Optical Engineering

SPIEDigitalLibrary.org/oe

No-training, no-reference image quality index using perceptual features

Chaofeng Li
Yiwen Ju
Alan C. Bovik
Xiaojun Wu
Qingbing Sang

SPIE

No-training, no-reference image quality index using perceptual features

Chaofeng Li
Jiangnan University
School of Internet of Things Engineering
Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education)
Wuxi 214122, China
and
Graduate University of Chinese Academy of Sciences
College of Earth Science
Laboratory of Computational Geodynamics
Beijing 100049, China
E-mail: wxlichaofeng@126.com

Yiwen Ju
Graduate University of Chinese Academy of Sciences
College of Earth Science
Laboratory of Computational Geodynamics
Beijing 100049, China

Alan C. Bovik
University of Texas at Austin
Laboratory for Image and Video Engineering
Austin, Texas 78712

Xiaojun Wu
Qingbing Sang
Jiangnan University
School of Internet of Things Engineering
Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education)
Wuxi 214122, China

Abstract. We propose a universal no-reference (NR) image quality assessment (QA) index that does not require training on human opinion scores. The new index utilizes perceptually relevant image features extracted from the distorted image. These include the mean phase congruency (PC) of the image, the entropy of the phase congruencyPC image, the entropy of the distorted image, and the mean gradient magnitude of the distorted image. Image quality prediction is accomplished by using a simple functional relationship of these features. The experimental results show that the new index accords closely with human subjective judgments of diverse distorted images. [DOI: 10.1117/1.OE.52.5.057003]

Subject terms: image quality assessment; no-reference; phase congruency; entropy; gradient.

Paper 130175 received Feb. 1, 2013; revised manuscript received Apr. 12, 2013; accepted for publication Apr. 16, 2013; published online May 7, 2013.

1 Introduction

Objective no-reference (NR) image quality assessment (IQA) refers to the design of algorithms that seek to judge the quality of distorted images without recourse to comparison with any reference image. Recently, several successful NR QA algorithms have been proposed. A new two-step framework for NR IQA called blind image quality index (BIQI) based on natural scene statistics (NSS) models was proposed in Ref. 1, then later refined to produce the distortion identification-based image verity and integrity evaluation (DIIVINE) index.[2] The DIIVINE index produces IQA results that accord very closely with human subjective judgments of quality when tested on large IQA databases.[2] The authors of Ref. 3 proposed a single-stage NR IQA algorithm called the BLind Image Integrity Notator using Discrete Cosine Transform Statistics (BLIINDS)-II index, which introduces a generalized parametric model of the natural statistics of local image discrete cosine transform (DCT) coefficients to predict image quality scores. Like DIIVINE, BLIINDS-II also produces predictions that compete very well with even full-reference (FR) IQA indices. The authors of Ref. 4 proposed a neural network–based NR IQA algorithm that was trained on perceptually relevant features including the mean phase congruency (PC) image, the entropy of the phase congruency PC image, the entropy of the distorted image, and the gradient of the distorted image. We use a similar set of features here to produce an IQA algorithm that does not require training. The algorithm in Ref. 5 extracts a set of low-level image features, such as texture statistics, then learns a mapping from these features to subjective image quality scores using a support vector machine (SVM). This algorithm also reports competitive results. A hybrid no-reference IQA model, which is based on a hybrid of curvelet, wavelet, and cosine transforms, was proposed in Ref. 6. It does well on a portion of the laboratory for image & video engineering (LIVE) IQA database.[7]

The above-mentioned IQA algorithms, several of which perform quite well, are all based on machine learning principles, and require training-testing sequences using human opinion scores of distorted images, such as difference mean opinion scores (DMOS). While the training procedures are fair and extensible, the algorithms produced still have a

0091-3286/2013/$25.00 © 2013 SPIE

connection to the database (and distortion types) on which they are trained. It is highly desirable to create algorithms that reduce or eliminate this dependence, since many possible deployments will require that algorithms not have available information of new or varying distortions [e.g., quality assessment (QA) agents operating in dynamic video networks].

