for comparison in our work. All the source codes of testing IQA methods were obtained from their authors or websites.

high SRCC, KRCC and PLCC values, as well as low RMSE values.

3.2. Comparison with General-Purpose NR-IQA Metrics

Recently, several effective general-purpose NR IQA metrics have been proposed to evaluate distorted images without knowing the distortion types. In this section, we will demonstrate and compare the performance of the proposed model with the top general-purpose NR IQA approaches, which are given as follows.

- BLIINDS-II [30], using Bayesian inference model to predict image quality scores based on features extracting from the natural scene statistics (NSS) model of the image discrete cosine transform coefficients (DCT). A total of 24 features were extracted over three scales with 8 features in each scale.
- BRISQUE [31], working based on the principle that natural images have certain regular statistical properties that are measurably modified by distortions. The author used spatial natural scene statistics of locally normalized luminance coefficients to quantify possible losses of naturalness in the distorted image. A total of 36 features were used to identify distortions. We also used those 36 features to test the performance of the BRISQUE on the CID2013.
- SISBLIM [33], inspiring by an early human visual model and free energy based brain theory. The algorithm first predict the noise variance, and apply a way to image denoising based on the above estimated noise level. Then utilizing free energy to measure the joint effect. Finally, the image quality score is acquired by an appropriate integration of estimates of noise, blur, JPEG compression artifacts and the joint effect.
- NFERM [36], constructing a collection of features based on free energy and

Table 2: The PLCC, SRCC, KRCC and RMSE results of our BQIC metric and state-of-the-art general-purpose NR-IQA methods on CID2013. We emphasize the best performed NR IQA algorithm by bold font.

CID2013	<i>PLCC</i>	<i>SRCC</i>	<i>KRCC</i>	<i>RMSE</i>
BLIINDS-II [30]	0.6393	0.6346	0.4539	17.4088
BRISQUE [31]	0.7810	0.7844	0.5902	14.1402
SISBLIM [33]	0.7010	0.6533	0.4762	16.0947
NFERM [36]	0.7933	0.7880	0.5943	13.7833
IL-NIQE [37]	0.4274	0.3065	0.2101	20.4687
BQIC (Proposed)	0.8285	0.8207	0.6291	12.6759

classical human visual system (HVS). A total of 23 features were used to evaluate the image quality.

- IL-NIQE [37], by integrating the features of natural image statistics derived from multiple cues, then a multivariate Gaussian model is created by learning a set of pristine natural images. A Bhattacharyya-like distance is used to measure the quality of each distorted image patch, then an overall quality score is obtained by average pooling.

A logistic nonlinear function is exploited before calculating PLCC and RMSE. Table II summarizes the performance results on the CID2013 database. One can see from Table II that the proposed BQIC metric has gained the highest PLCC, SRCC and KRCC values as well as the smallest RMSE value. None of those compared metrics performs better than the proposed metric, namely the proposed BQIC model correlates highly with human visual perception to image distortions. We also show the scatter plots of the subjective scores versus the predicted scores using different metrics in Fig. 6. A good metric is expected to produce scatter

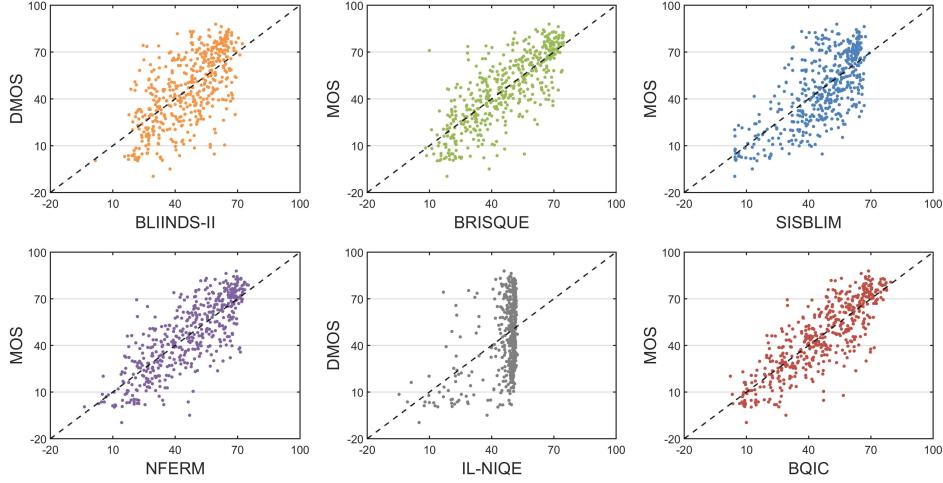


Figure 6: Scatter plots of objective scores generated by BLIINDS-II [30], BRISQUE [31], SISBLIM [33], NFERM [36], IL-NIQE [37] and our proposed BQIC metric versus subjective scores reported by CID2013 databases after nonlinear mapping.

points that are closed to the fitted curve. It can be easily found from Fig. 6 that the proposed metric produces the best fitting results on the CID2013 database.

