Making a "Completely Blind" Image Quality Analyzer

Anish Mittal, Rajiv Soundararajan, and Alan C. Bovik, *Fellow, IEEE*

***Abstract*—An important aim of research on the blind image quality assessment (IQA) problem is to devise perceptual models that can predict the quality of distorted images with as little prior knowledge of the images or their distortions as possible. Current state-of-the-art "general purpose" no reference (NR) IQA algorithms require knowledge about anticipated distortions in the form of training examples and corresponding human opinion scores. However we have recently derived a blind IQA model that only makes use of measurable deviations from statistical regularities observed in natural images, without training on human-rated distorted images, and, indeed without any exposure to distorted images. Thus, it is "completely blind." The new IQA model, which we call the Natural Image Quality Evaluator (NIQE) is based on the construction of a "quality aware" collection of statistical features based on a simple and successful space domain natural scene statistic (NSS) model. These features are derived from a corpus of natural, undistorted images. Experimental results show that the new index delivers performance comparable to top performing NR IQA models that require training on large databases of human opinions of distorted images. A software release is available at http://live.ece.utexas.edu/research/quality/niqe_release.zip.**

***Index Terms*—Completely blind, distortion free, image quality assessment, no reference.**

I. Introduction

AMERICANS captured 80 billion digital photographs in 2011 and this number is increasing annually [1]. More than 250 million photographs are being posted daily on facebook. Consumers are drowning in digital visual content and finding ways to review and control of the quality of digital photographs is becoming quite challenging.

At the same time, camera manufacturers continue to provide improvements in photographic quality and resolution. The raw captured images pass through multiple post processing steps in the camera pipeline, each requiring parameter tuning. A problem of great interest is to find ways to automatically evaluate and control the perceptual quality of the visual content as a function of these multiple parameters.

Manuscript received August 26, 2012; revised October 22, 2012; accepted November 04, 2012. Date of publication November 15, 2012; date of current version January 29, 2013. This work was supported by Intel and Cisco under the VAWN program and by the National Science Foundation under Grants CCF-0728748 and IIS-1116656. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Ce Zhu.

A. Mittal and A. C. Bovik are with the Laboratory for Image and Video Engineering (LIVE), The University of Texas, Austin, TX 78812 USA (e-mail: mittal.anish@gmail.com).

R. Soundararajan was with the The University of Texas, Austin, TX 78812 USA. He is now with Qualcomm Research India, Bangalore, India.

Digital Object Identifier 10.1109/LSP.2012.2227726

Objective image quality assessment refers to automatically predict the quality of distorted images as would be perceived by an average human. If a naturalistic reference image is supplied against which the quality of the distorted image can be compared, the model is called full reference (FR) [2]. Conversely, NR IQA models assume that only the distorted image whose quality is being assessed is available. Existing general purpose NR IQA algorithms are based on models that can learn to predict human judgments of image quality from databases of human-rated distorted images [3]–[7]. These kinds of IQA models are necessarily limited, since they can only assess quality degradations arising from the distortion types that they have been trained on.

However, it is also possible to contemplate subcategories of general-purpose NR IQA models having tighter conditions. A model is "opinion-aware" (OA) if it has been trained on a database(s) of human rated distorted images and associated subjective opinion scores. Thus algorithms like DIIVINE [4], CBIQ [6], LBIQ [7], BLIINDS [5] and BRISQUE [3] are OA IQA models. Given the impracticality of obtaining collections of distorted images with co-registered human scores, models that do not require training on databases of human judgments of distorted images, and hence are "opinion unaware" (OU), are of great interest. One such effort was made in this direction by the authors of [8]. However, their model requires knowledge of the expected image distortions.