Recently, the authors of Ref. 8 propose an unsupervised, training free, NR IQA model using latent quality factors. A newer IQA model, called the natural image quality evaluator (NIQE), was proposed by the authors of Ref. 9 based on the construction of a "quality aware" collection of statistical features based on a simple and successful space domain NSS model. Here we also develop an NR QA algorithm that can also handle multiple distortions without any training on human scores, but using a different complementary set of features. The new algorithm is not based on machine learning, but instead uses a set of perceptually relevant image features combined using a simple functional relationship to predict image quality. Its performance as tested on the LIVE IQA database suggests that it is quite promising.

2 Relevant Perceptual Features

We use several local features to measure complementary aspects of perceptual image content. These are described in the following sections.

2.1 *Phase Congruency*

Phase congruency (PC) measures the degree of coherency of the local frequencies comprising the image, which was developed by Morrone and Owens[10] based on the local energy model. The underlying principle of phase congruencyPC is that perceptually significant image features occur at spatial locations where the important Fourier components are maximally in phase with one another.[10]

Morrone and Owens[10] define the PC is equal to the ratio between the energy and the sum of the Fourier amplitudes.

$$\mathrm{PC}_1(x) = \frac{E(x)}{\sum_n A_n}, \tag{1}$$

where A_n is the amplitude of the n'th Fourier component, and $E(x) = \sqrt{F^2(x) + H^2(x)}$, $F(x)$ is the signal $I(x)$ with its DC component removed, and $H(x)$ is the Hilbert transform of $F(x)$.

Kovesi[11] developed another measure of PC that is easier to compute:

$$\mathrm{PC}_2(x) = \frac{\sum_n W(x)\lfloor A_n(x)\Delta\varphi_n(x) - T\rfloor}{\sum_n A_n(x) + \varepsilon}, \tag{2}$$

where $W(x)$ is a tapered weighting function, $\lfloor\rfloor$ is a floor function that leaves the argument unchanged if non-negative, and zero otherwise. $\Delta\varphi_n(x) = \cos[\varphi_n(x) - \bar{\varphi}(x)] - |\sin[\varphi_n(x) - \bar{\varphi}(x)]|$ is a sensitive measure of phase deviation, where $\phi_n(x)$ is the local phase of the Fourier component at x, and $\bar{\phi}(x)$ is the average phase at x. T is an estimate of the noise level, and ε is a small constant that avoids division by zero. Details on the property of PC can be found in Ref. 11.

PC appears to be perceptually relevant. For example, Fig. 1 shows the "parrots" image containing various degrees of blur and two corresponding PC images, which highlights many image regions that are of structural significance, and containing different degrees of blur. The first PC image was obtained by setting the wave parameter to 3 (the default value provided by the author), while the second is the result of fixing the wave parameter at 8, since a larger wave parameter delivers better PC on heavily distorted images (see Fig. 1). Note that the wave parameter denotes the wavelength of a low-pass filter constructed for log Gabor filters. As shown in Fig. 1(a), when the degree of blur is low, the first PC image is better articulated than the second PC image; when the blur increases, the information content of the first PC image degrades, while the second PC image improves relative to the first [in Fig. 1(c)]. This suggests that proper wave parameter selection yields more informative PC images on blurred images. Matlab code for calculating PC can be found in Ref. 12

PC is a relevant feature for IQA, which has been used for both FR IQA and (training-based) NR IQA.[5,13] Here we use PC as an important perceptual feature to construct a training-free NR IQA model.

2.2 *Image Entropy*

Image entropy measures the available local information content of the image. The sample entropy of the image I is

$$E_I = -\sum_n p(n)\log_2 p(n), \tag{3}$$

where $p(n)$ is the empirical probability of luminance value n. Image distortion leads to information loss, and image entropy appears to vary with the type and severity of distortion. Entropy has been used quite successfully for FR IQA[14] and for reduced-reference (RR) IQ.[15] A simple entropy measure has also been proposed for NR IQA.[16] In the following experiment, the MATLAB function entropy was used to compute entropy.