3.3. Comparison with Specific-Distortion NR-IQA Metrics

Sharpness is one of the most important factors in the problem of camera IQA. In this section, we demonstrate and compare the performance of the proposed model with 5 popular blind sharpness algorithms below.

- BLUR [15], measuring blur based on an analysis of the smoothing or smearing effect of filtering or compression on sharp edges and adjacent regions in an image. The algorithm first find strong vertical edges in the original image or the distorted image, then finding the start and end position of the edge for each corresponding edge in the processed image to calculate local blur. And the global blur is acquired by averaging the local blur values over all edges found.

Table 3: The PLCC, SRCC, KRCC and RMSE results of our BQIC metric and state-of-the-art distortion-specific blind algorithms on CID2013. We highlight the top one.

CID2013	<i>PLCC</i>	<i>SRCC</i>	<i>KRCC</i>	<i>RMSE</i>
BLUR [15]	0.5287	0.5515	0.4063	19.1543
S_3 [18]	0.3277	0.2936	0.2019	21.3902
FISH [19]	0.7038	0.6822	0.4956	16.0827
$FISH_{bb}$ [19]	0.7553	0.7383	0.5461	14.8375
ARISM [22]	0.4877	0.4408	0.3090	19.7651
BQIC (Proposed)	0.8285	0.8207	0.6291	12.6795

- S_3 [18], utilizing a weighted geometric mean to combine the adjusted measures of the slope of the magnitude spectrum and the total spatial variation of the distorted images.
- FISH and $FISH_{bb}$ [19], operating by first decomposing the input image via DWT, then computing the log-energies of the DWT subbands at each level, and finally, a scalar index is obtained by a weighted function of the three-level log-energies. It is also possible to operate the algorithm in a block-based fashion, namely $FISH_{bb}$, acquiring a certain performance gain.
- ARISM [22], calculating the energy and contrast differences in the locally estimated autoregressive coefficients, and then utilizing percentile pooling to acquire the overall score of the image sharpness.

To estimate the performance of the proposed BQIC metric and aforementioned specific-distortion blind algorithms, experiments are conducted on the CID2013 database. Table III lists the performance of the comparison methods and Fig.7 shows the scatter plots between the predicted scores and the corresponding MOSSs,

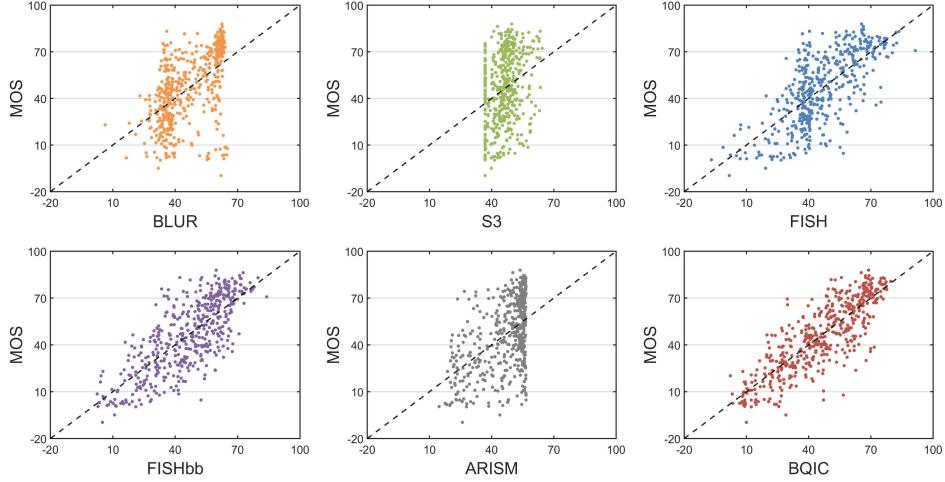


Figure 7: Scatter plots of objective scores generated by BLUR [15], S₃ [18], FISH and FISH_{bb} [19], ARISM [22] and our proposed metric BQIC versus subjective scores reported by CID2013 databases after nonlinear mapping.

where a point denotes one image. It can be clearly observed from both Table III and Fig. 7 that the proposed IQA algorithm outperforms the prevailing sharpness metrics compared in this paper.