Likewise, among algorithms derived from OU models, distorted images may or may not be available during IQA model creation or training. For example, in highly unconstrained environments, such as a photograph upload site, the *a priori* nature of distortions may be very difficult to know. Thus a model may be formulated as "distortion aware" (DA) by training on (and hence tuning to) specific distortions, or it may be "distortion unaware" (DU), relying instead only on exposure to naturalistic source images or image models to guide the QA process. While this may seem as an extreme paucity of information to guide design, it is worth observing that very successful FR IQA models (such as the structural similarity index (SSIM) [9]) are DU.

Our contribution in this direction is the development of a NSS-based modeling framework for OU-DU NR IQA design, resulting in a first of a kind NSS-driven blind OU-DU IQA model which does not require exposure to distorted images *a priori*, nor any training on human opinion scores. The new NR OU-DU IQA quality index performs better than the popular FR peak signal-to-noise-ratio (PSNR) and structural similarity (SSIM) index and delivers performance at par with top performing NR OA-DA IQA approaches.

1070-9908/$31.00 © 2012 IEEE

II. No Reference Opinion-Unaware Distortion-Unaware IQA Model

Our new NR OU-DU IQA model is based on constructing a collection of "quality aware" features and fitting them to a multivariate Gaussian (MVG) model. The quality aware features are derived from a simple but highly regular natural scene statistic (NSS) model. The quality of a given test image is then expressed as the distance between a multivariate Gaussian (MVG) fit of the NSS features extracted from the test image, and a MVG model of the quality aware features extracted from the corpus of natural images.

A. Spatial Domain NSS

Our "completely blind" IQA model is founded on perceptually relevant spatial domain NSS features extracted from local image patches that effectively capture the essential low-order statistics of natural images. The classical spatial NSS model [10] that we use begins by preprocessing the image by processes of local mean removal and divisive normalization:

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

where $i \in \{1, 2 \dots M\}, j \in \{1, 2 \dots N\}$ are spatial indices, M and N are the image dimensions, and

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

$$\sigma(i,j) = \sqrt{\sum_{k=-K}^{K} \sum_{l=-L}^{L} w_{k,l} [I(i+k, j+l) - \mu(i,j)]^2} \tag{3}$$

estimate the local mean and contrast, respectively, where $w = \{w_{k,l} | k = -K, \dots, K, l = -L, \dots L\}$ is a 2D circularly-symmetric Gaussian weighting function sampled out to 3 standard deviations ($K = L = 3$) and rescaled to unit volume.

The coefficients (1) have been observed to reliably follow a Gaussian distribution when computed from natural images that have suffered little or no apparent distortion [10]. This ideal model, however, is violated when the images do not derive from a natural source (e.g., computer graphics) or when natural images are subjected to unnatural distortions. The degree of modification can be indicative of perceptual distortion severity.

The NSS features used in the NIQE index are similar to those used in a prior OA-DA IQA model called BRISQUE [3]. However, NIQE only uses the NSS features from a corpus of natural images while BRISQUE is trained on features obtained from both natural and distorted images and also on human judgments of the quality of these images. Therefore, BRISQUE is limited to the types of distortions it has been tuned to. By comparison, the NIQE Index is not tied to any specific distortion type, yet, as will be shown, delivers nearly comparable predictive power on the same distortions the BRISQUE index has been trained on, with a similar low complexity.

B. Patch Selection

Once the image coefficients (1) are computed, the image is partitioned into $P \times P$ patches. Specific NSS features are then computed from the coefficients of each patch. However, only a subset of the patches are used for the following reason.

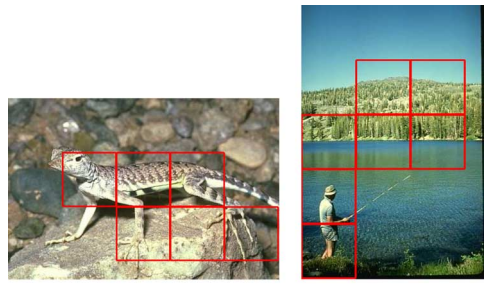

Fig. 1. The marked blocks in images (a) and (b) depict instances of natural image patches selected using a local sharpness measure.

