

No-reference Image Quality Assessment for Photographic Images Based on Robust Statistics

Zhengda Zeng^a, Wenming Yang^{a,*}, Wen Sun^a, Jing-Hao Xue^b, Qingmin Liao^a

^a*Shenzhen Key Lab. of Info. Sci&Tech/Shenzhen Engineering Lab. of IS&DCP,
Department of Electronic Engineering/Graduate School at Shenzhen, Tsinghua University,
China*

^b*Department of Statistical Science, University College London, UK*

Abstract

No-reference image quality assessment (NR-IQA) is developing rapidly, but there lacks of research on exploring robust statistics to improve the prediction accuracy and monotonicity of NR-IQA algorithms, in particular for assessing photographic images captured by different digital cameras where a variety of unknown distortions may happen. Hence this paper proposes a novel robust-statistics-based NR-IQA method (termed RSN) for photographic images. In RSN, we present three types of features based on robust statistics: robust natural scene statistics of multiple components, robust multi-order derivatives, and robust complementary features in the frequency domain. Then support vector regression is applied to predict image quality using the extracted features. Experimental results show that RSN remarkably outperforms state-of-the-art NR-IQA methods on the CID2013 database of photographic images, as well as on the popular LIVE and TID2013 databases.

Keywords: Image quality assessment (IQA), no-reference/blind IQA, camera image, robust statistics, natural scene statistics (NSS).

*Corresponding author.

Email addresses: zzhengda92@gmail.com (Zhengda Zeng), yangelmw@163.com (Wenming Yang), 1213755392@qq.com (Wen Sun), jinghao.xue@ucl.ac.uk (Jing-Hao Xue), liaoqm@sz.tsinghua.edu.cn (Qingmin Liao)

1. Introduction

With the rapid popularity of smart phones, images produced by digital cameras are becoming an increasingly important medium of access to information. However, they are still prone to be distorted during acquisition, compression, transmission, processing and restoration. Distortions impair the image quality and even the understanding of the information contained. Therefore, it is timely to automatically predict the perceptual quality of photographic images shown in 1.

Image quality assessment (IQA) is performed in two ways: subjective and objective. Subjective IQA is accurate to reflect human’s visual perception, but it is time-consuming, cumbersome and often impractical. Hence, it is necessary to develop objective IQA algorithms, which can evaluate image quality automatically and accurately. According to the information availability of pristine reference images, objective IQA methods can be divided into three categories: full-reference (FR), reduced-reference (RR) and no-reference (NR) [1–6]. When reference images are available, the FR-IQA methods, such as SSIM [1], GMSD [7], VSI [8], LLM [9], MvSSIM [10] and PQIC [11], can be applied to directly evaluate the difference between a distorted image and its reference image. For RR-IQA, part of the reference image information is available to calculate the quality scores. In reality, the reference images are often unavailable, hence FR-IQA and RR-IQA are infeasible and NR-IQA should be used instead. The state-of-the-art NR-IQA methods include DIIVINE [12], BLINDS-II [13], BRISQUE [14], CNN [15], HOSA [16], IDEAL [17], method in [18], etc. This paper concentrates on NR-IQA for photographic images.

The aforementioned NR-IQA methods, unfortunately, have been validated only on images affected by artificial distortions, such as those in the LIVE [19] and TID2013 [20] databases. Photographic images often mix a variety of unpredictable distortions for which a single type of features, e.g. sharpness, illumination and colour information, can predict poorly. Hence, some algorithms focusing on feature fusion are proposed for photographic images. In [21], a

hybrid approach is illustrated to utilise the natural scene statistics (NSS) modelling and the quality-aware interpretable features, such as sharpness and noise. The method in [22] combines sharpness measure and compound noise estimation. The BQIC method [23] addresses this problem from two aspects, NSS and
 35 local sharpness.



Figure 1: Examples of photographic images in the CID2013 database. The quality scores (MOS) of the images are 31.3, 47.8 and 87.9 in order.

Despite these promising methods, there is still a lack of research on exploring robust statistics to improve the prediction accuracy and monotonicity of IQA (including NR-IQA) algorithms for photographic images. Robust statistics [24], a valid discipline of statistical to make popular statistics less affected by
 40 outliers and perform better over a wider range of probability distributions, has been widely applied in digital image processing, such as image thresholding [25], image denoising [26], image registration [27] and face recognition [28]. Therefore in this paper, we propose a novel Robust Statistics-based NR-IQA method (termed RSN) for photographic images. In RSN, we propose three types of new
 45 robust statistical features: robust natural scene statistics of multiple components, robust multi-order derivatives, and robust complementary features in the frequency domain.

More specifically, these three types of new robust features are extracted as follows. Firstly, we construct five components of an image (luminance component, two chrominance components, maximum local variation component and
 50 minimum local variation component [29]), and compute the median subtracted contrast normalised coefficients (MDSCN), as the robust counterpart of the

mean subtracted contrast normalised coefficient (MSCN) [14, 30], using some robust summary statistics. From the MDSCN coefficients, the variance, shape, skewness, kurtosis and entropy are extracted as robust NSS features. Secondly, we propose the robust multi-order derivative features, which are composed of the medians and mean absolute deviations (MADs) of the colour image and its first three derivatives. Thirdly, we utilise some complementary features in the frequency domain, which include the robust saliency features that we derive from the statistical property of saliency maps and the sharpness measured by the FISH characteristic [31]. When a feature vector is constituted by the three types of features, support vector regression (SVR) is used to map the feature vector to image quality.

Besides our above-mentioned contributions to feature extraction, we shall use experiments to demonstrate that the proposed robust features and RSN method can achieve remarkable improvement in prediction monotonicity and accuracy on the CID2013 database of photographic images, as well as on the popular LIVE and TID2013 databases.