3.4. Statistical Significance Comparison

Apart from direct comparisons with general-purpose and distortion-specific metrics, we further evaluate the statistical significance via the F-test [54], which is based on the variance-based hypothesis testing shows additional information regarding the relative performance of different quality algorithms. In order to make a statistically sound judgment regarding superiority of one objective metric against another, the F-test computes the prediction residuals between the converted objective scores and the subjective ratings. Fig. 8 shows the computed F-test results on CID2013. A symbol of ‘1’ indicates that the method performs better

	M01	M02	M03	M04	M05	M06	M07	M08	M09	M10	M11	
M01	—	1	1	1	0	1	1	1	1	0	1	M01:BQIC
M02	-1	—	-1	0	-1	0	1	1	0	-1	1	M02:BLIINDS-II
M03	-1	1	—	0	0	1	1	1	0	0	1	M03:BRISQUE
M04	-1	0	0	—	-1	1	1	1	0	-1	1	M04:SISBLIM
M05	0	1	0	1	—	1	1	1	1	0	1	M05:NFERM
M06	-1	0	-1	-1	-1	—	0	-1	-1	-1	0	M06:IL-NIQE
M07	-1	-1	-1	-1	-1	0	—	0	-1	-1	0	M07:S3
M08	-1	-1	-1	-1	-1	1	0	—	-1	-1	0	M08:BLUR
M09	-1	0	0	0	-1	1	1	1	—	-1	1	M09:FISH
M10	0	1	0	1	0	1	1	1	1	—	1	M10:FISH _{bb}
M11	-1	-1	-1	-1	-1	0	0	0	-1	-1	—	M11:ARISM

Figure 8: Statistical significance comparison between our BQIC and other IQA methods with F-test.

than that of the column, '0' denotes that two objective methods are indistinguishable, and '-1' denotes the method is statistically worse than that of the column. We can find from Fig. 8, where we adopt different colors to label different types of results for readers' conveniences, that our model is statistically superior to most NR/blind models, except statistically indistinguishable from NFERM and FISH_{bb}.

3.5. Analysis of BQIC's Components

Considering that the proposed BQIC is composed of three groups of features, it is necessary to analyze the performance of each group of features. The first group consists of 4 features, the second group consists of 4 features, and the last group includes 3 features. In order to identify how well the features correlate with human judgment of quality, we list the values of PLCC, SRCC, KRCC and RMSE in Table IV and the associated scatter plots in Fig. 9.

We are able to draw important findings from performance comparisons above. Each group of features used in BQIC performs well. For example, the SRCC

Table 4: The PLCC, SROCC , KRCC and RMSE results of the three group features used by BQIC on the whole 474 images in CID2013 databases.

CID2013	<i>PLCC</i>	<i>SROCC</i>	<i>KRCC</i>	<i>RMSE</i>
BQIC1	0.6760	0.6625	0.4756	16.6828
BQIC2	0.6362	0.6380	0.4643	17.4681
BQIC3	0.7509	0.7296	0.5378	14.9512
BQIC	0.8285	0.8207	0.6291	12.6795

of the first group is 0.6625, the second group is 0.6380, and the third group is 0.7296. In comparison, the SROCC of combining the three groups features can reach to 0.8207, performance improvement is larger than 20.1%, 24.4% and 9.4% relative to each group of features. Moreover, it is necessary to point out that three groups of features in BQIC use different strategies. The first group of features is motivated by local AR model, and it is derived from a novel strategy of combining two effective reduced-reference IQA algorithms. The second group is based on global histogram to quantify the possible losses of naturalness in distorted images. And third group focuses on estimating the image sharpness. Thus, the whole three groups of features have even better performance.