Every image is subject to some kind of limiting distortion [11]. For instance, there is a loss of resolution due to defocus blur in parts of most images due to the limited depth of field (DOF) of any single-lens camera. Since humans appear to more heavily weight their judgments of image quality from the sharp image regions [12], more salient quality measurements can be made from sharp patches. Setting aside the question of the aesthetic appeal of having some parts of an image sharper than others, any defocus blur represents a potential loss of visual information.

We use a simple device to preferentially select from amongst a collection of natural patches those that are richest in information and less likely to have been subjected to a limiting distortion. This subset of patches is then used to construct a model of the statistics of natural image patches.

The variance field (3) has been largely ignored in the past in NSS based image analysis, but it is a rich source of structural image information that can be used to quantify local image sharpness. Letting the $P \times P$ sized patches be indexed $b = 1, 2, \dots, B$, a direct approach is to compute the average local deviation field of each patch indexed b:

$$\delta(b) = \sum \sum_{(i,j) \in \text{patch} b} \sigma(i,j) \tag{4}$$

where δ denotes *l*ocal activity/sharpness.

Once the sharpness of each patch is found, those having a suprathreshold sharpness $\delta > T$ are selected. The threshold T is picked to be a fraction p of the peak patch sharpness over the image. In our experiments, we used the nominal value $p = 0.75$. Examples of this kind of patch selection are shown in Fig. 1. We have observed only small variations in performance when p is varied in the range [0.6, 0.9].

C. Characterizing Image Patches

Given a collection of natural image patches selected as above, their statistics are characterized by "quality aware" NSS features computed from each selected patch [3]. Prior studies of NSS based image quality have shown that the generalized Gaussian distribution effectively captures the behavior of the coefficients (1) of natural and distorted versions of them [13].

The generalized Gaussian distribution (GGD) with zero mean is given by:

$$f(x; \alpha, \beta) = \frac{\alpha}{2\beta\Gamma(1/\alpha)} \exp\left(-\left(\frac{|x|}{\beta}\right)^{\alpha}\right) \tag{5}$$

where $\Gamma(\cdot)$ is the gamma function:

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

The parameters of the GGD (α, β), can be reliably estimated using the moment-matching based approach proposed in [14]. The signs of the transformed image coefficients (1) have been observed to follow a fairly regular structure. However, distortions disturb this correlation structure [3]. This deviation can be captured by analyzing the sample distribution of the products of pairs of adjacent coefficients computed along horizontal, vertical and diagonal orientations: $\hat{I}(i,j)\hat{I}(i,j+1), \hat{I}(i,j)\hat{I}(i+1,j), \hat{I}(i,j)\hat{I}(i+1,j+1)$ and $\hat{I}(i,j)\hat{I}(i+1,j-1)$ for $i \in \{1,2\ldots M\}$ and $j \in \{1,2\ldots N\}$ [3].

The products of neighboring coefficients are well-modeled as following a zero mode asymmetric generalized Gaussian distribution (AGGD) [15]:

$$f(x;\gamma,\beta_l,\beta_r) = \begin{cases} \frac{\gamma}{(\beta_l+\beta_r)\Gamma\left(\frac{1}{\gamma}\right)} \exp\left(-\left(\frac{-x}{\beta_l}\right)^\gamma\right) \forall x \le 0 \\ \frac{\gamma}{(\beta_l+\beta_r)\Gamma\left(\frac{1}{\gamma}\right)} \exp\left(-\left(\frac{x}{\beta_r}\right)^\gamma\right) \forall x \ge 0. \end{cases} \tag{7}$$

The parameters of the AGGD $(\gamma, \beta_l, \beta_r)$ can be efficiently estimated using the moment-matching based approach in [15]. The mean of the distribution is also useful:

$$\eta = (\beta_r - \beta_l)\frac{\Gamma\left(\frac{2}{\gamma}\right)}{\Gamma\left(\frac{1}{\gamma}\right)}. \tag{8}$$

By extracting estimates along the four orientations, 16 parameters are arrived at yielding 18 overall. All features are computed at two scales to capture multiscale behavior, by low pass filtering and downsampling by a factor of 2, yielding a set of 36 features.