The rest of the paper is organized as follows. Section 2 presents the details of the proposed camera image quality assessment algorithm. In Section 3, the experimental details and results are presented, and the effectiveness of our algorithm is validated. Finally, some conclusions and future work are presented in Section 4.

2. The proposed method

The framework of our new RSN method is shown in Fig.2, mainly including two parts: feature extraction and regression-model learning. In [21–23], when various features are fused, the IQA performance is largely improved, especially for photographic images affected by multiple unknown distortions. In this paper we shall show that using our proposed robust NSS features, robust multi-order derivative features and robust complementary features can offer even much further improvement.

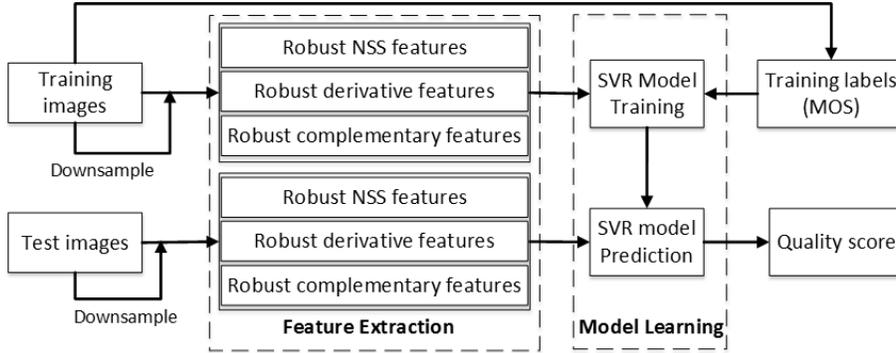


Figure 2: The framework of the proposed robust statistics-based no-reference IQA method for photographic images (RSN).

2.1. Robust NSS features

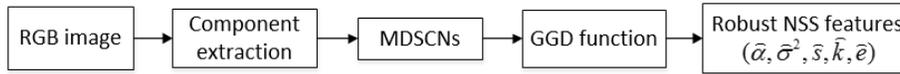


Figure 3: The process of extracting robust NSS features.

In this subsection, motivation and implementation of our robust NSS features are provided. There is a well-known hypothesis about NSS [14, 32] that natural images possess certain statistical properties that can be modified by the presence of distortions. The NSS approach has been successfully applied in NR-IQA through several image domains, such as the wavelet domain in DIVINE [12], the DCT domain in BLINDS-II [13] and the spatial domain in BRISQUE [14] and IDEAL [17]. Our robust NSS features are based on the NSS in the spatial domain because of the pursuit of lower computational complexity and better performance. Recently, the characteristics of chrominance components have been gradually exploited to complement the features in luminance components. The IDEAL [17] proposes four statistic properties from colour representations: hue, saturation, opponent angle and spherical angle. Hence, we attempt to exploit the statistical properties of luminance components I_L along with two different chrominance components (I_M, I_N) in the LMN colour

space [33]. Also, in MLV [29], the distribution of maximum local variation component I_{MaxL} can be parameterised with the generalised Gaussian distribution (GGD), which is easy to extend to minimum local variation component I_{MinL} . That is, we extract the robust NSS features on multiple components, the diagram of which is shown in Fig.3.

Specifically, given an RGB image of size $H \times W$, we first convert it into the LMN colour space [33], obtaining a luminance component I_L and two chrominance components I_M and I_N .

$$\begin{bmatrix} I_L \\ I_M \\ I_N \end{bmatrix} = \begin{bmatrix} 0.06 & 0.63 & 0.27 \\ 0.30 & 0.04 & -0.35 \\ 0.34 & -0.60 & 0.17 \end{bmatrix} \begin{bmatrix} I_R \\ I_G \\ I_B \end{bmatrix}. \quad (1)$$

Then, inspired by the statistical properties of MLV [29] to describe sharpness, we propose to augment the feature space by adding two new order-statistics-based components: the maximum local component (I_{MaxL}) and minimum local component (I_{MinL}) of the luminance component I_L :

$$I_{MaxL} = \begin{pmatrix} \psi(I_L(1, 1)) & \dots & \psi(I_L(1, W)) \\ \vdots & \ddots & \vdots \\ \psi(I_L(H, 1)) & \dots & \psi(I_L(H, W)) \end{pmatrix}, \quad (2)$$

where each entry $\psi(I_L(x, y))$ of pixel $I(x, y)$ is defined as

$$\psi(I_L(x, y)) = \max_{(i,j) \in N_{(x,y)}} \{|I(i, j) - I(x, y)|\}, \quad (3)$$

in which $N_{(x,y)}$ denotes a 3×3 neighbourhood of $I(x, y)$. The minimum local component I_{MinL} can be similarly obtained.

Then, for each of the five components I_k , we propose the median subtracted contrast normalised (MDSCN) coefficient \hat{I}_k as the new robust counterpart of MSCN:

$$\hat{I}_k(x, y) = \frac{I_k(x, y) - \text{med}_k(x, y)}{\text{RMD}_k(x, y) + C}, \quad (4)$$

where index $k \in \{L, M, N, MaxL, MinL\}$ denotes a component; C is a small constant preventing the denominator from tending to zero; $\text{med}_k(x, y)$ is the

local median, a robust order-statistic for data location; and $\text{RMD}_k(x, y)$ is the local root median square error, a statistic for data variability:

$$\text{RMD}_k(x, y) = \sqrt{\frac{1}{|N(x, y)|} \sum_{(i, j) \in N(x, y)} (I_k(i, j) - \text{med}_k(x, y))^2}, \quad (5)$$

where $|N(x, y)|$ denotes the size of the local image patch 3×3 .