2.3 *Image Gradient*

The image gradient $\Delta I = [\Delta I_x, \Delta I_y]$ measures the perceptually relevant rate of change of image luminance, which is large when there are significant luminance variations in

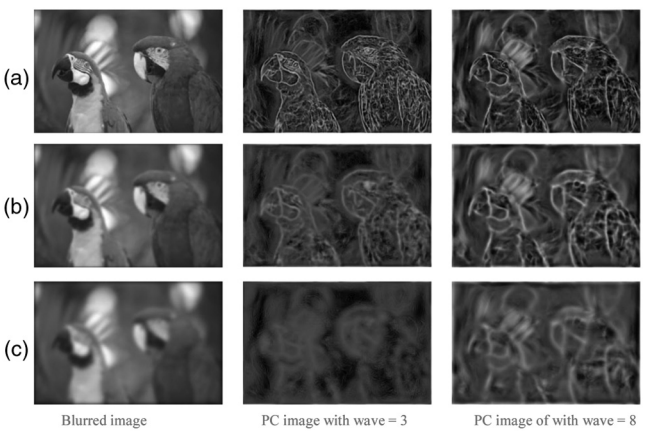

Fig. 1 Image "parrots" of with varying degrees of blur[7] and associated phase congruency maps. (a) Standard deviation of blurred image is 1.3; (b) standard deviation of blurred image is 4.0; (c) standard deviation of blurred image is 7.7.

the image I. A simple and robust measurement of the horizontal and vertical components of the gradient of I is generated by convolving I with the 3×3 Sobel operator masks.

As usual, the gradient magnitude is defined as the square root of the sum of the directional derivative estimates. The image gradient may be viewed as a local measure of image contrast, which has been used as an important ingredient in many IQA algorithms.[17]

3 Definition of the Novel NR IQA Index

Existing successful NR IQA methods use machine learning of DMOS to train the quality prediction models. In some sense, these approaches are not completely "distortion-independent," since they learn how to assess the quality of only the distortions contained in the database. Our goal is to develop a "universal" NR IQA index that can assess the quality of distorted images without relying on DMOS training. We call such an approach "opinion unaware."

We use four perceptually motivated features, which we have used before to construct an NR IQA algorithm that required training,[5] to construct a training-free NR IQA model: (1) the mean value of the phase congruency image (MPC); (2) the entropy of the phase congruency image (EPC); (3) the entropy of the distorted image (EDIS); and (4) the mean value of the gradient magnitude of the distorted image (MGDIS).

We propose three variants of NR QA indices that combine these features in different ways.

$$\mathrm{NRQI1} = \frac{\mathrm{MPC}_1 * \mathrm{EPC}_1}{\mathrm{EDIS}} \quad \mathrm{NRQI2} = \frac{\mathrm{MPC}_2 * \mathrm{EPC}_2}{\mathrm{EDIS} * \mathrm{MGDIS}} \quad (4)$$
$$\mathrm{NRQI} = \mathrm{Max}(\mathrm{NRQI1}, \mathrm{NRQI2} \times \mathrm{Scale}),$$

where MPC_1 and MPC_2 are the MPC *for different wave parameters* and *Scale* is a constant that reflects the proportional coefficient between NRQI1 and NRQI2. When Scale $= 0$, then NRQI $=$ NRQI1, and when Scale is large, NRQI is equivalent to NRQI2.

4 Performance Evaluation

We tested the no-reference quality index (NRQI) indices on the LIVE IQA database, which consists of 29 different reference images and 779 distorted images from five distortion categories—JPEG2000 (JP2K), JPEG, White Noise (WN), Gaussian blur (Gblur) and Fastfading noise (FF)—along with the associated DMOS, which represent human judgments of image quality. In the following experiment, we remove all reference images, leaving 779 distorted images.

The indices used to measure the performance of the algorithm are the Spearman rank ordered correlation coefficient (SROCC) and the linear correlation coefficient (LCC) between the algorithm scores and the DMOS. For the latter measures, the logistic function specified in Ref. 18 was used to fit the model predictions to the subjective data.

4.1 *Testing on LIVE Single Distortion Databases*

First, we test the four perceptual features and the above three NR indices on the LIVE JP2K, JPEG, WN, Gblur, and Fastfading databases individually, with the experimental results given in Tables 1 and 2. When computing NRQI, Scale $= 25$; changing this value by $\pm 20\%$ produces little change in performance.

Table 1 Spearman rank ordered correlation coefficient (SROCC) of perceptual features individually on LIVE distortion databases.