D. Multivariate Gaussian Model

A simple model of the NSS features computed from natural image patches can be obtained by fitting them with an MVG density, providing a rich representation of them:

$$f_X(x_1,\ldots,x_k) = \frac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}} \times \exp\left(-\frac{1}{2}(x-\nu)^T\Sigma^{-1}(x-\nu)\right) \tag{9}$$

where $(x_1,\ldots,x_k)$ are the NSS features computed in (5)–(8), and ν and Σ denote the mean and covariance matrix of the MVG model, which are estimated using a standard maximum likelihood estimation procedure [16]. We selected a varied set of 125 natural images with sizes ranging from 480×320 to 1280×720 to obtain the multivariate Gaussian model. Images were selected from copyright free Flickr data and from the Berkeley image segmentation database [17] making sure that no overlap occurs with the test image content. The images may be viewed at http://live.ece.utexas.edu/research/quality/pristinedata.zip.

E. NIQE Index

The new OU-DU IQA index, called NIQE, is applied by computing the 36 identical NSS features from patches of the same size $P \times P$ from the image to be quality analyzed, fitting them with the MVG model (9), then comparing its MVG fit to the natural MVG model. The sharpness criterion (4) is not applied to these patches because loss of sharpness in distorted images is indicative of distortion and neglecting them would lead to incorrect evaluation of the distortion severity. The patch size was set to 96×96 in our implementation. However, we observed stable performance across patch sizes ranging from 32×32 to 160×160.

Finally, the quality of the distorted image is expressed as the distance between the quality aware NSS feature model and the MVG fit to the features extracted from the distorted image:

$$D(\nu_1,\nu_2,\Sigma_1,\Sigma_2) = \sqrt{\left((\nu_1-\nu_2)^T\left(\frac{\Sigma_1+\Sigma_2}{2}\right)^{-1}(\nu_1-\nu_2)\right)} \tag{10}$$

where ν_1, ν_2 and Σ_1, Σ_2 are the mean vectors and covariance matrices of the natural MVG model and the distorted image's MVG model.

III. Performance Evaluation

A. Correlation With Human Judgments of Visual Quality

To test the performance of the NIQE index, we used the LIVE IQA database [2] of 29 reference images and 779 distorted images spanning five different distortion categories—JPEG and JPEG2000 (JP2K) compression, additive white Gaussian noise (WN), Gaussian blur (blur) and a Rayleigh fast fading channel distortion (FF). A difference mean opinion score (DMOS) associated with each image represents its subjective quality.

Since all of the OA IQA approaches that we compare NIQE to require a training procedure to calibrate the regressor module, we divided the LIVE database randomly into chosen subsets for training and testing. Although our blind approach and the FR approaches do not require this procedure, to ensure a fair comparison across methods, the correlations of predicted scores with human judgments of visual quality are only reported on the test set. The dataset was divided into 80% training and 20% testing—taking care that no overlap occurs between train and test content. This train-test procedure was repeated 1000 times to ensure that there was no bias due to the spatial content used for training. We report the median performance across all iterations.

We use Spearman's rank ordered correlation coefficient (SROCC), and Pearson's (linear) correlation coefficient (LCC) to test the model. The NIQE scores are passed through a logistic non-linearity [2] before computing LCC for mapping to DMOS space. We compared NIQE with three FR indices: PSNR, SSIM [9] and multiscale SSIM (MS-SSIM) [18], five general purpose OA-DA algorithms—CBIQ [6], LBIQ [7], BLIINDS-II [5], DIIVINE [4], BRISQUE [3] and the DA-OU approach TMIQ [8].