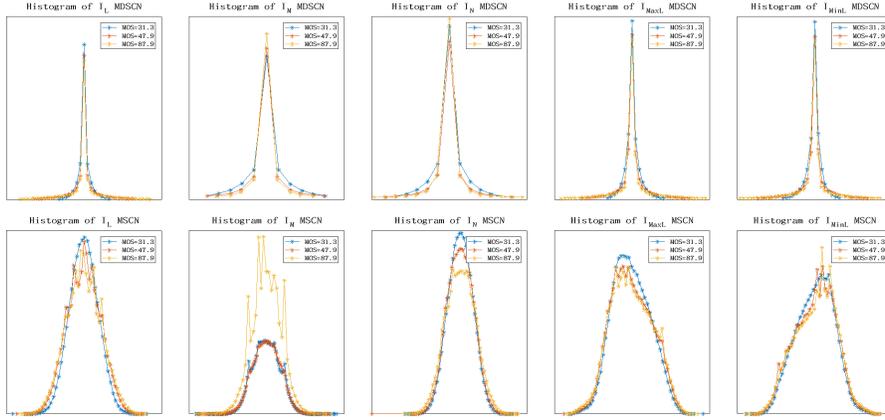


Figure 4: Histograms of the MDSCN and MSCN coefficients for five components of the three photographic images shown in Fig. 1.

120 We analyze histograms of the robust MDSCN and non-robust MSCN coefficients for different components. Three photographic images captured in the same scene are shown in Fig. 1. Their subjective quality scores (MOS) are 31.3, 47.8 and 87.9 in order, in which the higher the subjective score, the better the quality. The histograms of the MDSCN and MSCN coefficients for five components $\{\hat{I}_L, \hat{I}_M, \hat{I}_N, \hat{I}_{MaxL}, \hat{I}_{MinL}\}$ of the three photographic images are shown in Fig. 4. The first row corresponds to MDSCN coefficients (robust version) and the second row corresponds to MSCN coefficients (non-robust version). The blue, red and yellow dotted lines in each histogram correspond to the three images in Fig. 1, respectively. It can be seen that the histograms of MSCN coefficients not only contains many outliers, but also may tilt to one side, such as the distributions for L_{max} and L_{min} components. However, these two problems can be overcome in the robust MDSCN coefficients as illustrated in Fig 4.

125

130

In addition, the histograms of robust MDSCN coefficients for photographic images with different quality scores are more recognizable than the corresponding
135 non-robust version.

It was shown that the MSCN coefficients of a distorted grey-image can be effectively characterised by GGDs [14]. We also use GGDs to model the distributions of the MDSCN coefficients. Therefore, we finally use the parameters of GGDs to summarise the statistical properties of these new coefficients. Let us
140 take the luminance component as example to illustrate our process of feature extraction.

For an MDSCN coefficient $\hat{I}_L(x, y) = z_{\hat{I}_L}$, its probability density function of the GGD with zero mean is

$$f(z_{\hat{I}_L}; \alpha_{\hat{I}_L}, \beta_{\hat{I}_L}) = \frac{\alpha_{\hat{I}_L}}{2\beta_{\hat{I}_L} \Gamma(\frac{1}{\alpha_{\hat{I}_L}})} \exp \left\{ - \left(\frac{|z_{\hat{I}_L}|}{\beta_{\hat{I}_L}} \right)^{\alpha_{\hat{I}_L}} \right\}, \quad (6)$$

where shape parameter $\alpha_{\hat{I}_L}$ and scale parameter $\beta_{\hat{I}_L}$ satisfy $\beta_{\hat{I}_L} = \sigma_{\hat{I}_L} \sqrt{\frac{\Gamma(1/\alpha_{\hat{I}_L})}{\Gamma(3/\alpha_{\hat{I}_L})}}$.
145 Hence, two parameters $(\alpha_{\hat{I}_L}, \sigma_{\hat{I}_L}^2)$ for shape and variance are sufficient to represent the GGD of \hat{I}_L ; they are estimated as with [34] over each image scale.

Besides these two estimated parameters, we also choose three other sample statistics to make a more comprehensive set of robust NSS features from the MDSCN coefficients: $s_{\hat{I}_L}$, the sample skewness; $k_{\hat{I}_L}$, the sample kurtosis; and
150 $e_{\hat{I}_L}$, the entropy. As with [35], they are calculated as

$$d(\hat{I}_L) = \sqrt{\mu(\hat{I}_L - \mu(\hat{I}_L))^2}, \quad (7)$$

$$s_{\hat{I}_L} = \frac{\mu[(\hat{I}_L - \mu(\hat{I}_L))^3]}{d(\hat{I}_L)^3}, \quad (8)$$

$$k_{\hat{I}_L} = \frac{\mu[(\hat{I}_L - \mu(\hat{I}_L))^4]}{d(\hat{I}_L)^4} - 3 \quad (9)$$

and

$$e_{\hat{I}_L} = - \sum_j p_j(\hat{I}_L) \log_2(p_j(\hat{I}_L)), \quad (10)$$

where $\mu(\cdot)$ is the sample mean and $p(\cdot)$ is the histogram.

Hence, five sample estimators of $(\alpha, \sigma^2, s, k, e)$ for each of the five MDSCN components $\{\hat{I}_L, \hat{I}_M, \hat{I}_N, \hat{I}_{MaxL}, \hat{I}_{MinL}\}$ are chosen as the robust NSS features.