	JP2K	JPEG	WN	GBlur	FF
MPC	0.8430	0.7579	0.9477	0.7976	0.7954
EPC	0.6636	0.6369	0.9217	0.0782	0.3153
EDIS	0.1000	0.7591	0.0305	0.2138	0.0860
MGDIS	0.2054	0.1873	0.8748	0.6364	0.5986

Table 2 Linear correlation coefficient (LCC) of perceptual features individually on LIVE distortion databases.

	JP2K	JPEG	WN	GBlur	FF
MPC	0.8474	0.7654	0.9369	0.7672	0.8357
EPC	0.6657	0.6280	0.9507	0.2421	0.5563
EDIS	0.1607	0.8410	0.5939	0.2366	0.2284
MGDIS	0.2812	0.1607	0.8979	0.6255	0.5811

Table 3 SROCC on LIVE individual distortion databases.

	JP2K	JPEG	WN	GBlur	FF
NRQI1	0.8064	0.8713	0.9357	0.7528	0.7758
NRQI2	0.4689	0.5255	0.9350	0.7642	0.7343
NRQI	0.8074	0.8715	0.9357	0.8562	0.8181
LQF (four topics)[8]	0.84	0.87	0.86	0.84	0.76
NIQE[9]	0.9187	0.9422	0.9718	0.9329	0.8639
GRNN[5]	0.8106	0.9129	0.9769	0.7796	0.7167

From Tables 1 and 2, it can be seen that perceptual feature MPC is quite relevant across all five distortion categories, while the other three features (EPC, EDIS, and MGDIS) are relevant to two or three distortion types.

Since the NRQI NR QA indices do not require any learning process, we compared them with the same blind IQA without human training, the method of Ref. 8 using latent quality factors (LQF), and the newer NIQE,[9] using the LIVE individual distortion databases shown in Tables 3 and 4. The results of the GRNN-based learning method by the same perceptual features[5] is also given (the realigned DMOS were used as in Refs. 1–4, so the results are different from those reported in Ref. 5).

Table 4 LCC on LIVE individual distortion databases.

	JP2K	JPEG	WN	GBlur	FF
NRQI1	0.8112	0.8911	0.9482	0.7190	0.8124
NRQI2	0.5138	0.5845	0.9411	0.8154	0.7673
NRQI	0.8126	0.8911	0.9482	0.8765	0.8435
LQF (four topics)[8]	0.87	0.89	0.87	0.85	0.82
NIQE[9]	0.9264	0.9526	0.9773	0.9447	0.8804
GRNN[5]	0.8233	0.9392	0.9885	0.8145	0.7842

From Tables 3 and 4, it can be seen that the index NRQI1 correlates well with human DMOS on the JP2K, JPEG, and WN distortion categories, and the combination index (NRQI) delivers good performance against all five distortion categories, and outperforms the LQF method[8] on the WN, Gblur, and FF distortion categories, and also better than the GRNN learning method[5] on the Gblur and FF categories, but the results are a litter inferior to NIQE.

Figure 2(a)–2(e) show the scatter plots between the NRQI objective scores and DMOS on the five LIVE individual distortion databases, which also suggests that NRQI index is consistent with human subjective judgments of quality in these distortions.

4.2 *Testing on Entire LIVE IQA Database*

The computation of phase congruencyPC is based on energy. As such, the mean value of phase congruencyPC increases with the amount of noise, but decreases as the other distortions increase, as reflected by the opposite trends in Fig. 2. To make the shape of the scatterplots of the NRQI index consistent over all distortion categories, we adjusted the NRQI for noise distortion as in Eq. (5), which simply reverses the slope and returns the scatterplot to the proper range of values.

$$\mathrm{NRQIwn} = \frac{k}{100 \times (\mathrm{NRQI} + 0.01)}, \tag{5}$$

where k is a constant. In the next experiment, k is set to 1.

In order to make the NRQI and NRQIwn indices applicable to diverse distorted images, we use the subband coefficients from a wavelet transform over three scales and three orientations with Daubechies 9/7 wavelet basis as input features, and a multiclass SVM to classify a given image into noise and other distortion categories (details available in Ref. 1). Thus, we use the first stage ("distortion identification") of the BIQI algorithm. This stage does not utilize human opinion scores. Hence, we regard our algorithm NRQI as "distortion aware." The correlations of NRQI against DMOS on the entire LIVE IQA Database (all distortions) are shown in Table 5.