As can be seen from Tables I and II, NIQE performs better than the FR PSNR and SSIM and competes well with all of the top performing OA-DA NR IQA algorithms. This is a fairly remarkable demonstration of the relationship between quantified image naturalness and perceptual image quality.

TABLE I
MEDIAN SPEARMAN RANK ORDERED CORRELATION COEFFICIENT (SROCC) ACROSS 1000 TRAIN-TEST COMBINATIONS ON THE LIVE IQA DATABASE. *ITALICS* INDICATE (OA/OU)-DA NO-REFERENCE ALGORITHMS AND **BOLD FACE** INDICATES THE NEW OU-DU MODEL ALGORITHM

	JP2K	JPEG	WN	Blur	FF	All
PSNR	0.8646	0.8831	0.9410	0.7515	0.8736	0.8636
SSIM	0.9389	0.9466	0.9635	0.9046	0.9393	0.9129
MS-SSIM	0.9627	0.9785	0.9773	0.9542	0.9386	0.9535
CBIQ	*0.8935*	*0.9418*	*0.9582*	*0.9324*	*0.8727*	*0.8954*
LBIQ	*0.9040*	*0.9291*	*0.9702*	*0.8983*	*0.8222*	*0.9063*
BLIINDS-II	*0.9323*	*0.9331*	*0.9463*	*0.8912*	*0.8519*	*0.9124*
DIIVINE	*0.9123*	*0.9208*	*0.9818*	*0.9373*	*0.8694*	*0.9250*
BRISQUE	*0.9139*	*0.9647*	*0.9786*	*0.9511*	*0.8768*	*0.9395*
TMIQ	*0.8412*	*0.8734*	*0.8445*	*0.8712*	*0.7656*	*0.8010*
NIQE	**0.9172**	**0.9382**	**0.9662**	**0.9341**	**0.8594**	**0.9135**

TABLE II
MEDIAN LINEAR CORRELATION COEFFICIENT ACROSS 1000 TRAIN-TEST COMBINATIONS ON THE LIVE IQA DATABASE. *ITALICS* INDICATE (OA/OU)-DA NO-REFERENCE ALGORITHMS AND **BOLD FACE** INDICATES THE NEW OU-DU MODEL ALGORITHM

	JP2K	JPEG	WN	Blur	FF	All
PSNR	0.8762	0.9029	0.9173	0.7801	0.8795	0.8592
SSIM	0.9405	0.9462	0.9824	0.9004	0.9514	0.9066
MS-SSIM	0.9746	0.9793	0.9883	0.9645	0.9488	0.9511
CBIQ	*0.8898*	*0.9454*	*0.9533*	*0.9338*	*0.8951*	*0.8955*
LBIQ	*0.9103*	*0.9345*	*0.9761*	*0.9104*	*0.8382*	*0.9087*
BLIINDS-II	*0.9386*	*0.9426*	*0.9635*	*0.8994*	*0.8790*	*0.9164*
DIIVINE	*0.9233*	*0.9347*	*0.9867*	*0.9370*	*0.8916*	*0.9270*
BRISQUE	*0.9229*	*0.9734*	*0.9851*	*0.9506*	*0.9030*	*0.9424*
TMIQ	*0.8730*	*0.8941*	*0.8816*	*0.8530*	*0.8234*	*0.7856*
NIQE	**0.9370**	**0.9564**	**0.9773**	**0.9525**	**0.9128**	**0.9147**

B. Number of Natural Images

We addressed the question: 'How many natural images are needed to obtain a stable model that can correctly predict image quality?' Such an analysis provides an idea of the quality prediction power of the NSS features and how well they generalize with respect to image content.

To undertake this evaluation, we varied the number of natural images K from which patches are selected and used for model fitting. Fig. 2 shows the performance against the number of images. An error band is drawn around each point to indicate the standard deviation in performance across 100 iterations of different sample sets of K images. It may be observed that a stable natural model can be obtained using a small set of images.