2.2. Robust multi-order derivative features

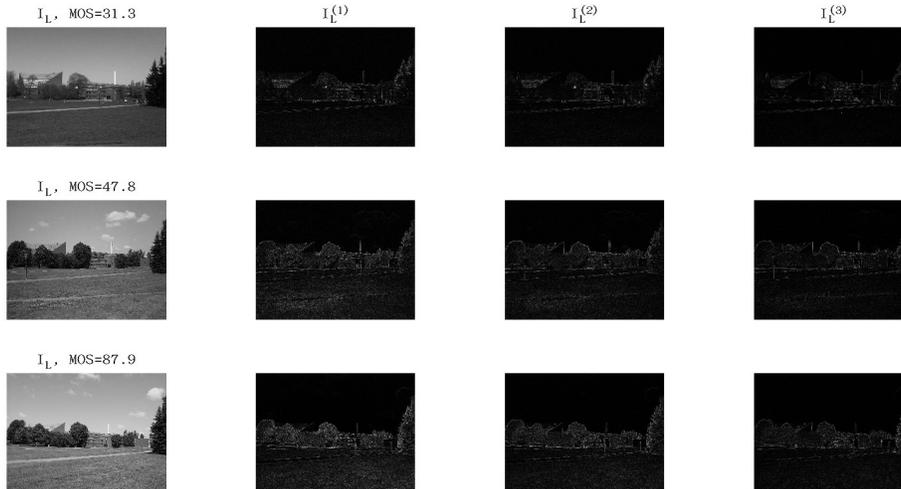


Figure 5: Examples of multi-order derivatives. Different rows correspond to photographic images with different quality scores (31.3, 47.8, 87.9) shown in Fig. 1 and different columns correspond to different order derivatives ($I_L^{(0)}, I_L^{(1)}, I_L^{(2)}, I_L^{(3)}$).

The multi-order derivatives of luminance components (gray images) are shown in Fig. 5. Different rows correspond to photographic images with different quality scores (31.3, 47.8, 87.9) shown in Fig. 1 and different columns correspond to the luminance component and its first three order derivatives ($I_L^{(1)}, I_L^{(2)}, I_L^{(3)}$). We treat I_L as its own zero-order derivative $I_L^{(0)}$. According to [36], the first order derivative $I_L^{(1)}$ is usually related to the slope and elasticity of image surface, the second order derivative $I_L^{(2)}$ is used to capture the curvature, and higher order derivatives can capture detailed discriminative information [37]. For photographic images with different quality scores in the same scene, the characteristics of the same derivative, such as structure, edge, texture are different. Thus, higher order derivatives can capture the details of image structure in the spatial domain, and have been successfully applied to

170 FR-IQA [7] and NR-IQA [37]. In our proposed method, we use the Prewitt
 filter of size 3×3 to calculate the derivatives for its simplicity.

It is well known that, when the histogram is skew or heavy-tailed or when
 there are outliers, the median and mean absolute deviation (MAD) are more
 robust estimators of location and variation than the mean and standard deviation,
 175 respectively [25, 38]. Hence we propose to use the medians and MADs of
 the colour image (I_L, I_M, I_N) and the first three derivatives $(I_L^{(1)}, I_L^{(2)}, I_L^{(3)})$ of
 the luminance component I_L as our robust multi-order derivative features.

Given a grey-image I_L (or say $I_L^{(0)}$), the n th order derivative image $I_L^{(n)}$ is
 calculated from the $(n - 1)$ th order one $I_L^{(n-1)}$:

$$I_L^{(n)} = \sqrt{(I_L^{(n-1)} * h_x)^2 + (I_L^{(n-1)} * h_y)^2}, \quad (11)$$

180 where h_x and h_y are the Prewitt filters along the horizontal and vertical direc-
 tions, respectively.

Then we choose the medians and MADs of the $\{I_L, I_M, I_N, I_L^{(1)}, I_L^{(2)}, I_L^{(3)}\}$
 as the robust multi-order derivative features; MAD is defined as

$$\text{MAD}_k^{(n)} = \frac{1}{|I_k^{(n)}|} \sum_{(i,j) \in I_k^{(n)}} |I_k^{(n)}(i,j) - \text{med}_k^{(n)}|, \quad (12)$$

where index $k \in \{L, M, N\}$ denotes a component, index $n \in \{0, 1, 2, 3\}$ denotes
 185 the derivative order, and $|I_k^{(n)}|$ denotes the size of the component $I_k^{(n)}$.

2.3. Robust complementary features

The above-defined robust NSS and derivative features are extracted from
 the spatial domain. To make the feature space more complete, we propose to
 extract the saliency features and sharpness from the frequency domain as a
 190 complement to the robust spatial features.

Firstly, the saliency features have been successfully applied in the FR-IQA
 algorithm VSI [8]. Considering the assumption that an image's visual saliency
 map has a close relationship with its perceptual quality since both of them dep-
 end on how the human visual system perceives an image [39], we also integrate

195 the saliency features into our algorithm. Specifically, we first choose SDSP [40], a well-performing saliency detection method, to get the saliency map S . Then we calculate the robust-statistics-based median med_S of S as a saliency feature. In addition, we compute the MDSCN \hat{S} of S and model it with the Laplace distribution with zero mean:

$$f(z_{\hat{S}}; \beta_{\hat{S}}) = \frac{1}{2\beta_{\hat{S}}} \exp\left(-\frac{|z_{\hat{S}}|}{\beta_{\hat{S}}}\right), \quad (13)$$

200 where $\beta_{\hat{S}}$ is the scale parameter. We choose the moment estimator of $\beta_{\hat{S}}$ of the \hat{S} map as another saliency feature:

$$\beta_{\hat{S}} = \sqrt{\frac{E\hat{S}^2 - (E\hat{S})^2}{2}}, \quad (14)$$

where $E\hat{S}$ is the first raw moment of \hat{S} and $E\hat{S}^2$ the second raw moment.

Secondly, the sharpness of an image is a useful feature for NR-IQA, because a sharp region is where fine details are resolvable and edges and object boundaries appear to be of high contrast [41]. We choose the global Fast Image Sharpness (FISH) [31] as the sharpness feature in the frequency domain.