From Table 5, it can be seen that NRQI outperforms the training-free LQF method,[8] but is inferior to the recent NIQE method.[9] NRQI is a training-free "opinion unaware" method, so it does not attain the performance of the GRNN-based learning (the results of the learning method are obtained using just a few test images, not the entire dataset as used by NRQI, so it is an unfair comparison). Figure 3 shows the scatter plot of objective NRQI scores against DMOS over the entire LIVE IQA database, which also indicates that the NRQI index agrees with human subjective judgments of quality.

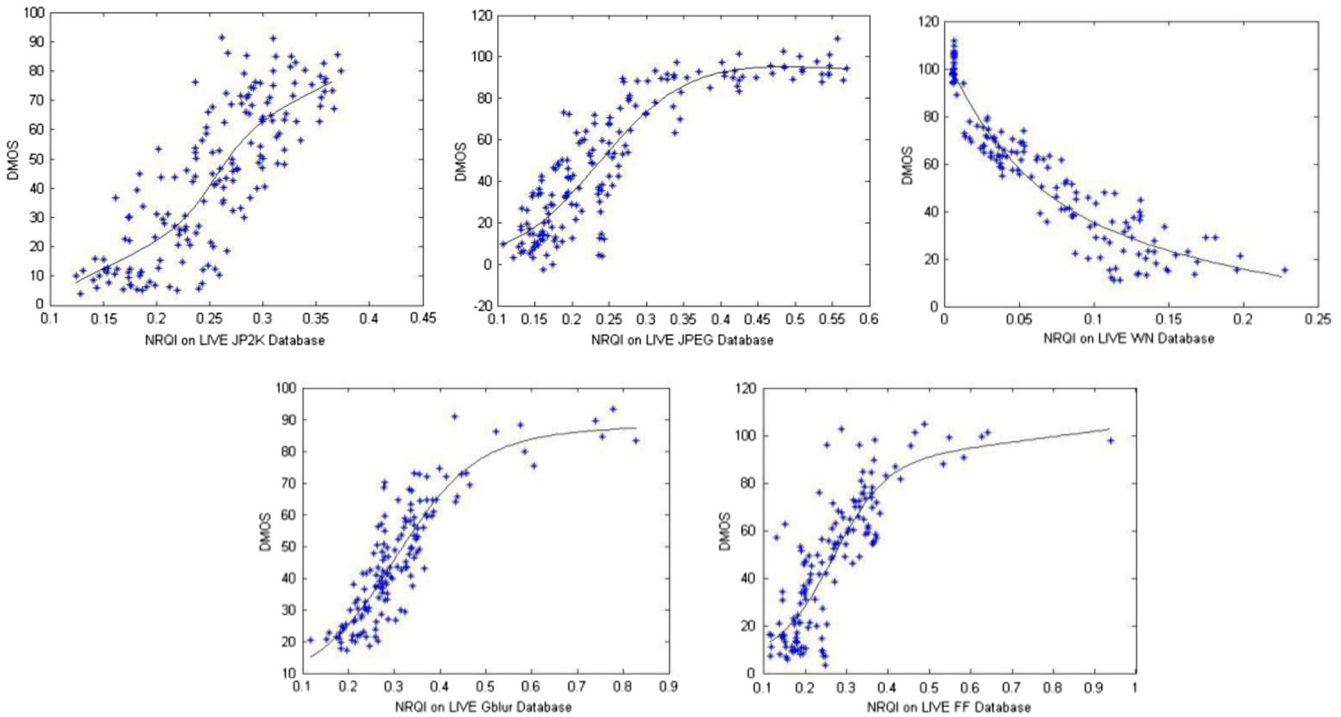

Fig. 2 No-reference quality index (NRQI) against difference mean opinion scores (DMOS) on LIVE individual distortion databases. (a) JP2K; (b) JPEG; (c) WN; (d) Gblur; (e) FF.

Table 5 Image quality assessment (IQA) results on LIVE entire database.