3.6. Extension to Assessment of Medical Images

In medical imaging, the image quality is also important. However, the image is commonly noise for some medical imaging. Therefore, accurate medical images quality assessment methods are highly desired which can validly control and monitor the perception quality of medical images. In this paper we further extend our proposed method to the medical images assessment. We conduct two group

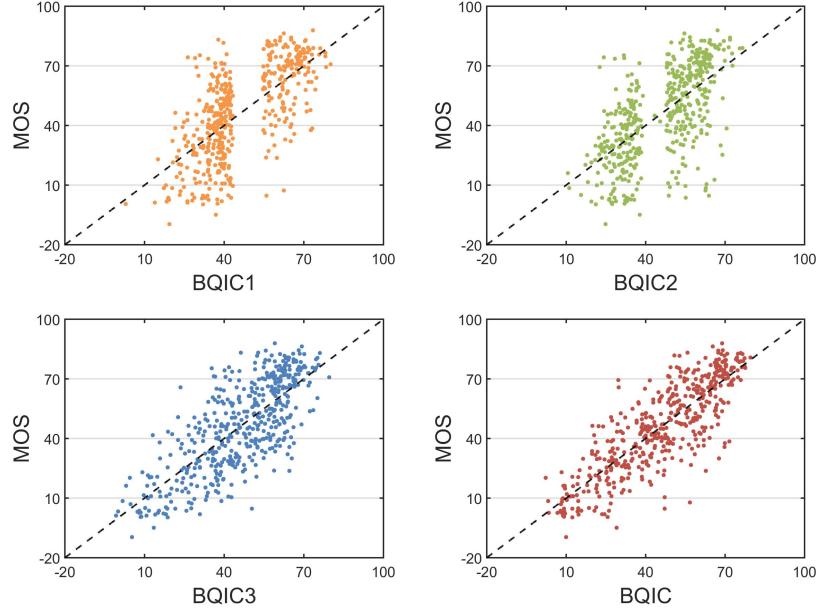
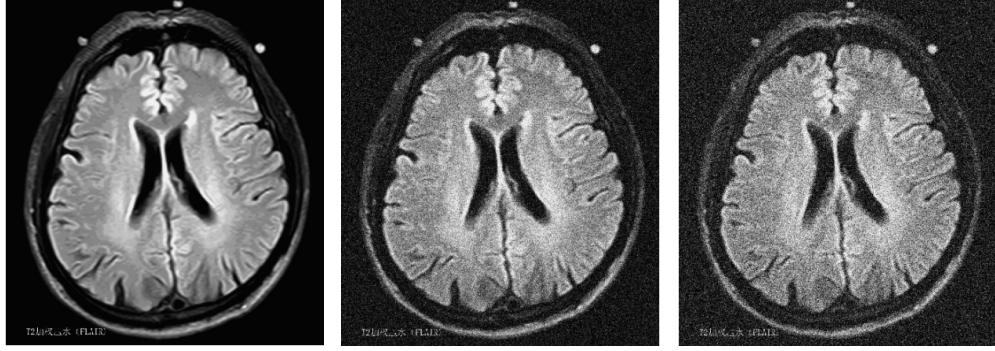


Figure 9: Scatter plots of objective scores generated by the first group features, the second group features, the third group features and all the total features versus subjective scores reported by CID2013 databases after nonlinear mapping.

experiments on the medical images to validate the proposed method.

The first group experiment is that we employ a standard normal pdf of variance σ_i^2 to the original medical images. Then using our method to predict objective quality scores. Fig. 10 shows the objective predicted scores of the original medical image and the corresponding distorted version ($\sigma_1^2 = 0.04, \sigma_2^2 = 0.08$) and Fig. 11 shows the predicted objective predicted scores of the original medical image and the corresponding distorted version ($\sigma_1^2 = 0.02, \sigma_2^2 = 0.06$). It can be obviously observed from the Figs. 10-11 that the proposed method produces the objective scores highly consistent with the HVS.

The second group experiment is that we utilize our method to assess the realistic Anterior Segment Optical Coherence Tomography (AS-OCT) images [55]



(a) Original $S = 54.5050$ (b) $\sigma_1^2 = 0.04 S = 51.3103$ (c) $\sigma_2^2 = 0.08 S = 50.2752$

Figure 10: Three medical images and their objective predicted scores for images with different level noise.