2.4. Regression model

As mentioned in the previous sections, we develop three types of features: 25 robust NSS features, 12 robust multi-order derivative features, and three complementary features. Moreover, the past IQA research suggests that adopting some multi-scale information can improve the performance [14]; hence we extract all features over two scales, the original scale and the down-sampled scale by a factor of 2. Overall, we form a feature vector of size 40×2 .

To map from the feature vector to the subjective quality scores (MOS/DMOS), 215 we train an SVR model, as commonly adopted in NR-IQA [14, 17, 21].

3. Experimental results

3.1. Experimental settings

We use the newly released CID2013 [20] database of photographic images to evaluate our method and compare it with existing state-of-the-art NR-IQA

220 algorithms. This database is mainly designed to validate the NR-IQA algorithms
for multi-degraded images without references. It consists of 474 photographic
images obtained by 79 different types of cameras in 8 different shooting scenes.
Moreover, in order to validate the versatility of our method, experiments on
two popular databases, the LIVE database [19] and the TID2013 database [42],
225 are also conducted. The LIVE database contains 779 distorted images with
JPEG2000, JPEG, Gaussian blur, white noise and fast fading, of 29 original
images. The TID2013 database is constructed from 25 reference images and
includes 3000 images with 24 types of distortion at five distortion levels.

Considering the non-linear relationship between the predicted score and the
230 subjective score, logistic functions are usually applied to convert the predicted
scores into the same scale as the subjective scores [19, 20, 43]. In this paper, we
adopt the five-parameter logistic fit function used in [20]:

$$f(\hat{x}) = \beta_1 \left\{ \frac{1}{2} - \frac{1}{1 + \exp(\beta_2(\hat{x} - \beta_3))} \right\} + \beta_4 \hat{x} + \beta_5, \quad (15)$$

where \hat{x} is the predicted score, and β_1 to β_5 are the function parameters to be
determined by minimising the MSE between $f(\hat{x})$ and the true subjective score.

235 In order to evaluate the performance of algorithms, three common evaluation
criteria are used: Spearman’s rank ordered correlation coefficient (SROCC),
Pearson’s linear correlation coefficient (PLCC) and the root mean square er-
ror (RMSE). Among them, SROCC is to assess prediction monotonicity, while
PLCC and RMSE are to assess prediction accuracy.

240 3.2. Comparison on the CID2013 database

We compare our method with some state-of-the-art NR-IQA methods in-
cluding the present best method on the CID2013 database. These methods are:
BRISQUE [14], BLIINDS-II [13], IDEAL [17], RISE [44], the method in [21],
CIQM [45], BQIC [23, 46] and the method in [22]. The experimental setting
245 with leave-one-out cross validation [47] is commonly used in many NR-IQA
method[18, 21, 35]. Since the CID2013 database is divided into six sets based
on different image scenes, we adopted six-fold leave-one-out cross validation [21]:

Table 1: Performance on the CID2013 database. NSRN: the non-robust counterpart of RSN.

	PLCC	SROCC	RMSE
BRISQUE [14]	0.781	0.784	14.140
BLIINDS-II [13]	0.639	0.635	17.409
RISE [44]	0.793	0.769	13.782
Method in [21]	0.799	0.815	
Method in [22]	0.818	0.793	12.948
CIQM [45]	0.817	0.805	12.887
BQIC [23]	0.829	0.821	12.680
IDEAL [17]	0.814	0.817	13.302
NRSN	0.834	0.823	12.043
RSN	0.909	0.890	9.341

we divided the image database into six sets based on image scenes, and used five image sets for training the SVR model and the remaining one for testing. The procedure repeats six times, each time with a different test set, and the mean results of the six-fold evaluation of SROCC, PLCC and RMSE are taken.

The experimental results are listed in Table 1. Our method performs remarkably better than other methods in terms of all three measures, PLCC, SROCC and RMSE, as well as its non-robust counterpart (NRSN). Even compared with the latest method BQIC [23], our PLCC and SROCC are more than 9% and 8% better, respectively. This indicates a better correlation of our method with the human visual perception of the distorted images than other state-of-the-art NR-IQA methods.

It is necessary to further analyse the effectiveness of each type of robust features and the different combinations of them. The experimental results are listed in Table 2. It shows that the robust NSS features (NSS*), that use the robust MDSCN to normalise components, are superior to the NSS features that use the ordinary MSCN. Similarly, the robust multi-order derivative features (Derivatives*), that extract the medians and MADs as features, achieve bet-

Table 2: Analysis of each type of features on the CID2013 database. NSS*: robust NSS features; Derivatives*: robust multi-order derivative features; Complement*: robust complementary features (saliency and sharpness).

features used	PLCC	SROCC	RMSE
NSS	0.718	0.712	15.402
NSS*	0.875	0.866	10.655
Derivatives	0.829	0.812	12.339
Derivatives*	0.860	0.854	10.947
Complement*	0.822	0.803	12.511
NSS*+Derivatives*	0.885	0.877	10.128
NSS*+Complement*	0.877	0.870	10.454
Derivatives*+Complement*	0.873	0.864	10.753
NRSN	0.834	0.823	12.043
RSN	0.909	0.890	9.341

265 ter performance than their ordinary counterparts (Derivatives) that use the mean and standard variation. Furthermore, the performance of combined features (NSS* and Derivatives*) is better than that of separate features (NSS* or Derivatives*). Similar patterns are with other combined features. In the end, when all three types of features are used together, we can obtain the best
270 performance. This indicates that these three types of robust features are complementary to some extent and they are superior to their non-robust counterparts.

Moreover, in order to demonstrate the collective added value of the robust features, we also implement a non-robust version of our RSN, by replacing the involved robust statistics with their ordinary counterparts. We call this non-robust version ‘NRSN’. As shown in Table 1 (or Table 2), NRSN performs worse
275 than RSN, its robust version.

3.3. Comparison on the LIVE and TID2013 databases

To verify the versatility of our approach, we also compare it with some state-of-the-art NR-IQA approaches on the popular LIVE and TID2013 databases.