Method	NRQI1	NRQI	LQF[8]	GRNN[5]	NIQE[9]
SROCC	0.6792	0.8263	0.80	0.8651	0.9086
LCC	0.6519	0.8350	0.78	0.8804	0.9069

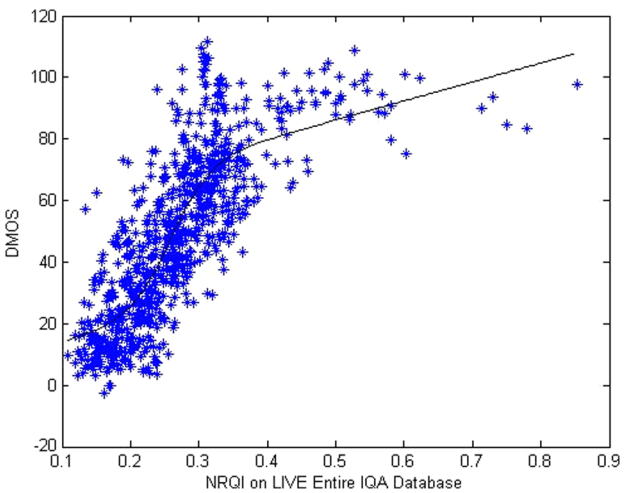

Fig. 3 NRQI against DMOS on entire LIVE image quality assessment (IQA) database.

4.3 *Testing on Part of TID2008 Database*

In order to further evaluate the performance of NRQI, we tested it on the same (available) distortions in an alternate database—the TID2008 (see Ref. 19). The TID database consists of 25 reference images and 1700 distorted images over 17 distortion categories. Of these 25 reference images, only 24 are natural images, and we test our algorithm only on these 24 images. In order to show database independence, the value of k was also set to 1 when computing NRQIwn. The results are shown in Table 6.

From Table 6, it may be seen that NRQI correlates well with many categories of MOS scores, and is competitive with NIQE and the FR PSNR metric over all distortion images and on the JP2K distorted category distortions in particular. However, this method does not yet match NIQE and PSNR on the other single distorted categories.

4.4 *Statistical Significance and Hypothesis Testing Based on DMOS*

In order to ascertain wether the apparent differences observed in image quality metric (IQM) performance are statistically significant, we applied a variance-based hypothesis test (HT) using the residuals between DMOS and the quality predicted by the IQM (after the nonlinear mapping). This was done on the portion of the TID2008 Database interesting with distortions in the LIVE database.

The test we used is based on an assumption of Gaussianity of the residual differences between the IQM prediction and DMOS. It uses the F-statistic to compare the variance of two sets of sample points. The Null Hypothesis is that the residuals from one IQM come from the same distribution and are statistically indistinguishable (with 95% confidence) from the residuals from another IQM. The results of the HT using the residuals between the DMOS and IQM predictions are shown in Table 7.

Table 6 SROCC on part of TID2008 database.

Method	JP2K	JPEG	WN	GBlur	ALL
NRQI	0.8765	0.7490	0.6082	0.7812	0.8002
NIQE	0.8939	0.8756	0.7775	0.8249	0.8006
GRNN	0.6478	0.8528	0.7532	0.6740	0.7625
PSNR	0.8823	0.9215	0.9230	0.9342	0.799

Table 7 Statistical significance matrix based on IQM-DMOS residuals on part of TID2008 database.

Method	NRQI	NIQE	GRNN
NRQI	—	0 −1 −1 −1 0	1 0 1 1 1
NIQE	0 1 1 1 0	—	1 1 1 1 1
GRNN	−1 0 −1 −1 −1	−1 −1 −1 −1 −1	—

Each entry in Table 7 is a codeword consisting of five symbols. The position of the symbol in the codeword represents the following datasets (from left to right): JP2K, JPEG, WN, Gblur, ALL data. Each symbol gives the results of the HT on the dataset represented by the symbol's position: A value of "1" in the table indicates that the row algorithm is statically superior to the column algorithm, whereas a "−1" indicates that the row is statistically worse than the column. A value of "0" indicates that the row and column are statistically indistinguishable (or equivalent).