(a) Original $S = 56.3590$ (b) $\sigma_1^2 = 0.02 S = 53.7173$ (c) $\sigma_2^2 = 0.06 S = 52.9580$

Figure 11: Three medical images and their objective predicted scores for images with different level noise.

which were provided by the Department of Ophthalmology in the National University Hospital, Singapore (NUHS). The total number of AS-OCT images used in the test is 209 which can be categorized as good, fair and poor. The number of poor AC-OCT images is 29, and the number of good and fair AC-OCT images is 180. The implement of subjective quality assessment is similar as [55]. Except the images with artifacts resulting from movements of eyelids and corneal scars, we think the images are acceptable for testing anterior segment characteristics. Therefore, we test the accuracy of our proposed method on differentiate the poor

AS-OCT images from the fair and good images. To specify, for every AS-OCT image, we extract 11 features using the proposed BQIC method, then the predicted rating for each AS-OCT image was determined by training an SVR on other 208 images via a leave-one-out cross-validation methodology. We categorize the images into two classes. One includes images marked with “good” and “fair” which we consider them as “feasible”, while the other one includes other images marked with “poor” denoted as “infeasible”. The proposed method achieves 88.04% accuracy for identifying the feasibility of AS-OCT images. It confirms that the proposed algorithm is effective for evaluating of realistic medical images.

4. Conclusion

With the development of networked hand-held devices, a large mount of visual data are presented to users. Many efforts have been made to ensure the end consumers is presented with a satisfactory quality of experience (QoE). Therefore, assessment of camera images is a significant and meaningful topic in scientific research and applicational development of digital image processing. However, it is struggle to handle the images with many concurrent distortion types for current blind quality metrics. Effective objective quality metrics are expected.

In this paper we have put forward a blind quality index for camera images with natural scene statistics and patch-based sharpness assessment. A comparison of our BQIC with state-of-the-art general-purpose NR-IQA methods and popular blind distortion-specific measures is conducted on CID2013 database. The experiment results have proved the superior performance of the proposed blind quality measure on the CID2013 database. Besides the substantially high prediction ac-

curacy, it is worthy to emphasize three points below. First, experimental results prove the superiority of our proposed method on CID2013 over popular NR-IQA models and blind sharpness measures. Second, the proposed BQIC needs merely 11 features, far less than the majority of general-purpose train-based NR-IQA metrics. Third, to the best of our known, our proposed method is the first to propose modular framework for camera images based on nature scene statistics and sharpness assessment.

Acknowledgment

This work is kindly supported in part by National Natural Science Foundation of China (61379143), the Fundamental Research Funds for the Central Universities (2015QNA66), Science and Technology Planning Project of Nantong (BK2014022) and the Qing Lan Project of Jiangsu Province.

5. References

- [1] C. Yan, Y. Zhang, J. Xu, F. Dai, L. Li, Q. Dai, and F. Wu, “A highly parallel framework for HEVC coding unit partitioning tree decision on many-core processors,” *IEEE Signal Process. Lett.*, vol.21, no.5, pp. 573-576, May. 2014.
- [2] C. Yan, Y. Zhang, J. Xu, F. Dai, J. Zhang, Q. Dai and F. Wu, “Efficient parallel framework for HEVC motion estimation on many-core processors,” *IEEE Trans. Circuits Syst. Video Technol.*, vol.24, no.12, pp. 2077-2089, Dec. 2014.

- [3] Z. Q. 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.
- [4] H. R. Wu, A. Reibman, W. Lin, F. Pereira, and S. S. Hemami, “Perceptual visual signal compression and transmission,” *Proc. IEEE*, vol. 101, no. 9, pp. 2025-2043, Sep. 2013.
- [5] K. Gu, G. Zhai, X. Yang, W. Zhang, and C. W. Chen, “Automatic contrast enhancement technology with saliency preservation,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 9, pp. 1480-1494, Sep. 2015.
- [6] S. Wang, K. Gu, S. Ma, W. Lin, X. Liu, and W. Gao, “Guided image contrast enhancement based on retrieved images in cloud,” *IEEE Trans. Multimedia*, vol. 18, no. 2, pp. 219-232, Feb. 2016.
- [7] K. Gu, G. Zhai, W. Lin, and M. Liu, “The analysis of image contrast: From quality assessment to automatic enhancement,” *IEEE Trans. Cybernetics*, vol. 46, no. 1, pp. 284-297, Jan. 2016.
- [8] K. Gu, S. Wang, G. Zhai, S. Ma, X. Yang, W. Lin, W. Zhang, and W. Gao, “Blind quality assessment of tone-mapped images via analysis of information, naturalness and structure,” *IEEE Trans. Multimedia*, vol. 18, no. 3, pp. 432-443, Mar. 2016.
- [9] J. Li, X. L. Li, B. Yang, and X. M. Sun, “Segmentation-based Image Copy-move Forgery Detection Scheme,” *IEEE Trans. Information Forensics and Security*, vol. 10, no. 3, pp. 507-518, Mar. 2015.