Table 3: Performance on the LIVE and TID2013 databases.

	LIVE		TID2013	
	PLCC	SROCC	PLCC	SROCC
BRISQUE [14]	0.939	0.940	0.800	0.719
BLIINDS-II [13]	0.930	0.931	0.736	0.644
DIIVINE [12]	0.917	0.916	0.641	0.558
IDEAL [17]	0.946	0.941	0.767	0.719
UNIQUE [48]	0.956	0.952	0.868	0.860
NRSN	0.974	0.971	0.896	0.884
RSN	0.975	0.976	0.906	0.906

280 The approaches chosen include the mainstream BLIINDS-II [13], DIIVINE [12], and BRISQUE [14], as well as the recently published IDEAL [17] and UNIQUE [48] methods. PLCC and SROCC are adopted as evaluation measures, and the experimental settings remain unchanged.

The experimental results in Table 3 show that the proposed method (RSN) 285 is the best in terms of both SROCC and PLCC on the LIVE and TID2013 databases; that is, it is superior to the compared NR-IQA methods and its non-robust version (NRSN) in prediction monotonicity and accuracy.

In summary, our robust statistics-based RSN method performs strongly for both the complex mixed-distortion photographic images in the CID2013 290 database and the simulated-distortion images in the LIVE and TID2013 databases.

3.4. Computational complexity

We now demonstrate that our method does not expend substantial computation. We compare the overall computational complexity of our RSN with other mainstream NR-IQA methods on the CID2013 database, such as DIIVINE [12], 295 BLIINDS-II [13], BRIQUE [14], IDEAL [17], and Method in [22]. In Table 4, we list the time taken (in seconds) to calculate each quality measure on a photographic image of resolution 1200 1400 on a 2.4Ghz Inter Xeon CPU with 64 GB RAM system running Xubuntu 16.04 OS. We use non-optimized MAT-

LAB codes for all of these algorithms to ensure a fair comparison. It can be
 300 observed that the computational time of our method is ranked under the middle
 among the compared methods although it is higher than its non-robust version
 (NRSN).

Table 4: Complexity analysis. NSRN: the non-robust counterpart of RSN.

Algorithm	Time(seconds)
DIIVINE [12]	85.836
BLIINDS-II [13]	400.178
BRIQUE [14]	1.027
IDEAL [17]	5.336
Method in [22]	7.150
NRSN	1.908
RSN	4.485

The computational time of different stages of our algorithm is shown in
 Table 5. We can observe that the time consumption of prediction with the SVR
 305 model is almost negligible, and most computational time is consumed in feature
 extraction. The extractions of three robust features take up different proportions
 of computational time, of which the extraction of robust NSS feature accounts
 for the largest proportion.

Table 5: Percentage of computational time on individual stages. NSS*: robust NSS features;
 Derivatives*: robust multi-order derivative features; Complement*: robust complementary
 features (saliency and sharpness).

Processing stage	Percentage of Delay
NSS*	77.40%
Derivatives*	8.70%
Complement*	13.85%
SVR model	0.05 %

3.5. Discussion

310 In this subsection, we discuss in depth why our algorithm (RSN) can be significantly superior to other NR-IQA algorithms on the CID2013 database.

Firstly, the non-robust version NRSN with the NSS feature, the derivative feature and the complementary feature in the frequency domain lays the foundation. The performance of the NRSN algorithm is slightly better than other
315 algorithms as shown in Table 1.

Secondly, the strategy based on robust statistics greatly improves the performance of non-robust features. Photographic images are impaired with a variety of unknown distortions, which causes that distributions of photographic images are scattered and there are many outliers in the distributions. Thus, the
320 robust statistics are proposed to replace general statistics, which significantly improves the non-robust version (NRSN). Specifically, the MDSCN coefficients are used to replace the MSCN coefficients in the extraction of the NSS feature and the complementary feature. The median and MAD are used to replace the mean and variance in the extraction of derivative feature, respectively. The experiment shown in Table 2 has demonstrated that any single robust statistical
325 feature is significantly superior to its non-robust statistical feature, as well as the algorithm (RSN) based on robust statistics is significantly superior to its corresponding non-robust version (NRSN).

Finally, in many cases, a single feature is not enough to characterize the
330 quality of photographic images with multiple unknown distortions. Thus, we integrate three features based on robust statistics (robust NSS feature, robust derivative feature and robust complementary feature) in our algorithm. The effectiveness of fusion strategy is proved good in the above section.

4. Conclusion

335 In this paper, we proposed a novel NR-IQA method termed RSN for photographic images based on robust statistics. Three new types of robust features

are constructed: 1) robust NSS features on multiple components by normalising each component to median subtracted contrast normalised (MDSCN) coefficients and incorporating the information of skewness, kurtosis and entropy; 2) robust multi-order derivative features by utilising their medians and mean absolute derivations (MADs); and 3) the complementary characteristics in the frequency domain - robust saliency and sharpness. Experimental results demonstrated that the proposed RSN achieved remarkable improvement in prediction monotonicity and accuracy on the CID2013 database of photographic images, and the combination of the three types of features was more effective than any other combination.

Our future work is to investigate how to combine deep learning with NR-IQA for photographic images. CNN has been used in NR-IQA [15], but there remains no deep-learning algorithm for resolving NR-IQA for photographic images.

Acknowledgment

This work was supported by the Natural Science Foundation of China (No.61471216 and 61771276), the National Key Research and Development Program of China (No.2016YFB0101001 and 2017YFC0112500) and the Special Foundation for the Development of Strategic Emerging Industries of Shenzhen (No.JCYJ20170307153940960 and No.JCYJ20150831192224146).