IV. CONCLUSION

We have created a first of a kind blind IQA model that assesses image quality without knowledge of anticipated distortions or human opinions of them. The quality of the distorted image is expressed as a simple distance metric between the model statistics and those of the distorted image. The new model outperforms FR IQA models and competes with top performing NR IQA trained on human judgments of known distorted images. Such a model has great potential to be applied in unconstained environments.

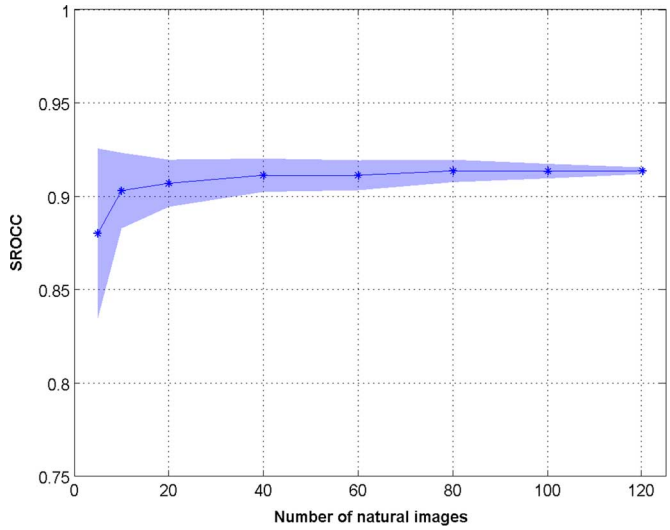

Fig. 2. Variation of performance with the number of natural images K. Error schmears around each point indicate the standard deviation in performance across 100 iterations for $5 < K < 125$.

REFERENCES

[1] "Image obsessed," *National Geographic* vol. 221, p. 35, 2012.
[2] 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, 2006.
[3] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, 2012, to be published.
[4] 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, 2011.
[5] M. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3339–3352, 2012.
[6] P. Ye and D. Doermann, "No-reference image quality assessment using visual codebook," in *IEEE Int. Conf. Image Process.*, 2011.
[7] H. Tang, N. Joshi, and A. Kapoor, "Learning a blind measure of perceptual image quality," in *Int. Conf. Comput. Vision Pattern Recognition*, 2011.
[8] A. Mittal, G. S. Muralidhar, J. Ghosh, and A. C. Bovik, "Blind image quality assessment without human training using latent quality factors," *IEEE Signal Process. Lett.*, vol. 19, pp. 75–78, 2011.
[9] 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, 2004.
[10] D. L. Ruderman, "The statistics of natural images," *Netw. Comput. Neural Syst.*, vol. 5, no. 4, pp. 517–548, 1994.
[11] A. C. Bovik, "Perceptual image processing: Seeing the future," *Proc. IEEE*, vol. 98, no. 11, pp. 1799–1803, 2010.
[12] R. Hassen, Z. Wang, and M. Salama, "No-reference image sharpness assessment based on local phase coherence measurement," in *IEEE Int. Conf. Acoust. Speech Signal Process.*, 2010, pp. 2434–2437.
[13] A. K. Moorthy and A. C. Bovik, "Statistics of natural image distortions," in *IEEE Int. Conf. Acoust. Speech Signal Process.*, pp. 962–965.
[14] 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, 1995.
[15] N. E. Lasmar, Y. Stitou, and Y. Berthoumieu, "Multiscale skewed heavy tailed model for texture analysis," in *IEEE Int. Conf. Image Process*, 2009, pp. 2281–2284.
[16] C. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer, 2006, vol. 4.
[17] 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 *Int. Conf. Comput. Vision*, 2001, vol. 2, pp. 416–423.
[18] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Asilomar Conf. Sig., Syst. Comput.*, 2003, vol. 2, pp. 1398–1402, IEEE.