- [10] Z. H. Xia, X. H. Wang, X. M. Sun, Q. S. Liu, and N. X. Xiong, “Steganalysis of LSB matching using differences between nonadjacent pixels,” *Multimed. Tools Appl.*, vol. 75, no. 4, pp. 1947-1962, Feb. 2016.
- [11] Z. H. Xia, X. H. Wang, X. M. Sun, and B. W. Wang, “Steganalysis of least significant bit matching using multi-order differences,” *Secur. Commun. Netw.*, vol. 7, no. 8, pp. 1283-1291, Aug. 2014.
- [12] W. Lin and C.-C. Jay Kuo, “Perceptual visual quality metrics: A survey,” *J. Vis. Commun. Image Represent.*, vol. 22, no. 4, pp. 297-312, May 2011.
- [13] H. Liu, N. Klomp, and I. Heynderickx, “A no-reference metric for perceived ringing artifacts in images,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 4, pp. 529-539, Apr. 2010.
- [14] S. A. Golestaneh and D. M. Chandler, “No-reference quality assessment of JPEG images via a quality relevance map,” *IEEE Sig. Process. Lett.*, vol. 21, no. 2, pp. 155-158, Feb. 2014.
- [15] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, “Perceptual blur and ringing metrics: Application to JPEG2000,” *Sig. Process. Image Commun.*, vol. 19, no. 2, pp. 163-172, Feb. 2004.
- [16] R. Ferzli and L. J. Karam, “A no-reference objective image sharpnessmetric based on the notion of just noticeable blur (JNB),” *IEEE Trans. Image Process.*, vol. 18, no. 4, pp. 717-728, Apr. 2009.

- [17] 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.
- [18] 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, May 2012.
- [19] P. V. Vu and D. M. Chandler, “A fast wavelet-based algorithm for global and local image sharpness estimation,” *IEEE Sig. Process. Lett.*, vol. 19, no. 7, pp. 423-426, Jul. 2012.
- [20] K. Bahrami and A. C. Kot, “A fast approach for no-reference image sharpness assessment based on maximum local variation (MLV),” *IEEE Sig. Process. Lett.*, vol. 21, no. 6, pp. 751-755, Jun. 2014.
- [21] Q. B. Sang, H. X. Qi, X. J. Wu, C. F. Li, and A. C. Bovik, “No-reference image blur index based on singular value curve (SVC),” *J. Vis. Communun. Image Represent.*, vol. 25, no. 7, pp. 1625-1630, 2014.
- [22] 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.
- [23] 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. Cybernetics.*, vol. 46, no. 1, pp. 39-50, Jan. 2016.

- [24] M. Nutinen, T.Virtanen, and P. Oittinen, “Features for predicting quality of images captured by digital cameras,” in *Proc. IEEE Int. Symp. Multimedia*, pp. 165-168, 2012.
- [25] M. Nutinen, T.Virtanen, and P. Oittinen, “Image feature subsets for predicting the quality of consumer camera images and identifying quality dimensions,” *Journal of Electronic Imaging*, vol. 23, no. 6, Nov. 2014.
- [26] M. A. Saad, P. Corriveau, and R. Jaladi, “Objective consumer device photo quality evaluation,” *IEEE Sig. Process. Lett.*, vol. 22, no. 10, pp. 1516-1520, Oct. 2015.
- [27] M. A. Saad, P. Corriveau, and R. Jaladi, “Consumer content framework for blind photo quality evaluation,” *Ninth International Workshop on Video Processing and Quality Metrics for Consumer Electronics.*, 2015.
- [28] A. K. Moorthy and A. C. Bovik, “A two-step framework for constructing blind image quality indices,” *IEEE Sig. Process. Lett.*, vol. 17, no. 5, pp. 513-516, May 2010.
- [29] A. Mittal, R. Soundararajan, and A. C. Bovik, “Making a completely blind Image quality analyzer,” *IEEE Sig. Process. Lett.*, vol. 20, no. 3, pp. 209-212, Mar. 2013.
- [30] M. A. Saad and A. C. Bovik, “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.