References

- [1] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, Image quality assessment: from error visibility to structural similarity, *IEEE Transactions on Image Processing* 13 (4) (2004) 600–612.
- [2] L. Li, Y. Zhou, W. Lin, J. Wu, X. Zhang, B. Chen, No-reference quality assessment of deblocked images, *Neurocomputing* 177 (2016) 572 – 584.
- [3] C. Zhang, J. Pan, S. Chen, T. Wang, D. Sun, No reference image quality assessment using sparse feature representation in two dimensions spatial correlation, *Neurocomputing* 173, Part 2 (2016) 462 – 470.

- 365 [4] X. Xie, Y. Zhang, J. Wu, G. Shi, W. Dong, Bag-of-words feature representation for blind image quality assessment with local quantized pattern, *Neurocomputing* 266 (2017) 176 – 187.
- [5] W. Zhou, S. Zhang, T. Pan, L. Yu, W. Qiu, Y. Zhou, T. Luo, Blind 3d
370 image quality assessment based on self-similarity of binocular features, *Neurocomputing* 224 (2017) 128 – 134.
- [6] Q. Li, W. Lin, Y. Fang, Bsd: Blind image quality assessment based on structural degradation, *Neurocomputing* 236 (2017) 93 – 103.
- [7] W. Xue, L. Zhang, X. Mou, A. C. Bovik, Gradient magnitude similarity deviation: a highly efficient perceptual image quality index, *IEEE Transactions on Image Processing* 23 (2) (2014) 684–695.
375
- [8] L. Zhang, Y. Shen, H. Li, VSI: A visual saliency-induced index for perceptual image quality assessment, *IEEE Transactions on Image Processing* 23 (10) (2014) 4270–4281.
- [9] H. Wang, J. Fu, W. Lin, S. Hu, C. C. J. Kuo, L. Zuo, Image quality
380 assessment based on local linear information and distortion-specific compensation, *IEEE Transactions on Image Processing* 26 (2) (2017) 915–926. doi:10.1109/TIP.2016.2639451.
- [10] R. Zhu, F. Zhou, J.-H. Xue, Mvssim: A quality assessment index for hyperspectral images, *Neurocomputing*.
- 385 [11] S. Wang, K. Ma, H. Yeganeh, Z. Wang, W. Lin, A patch-structure representation method for quality assessment of contrast changed images, *IEEE Signal Processing Letters* 22 (12) (2015) 2387–2390.
- [12] A. K. Moorthy, A. C. Bovik, Blind image quality assessment: From natural scene statistics to perceptual quality, *IEEE Transactions on Image
390 Processing* 20 (12) (2011) 3350–3364.

- [13] M. A. Saad, A. C. Bovik, C. Charrier, Blind image quality assessment: A natural scene statistics approach in the DCT domain, *IEEE Transactions on Image Processing* 21 (8) (2012) 3339–3352.
- [14] A. Mittal, A. K. Moorthy, A. C. Bovik, No-reference image quality assessment in the spatial domain, *IEEE Transactions on Image Processing* 21 (12) (2012) 4695–4708.
- [15] L. Kang, P. Ye, Y. Li, D. Doermann, Convolutional neural networks for no-reference image quality assessment, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1733–1740.
- [16] J. Xu, P. Ye, Q. Li, H. Du, Y. Liu, D. Doermann, Blind image quality assessment based on high order statistics aggregation, *IEEE Transactions on Image Processing* 25 (9) (2016) 4444–4457.
- [17] D. Lee, K. N. Plataniotis, Toward a no-reference image quality assessment using statistics of perceptual color descriptors, *IEEE Transactions on Image Processing* 25 (8) (2016) 3875–3889.
- [18] K. Gu, D. Tao, J.-F. Qiao, W. Lin, Learning a no-reference quality assessment model of enhanced images with big data, *IEEE Transactions on Neural Networks and Learning Systems*.
- [19] H. R. Sheikh, M. F. Sabir, A. C. Bovik, A statistical evaluation of recent full reference image quality assessment algorithms, *IEEE Transactions on Image Processing* 15 (11) (2006) 3440–3451.
- [20] T. Virtanen, M. Nuutinen, M. Vaahteranoksa, P. Oittinen, J. Häkkinen, CID2013: A database for evaluating no-reference image quality assessment algorithms, *IEEE Transactions on Image Processing* 24 (1) (2015) 390–402.
- [21] M. A. Saad, P. Corriveau, R. Jaladi, Objective consumer device photo quality evaluation, *IEEE Signal Processing Letters* 22 (10) (2015) 1516–1520.

- [22] Y. Zhu, G. Zhai, K. Gu, Z. Che, No-reference image quality assessment for photographic images of consumer device, in: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2016, pp. 1085–1089.
- [23] L. Tang, L. Li, K. Gu, X. Sun, J. Zhang, Blind quality index for camera images with natural scene statistics and patch-based sharpness assessment, *Journal of Visual Communication and Image Representation* 40 (2016) 335–344.
- [24] P. J. Huber, Robust statistics, in: *International Encyclopedia of Statistical Science*, Springer, 2011, pp. 1248–1251.
- [25] J.-H. Xue, D. M. Titterington, Median-based image thresholding, *Image and Vision Computing* 29 (9) (2011) 631–637.
- [26] K. N. Chaudhury, A. Singer, Non-local patch regression: Robust image denoising in patch space, in: *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, IEEE, 2013, pp. 1345–1349.
- [27] K. Arya, P. Gupta, P. K. Kalra, P. Mitra, Image registration using robust m-estimators, *Pattern Recognition Letters* 28 (15) (2007) 1957–1968.
- [28] M. Yang, L. Zhang, J. Yang, D. Zhang, Regularized robust coding for face recognition, *IEEE Transactions on Image Processing* 22 (5) (2013) 1753–1766. doi:10.1109/TIP.2012.2235849.
- [29] K. Bahrami, A. C. Kot, A fast approach for no-reference image sharpness assessment based on maximum local variation, *IEEE Signal Processing Letters* 21 (6) (2014) 751–755.
- [30] D. L. Ruderman, The statistics of natural images, *Network: computation in neural systems* 5 (4) (1994) 517–548.