From Table 7, we can conclude that NRQI is statistically better than the machine learning GRNN method, and that it is also statistically equivalent with NIQE in regard to quality prediction of JP2K distorted images and over ALL types of distortion.

5 Conclusions

We proposed a new training-free NR IQA model based on several complementary and perceptually relevant image features, namely the mean phase congruency of the image (MPC), the entropy of the phase congruency image (EPC), the entropy of the distorted image (EDIS), and the mean gradient magnitude of the distorted image (MGDIS). The MPC is quite relevant across all five distortion categories on LIVE, while the other three features (EPC, EDIS, and MGDIS) are relevant to two or three distortion types. The new index does not need to be trained on any human opinion scores (DMOS). Reasonably accurate image quality estimation is accomplished using a simple functional relationship between those features.

Acknowledgments

This research is supported in part by the National Natural Science Foundation of China (No. 61170120), Program for New Century Excellent Talents in University (NCET-12-0881), the 111 Project (B12018), the National Science and Technology Major Project (2011ZX05039-004), the Natural Science Foundation of Jiangsu Province (No. BK2011147), and China Postdoctoral Science Foundation (No. 2011M500431).

References

1. A. K. Moorthy and A. C. Bovik, "A two-step framework for constructing blind image quality indices," *IEEE Signal Process. Lett.* **17**(5), 513–516 (2010).
2. A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: from scene statistics to perceptual quality," *IEEE Trans. Image Process.* **20**(12), 3350–3364 (2011).
3. M. A. Saad, A. C. Bovik, and C. Charrier, "Model-based blind image quality assessment: a natural scene statistics approach in the DCT domain," *IEEE Trans. Image Process.* **21**(8), 3339–3352 (2012).
4. C. Li, A. C. Bovik, and X. Wu, "Blind image quality assessment using a general regression neural network," *IEEE Trans. Neural Netw.* **22**(5), 793–799 (2011).
5. H. Tang, N. Joshi, and A. Kapoor, "Learning a blind measure of perceptual image quality," in *IEEE Int. Conf. Comput. Vision Pattern Recognition*, Colorado Springs (2011).
6. J. Shen, Q. Li, and G. Erlebacher, "Hybrid no-reference natural image quality assessment of noisy, blurry, JPEG2000, and JPEG images," *IEEE Trans. Image Process.* **20**(8), 2089–2098 (2011).
7. H. R. Sheikh et al., LIVE Image Quality Assessment Database, http://live.ece.utexas.edu/research/quality (2006).
8. A. Mittal, G. S. Muralidhar, and A. C. Bovik, "Blind image quality assessment without human training using latent quality factors," *IEEE Signal Process. Lett.* **19**(2), 75–78 (2012).
9. A. Mittal, G. S. Muralidhar, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.* **20**(3), 209–212 (2013).
10. M. C. Morrone and R. A. Owens, "Feature detection from local energy," *Pattern Recogn. Lett.* **6**(5), 303–313 (1987).
11. P. Kovesi, "Image features from phase congruency," *J. Comput. Vis. Res.* **1**, 1–26 (1999).
12. MATLAB and Octave Functions for Computer Vision and Image Processing, http://www.csse.uwa.edu.au/~pk/Research/MatlabFns/index.html (2006).
13. L. Zhang et al., "FSIM: a feature similarity index for image quality assessment," *IEEE Trans. Image Process.* **20**(8), 2378–2386 (2011).
14. H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.* **15**(2), 430–444 (2006).
15. R. Soundararajan and A. C. Bovik, "RRED indices: reduced reference entropic differencing for image quality assessment," *IEEE Trans. Image Process* **21**(2), 517–526 (2012).
16. S. Gabarda and G. Cristóbal, "Blind image quality assessment through anisotropy," *J. Opt. Soc. Am. A* **24**(12), B42–B51 (2007).
17. C. Li and A. C. Bovik, "Content-partitioned structural similarity index for image quality assessment," *Signal Process. Image Commun.* **25**(7), 517–526 (2010).
18. VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment," http://www.vqeg.org/ (2000).
19. N. Ponomarenko et al., "Tid2008: a database for evaluation of full reference visual quality assessment metrics," *Adv. Mod. Radioelectron.* **10**(1), 30–45 (2009).

Biographies and photographs of the authors are not available.