- [31] 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.
- [32] L. Li, H. Cai, Y. Zhang, W. Lin, Alex C. Kot and X. Sun, “Sparse representation based image quality index with adaptive sub-dictionaries”, *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3775-3786, 2016.
- [33] K. Gu, G. Zhai, X. Yang, and W. Zhang, “Hybrid no-reference quality metric for singly and multiply distorted images,” *IEEE Trans. Broadcasting*, vol. 60, no. 3, pp. 555-567, Sep. 2014.
- [34] 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.
- [35] G. Zhai, X. Wu, X. Yang, W. Lin, and W. Zhang, “A psychovisual quality metric in free-energy principle,” *IEEE Trans. Image Process.*, vol. 21, no.1, pp. 41-52, Jan. 2012.
- [36] 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.
- [37] L. Zhang, L. Zhang, and A. C. Bovik, “A feature-enriched completely blind image quality evaluator,” *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2579-2591, Aug. 2015.

- [38] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, “LIVE image quality assessment Database Release 2,” [Online]. Available:<http://live.ece.utexas.edu/research/quality>.
- [39] T. Virtanen, M. Nuutinen, M. Vaaheranoksa, P. Oittinen, and J. Hakkinen, “CID2013: A database for evaluating no-reference image quality assessment algorithms,” *IEEE Trans. Image Process.*, vol. 24, no. 1, Jan. 2015.
- [40] K. Gu, G. Zhai, X. Yang, and W. Zhang, “A new reduced-reference image quality assessment using structural degradation model,” in *Proc. IEEE Int. Symp. Circuits and Syst.*, pp. 1095-1098, May. 2013.
- [41] H. Attias, “A variational bayesian framework for graphical models,” in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 12, pp. 209-215, 2000.
- [42] 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.
- [43] K. Gu, G. Zhai, X. Yang, W. Zhang, and L. Liang, “No-reference image quality assessment metric by combining free energy theory and structural degradation model,” in *Proc. IEEE Int. Conf. Multimedia and Expo.*, pp. 1-6, Jul. 2013.
- [44] D. Martin, C. Fowlkes, D. Tal, and J. Malik, “A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics,” in *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 416-423, 2001.

- [45] D. L. Ruderman, “The statistics of natural images,” *Netw. Comput. Neural Syst.*, vol. 5, no. 4, pp. 517-548, 1994.
- [46] K. Sharifi and A. Leon-Garcia, “Estimation of shape parameter for generalized Gaussian distributions in subband decompositions of video,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, no. 1, pp. 52-56, Feb.1995.
- [47] A. Cohen, I. Daubechies, and J.-C. Feauveau, “Biorthogonal bases of compactly supported wavelets,” *Commun. Pure Appl. Math.*, vol. 45, Jun. 1992.
- [48] K. G. Rottach *et al.*, “Comparison of horizontal, vertical and diagonal smooth pursuit eye movements in normal human subjects,” *Vision Research*, vol. 36, no. 14, pp. 2189-2195, Jul. 1996.
- [49] H. Grönqvist, G. Gredebäck and C. V. Hofsten, “Developmental asymmetries between horizontal and vertical tracking,” *Vision Research*, vol.46, no.11, pp.1754-1761, May 2006.
- [50] B. Schokopf, A. J. Smola, R. C. Williamson, and P. L. Bartlett, “New support vector algorithms,” *Neural Comput.*, vol. 12, no. 5, pp. 1207-1245, 2000.
- [51] C-C. Chang and C-J. Lin, “LIBSVM: a library for support vector machines,” *ACM Trans. Intelligent Symp. Technol.*, vol. 2, no. 3, 2011. Software available at <http://www.csie.ntu.edu.tw/cjlin/libsvm>.
- [52] E. Kee and H. Farid, “A perceptual metric for photo retouching,” *Proceedings of the National Academy of Sciences of the United States of America.*, vol. 108, no. 50, pp. 19907-19912, Dec. 2011.

- [53] VQEG, “Final report from the video quality experts group on the validation of objective models of video quality assessment,” Mar. 2000, <http://www.vqeg.org/>.
- [54] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, “A statistical evaluation of recent full reference image quality assessment algorithms,” *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440-3451, Nov. 2006.
- [55] S. I. Niwas, V. Jakhetiya, W. Lin, C. K. Kwoh, C. C. Sng, M. C. Aquino, K. Victor, and P. T. K. Chew, “Complex wavelet based quality assessment for as-oct images with application to angle closure glaucoma diagnosis,” *Compu. Methods Programs in Biomedicine*, vol. 130, pp. 13-21, 2016.