- [31] P. V. Vu, D. M. Chandler, A fast wavelet-based algorithm for global and
445 local image sharpness estimation, *IEEE Signal Processing Letters* 19 (7)
(2012) 423–426.
- [32] E. P. Simoncelli, B. A. Olshausen, Natural image statistics and neural
representation, *Annual review of neuroscience* 24 (1) (2001) 1193–1216.
- [33] J.-M. Geusebroek, R. Van den Boomgaard, A. W. M. Smeulders, H. Geerts,
450 Color invariance, *IEEE Transactions on Pattern Analysis and Machine In-*
telligence 23 (12) (2001) 1338–1350.
- [34] K. Sharifi, A. Leon-Garcia, Estimation of shape parameter for generalized
Gaussian distributions in subband decompositions of video, *IEEE Trans-*
actions on Circuits and Systems for Video Technology 5 (1) (1995) 52–56.
- 455 [35] Y. Fang, K. Ma, Z. Wang, W. Lin, Z. Fang, G. Zhai, No-reference quality
assessment of contrast-distorted images based on natural scene statistics,
IEEE Signal Processing Letters 22 (7) (2015) 838–842.
- [36] K. Ghosh, S. Sarkar, K. Bhaumik, Understanding image structure from a
new multi-scale representation of higher order derivative filters, *Image and*
460 *Vision Computing* 25 (8) (2007) 1228–1238.
- [37] Q. Li, W. Lin, Y. Fang, No-reference image quality assessment based on
high order derivatives, in: *Multimedia and Expo (ICME), 2016 IEEE In-*
ternational Conference on, IEEE, 2016, pp. 1–6.
- [38] P. Hall, D. Titterton, J.-H. Xue, Median-based classifiers for high-
465 dimensional data, *Journal of the American Statistical Association* 104 (488)
(2009) 1597–1608.
- [39] U. Engelke, H. Kaprykowsky, H.-J. Zepernick, P. Ndjiki-Nya, Visual atten-
tion in quality assessment, *IEEE Signal Processing Magazine* 28 (6) (2011)
50–59.

- 470 [40] L. Zhang, Z. Gu, H. Li, SDSP: A novel saliency detection method by combining simple priors, in: 2013 IEEE International Conference on Image Processing, IEEE, 2013, pp. 171–175.
- [41] C. T. Vu, D. M. Chandler, S3: a spectral and spatial sharpness measure, in: Advances in Multimedia, 2009. MMEDIA'09. First International Conference on, IEEE, 2009, pp. 37–43.
- 475 [42] N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, et al., Image database TID2013: Peculiarities, results and perspectives, *Signal Processing: Image Communication* 30 (2015) 57–77.
- 480 [43] L. Ma, W. Lin, C. Deng, K. N. Ngan, Image retargeting quality assessment: A study of subjective scores and objective metrics, *IEEE Journal of Selected Topics in Signal Processing* 6 (6) (2012) 626–639.
- [44] L. Li, W. Xia, W. Lin, Y. Fang, S. Wang, No-reference and robust image sharpness evaluation based on multiscale spatial and spectral features, *IEEE Transactions on Multimedia* 19 (5) (2017) 1030–1040.
- 485 [45] L. Tang, L. Li, K. Sun, Z. Xia, K. Gu, J. Qian, An efficient and effective blind camera image quality metric via modeling quaternion wavelet coefficients, *Journal of Visual Communication and Image Representation* 49 (2017) 204–212.
- 490 [46] L. Tang, L. Li, K. Gu, J. Qian, J. Zhang, No-reference quality assessment of camera-captured distortion images, in: Pacific Rim Conference on Multimedia, Springer, 2016, pp. 590–599.
- [47] E. Kee, H. Farid, A perceptual metric for photo retouching, *proceedings of the national academy of sciences* 108 (50) (2011) 19907–19912.
- 495 [48] D. Temel, M. Prabhushankar, G. AlRegib, Unique: Unsupervised image quality estimation, *IEEE Signal Processing Letters* 23 (10) (2016) 1414–1418.



Zhengda Zeng received the B.S. degree in Electronic Information Engineering from Huazhong University of Science and Technology in 2015. He is currently pursuing the M.S. degree with the Department of Electronic Engineering, Tsinghua University. His research interests include image quality assessment and image understanding.



Wenming Yang received his Ph.D. degree in electronic engineering from Zhejiang University in 2006. He is an associate professor in the Department of Electronic Engineering, Graduate School at Shenzhen, Tsinghua University. His research interests include image processing, pattern recognition, computer vision, biometrics, video surveillance and image super-resolution.



Wen Sun received the B.Eng. degree from the Department of Information Science and Engineering, Southeast University in 2013. He is currently working towards the Ph.D. degree in electronic engineering at Tsinghua University. His research interests include image quality assessment and image understanding.



Jing-Hao Xue received the Dr.Eng. degree in signal and information processing from Tsinghua University in 1998 and the Ph.D. degree in statistics from the University of Glasgow in 2008. He is a senior lecturer in the Department of Statistical Science, University College London. His research interests include statistical classification, high-dimensional data analysis, pattern recognition and image analysis.



Qingmin Liao received his Ph.D. degree in signal processing and telecommunications from the University of Rennes 1, France, in 1994. He is a Professor in the Department of Electronic Engineering of Tsinghua University, in 2002. Since 2010, he has been the Director of the Division of Information Science and Technology in the Graduate School at Shenzhen, Tsinghua University. His research interests include image/video processing, transmission, analysis, biometrics and their applications.