BLIND/REFERENCELESS IMAGE SPATIAL QUALITY EVALUATOR

Anish Mittal, Anush K. Moorthy and Alan C. Bovik

Department of Electrical and Computer Engineering
University of Texas at Austin
Austin, Texas 78712

ABSTRACT

We propose a natural scene statistic based **B**lind/**R**eferenceless **I**mage **S**patial **QU**ality **E**valuator (BRISQUE) which extracts the point wise statistics of local normalized luminance signals and measures image naturalness (or lack there of) based on measured deviations from a natural image model. We also model the distribution of pairwise statistics of adjacent normalized luminance signals which provides distortion orientation information. Although multi scale, the model uses *easy to compute* features making it computationally fast and time efficient. The frame work is shown to perform **statistically better** than other proposed no reference algorithms and full reference structural similarity index (SSIM) .

1. INTRODUCTION

Personal digital assistants (PDA), hand held displays and smart phones capable of connecting with internet have brought a revolution in communication and networking technologies. This has invited new challenges for service providers to optimize networks with respect to structured multimedia data such as text, images and videos given limited bandwidth. Traditional delivery methods have mainly focused on optimizing throughput, buffer lengths and capacity. While these approaches have shown great performance, one would expect that accounting for the nature of multimedia data and modeling the behaviour of the end consumer could lead to appreciable gains in performance. This suggests the idea of using human perception models for service quality optimization. With this application mindset, we propose a model based approach to predict visual quality of distorted images. It is a blind/no reference approach that does not require any reference image to compare the quality of, distorted image against. While auxiliary information about a reference certainly is helpful and makes the problem much simpler, sending that information over the network causes an additional bandwidth overhead and this information may be difficult to retrieve in real scenarios. Also, assumptions on the presence of a *pristine* reference is questionable [1] since all captured images are afflicted with distortions normally occuring during a quality photo shoot.

The preliminary work in the area of no reference quality assessment started by modeling blur, noise or compression artifacts to address the blurriness, graininess, blockiness and ringing in images [4, 10, 19, 13, 3, 18]. However, it is of great interest to develop dual approaches which model the statistics of natural images (or lack there of). In this way, it may be possible to avoid detailed modeling of distortions, making possible approaches that are independent from the distortions images are afflicted with. Two recently developed algorithms **BL**ind **I**mage **N**otator using **DCT S**tatistics (BLIINDS) - II [12] and the **D**istortion **I**dentification-based **I**mage **V**erity and **IN**tegrity **E**valuation (DIIVINE) were developed along these lines, making no assumption on the kind of distortion medium image is passed through. These proposed algorithms have some associated pitfalls however, that might make them ineffective for some applications. DIIVINE [9] is expensive since it involves the computation of a relatively large number of features. BLINDS-II [12], computes the block statistics of discrete cosine transform coefficients which is also time consuming. Both DIIVINE and BLIINDS-II deliver top NR QA performance (to date), but it is of interest to consider whether purely spatial NR QA algorithms not requiring local transforms can be developed towards performance and computational efficiency.

In this paper, we explain a model which makes use of normalized luminance coefficients in the spatial domain, as well as pairwise products of adjacent normalized luminance coefficients which provides distortion orientation information. We then demonstrate how parameterizing the respective distributions of these coefficients can be used as statistical features that correlate well with human judgements of image quality. These features are used for distortion identification and distortion specific quality assessment, which when combined produces a distortion independent measure of image quality.

2. NATURAL SCENE STATISTICS MODEL IN THE SPATIAL DOMAIN

Our approach is built on a foundation of models of statistical regularities observed in natural scenes. 'Natural' scenes subsumes any image taken using an optical camera with no artificial processing. Thus, natural images are not necessar-

ily images of outdoor environments such as trees or skies. Natural scene statistics (NSS) have been analysed by psychophysicists in both the spatial domain [11], and the wavelet domain [17]. In the present development, we concentrate on spatial domain statistics. Ruderman observed that local mean subtraction and variance normalization on log-contrast values [11] produces decorrelated coefficients which follow gaussian like distributions. This process is similar to center surround mechanism followed by adaptive gain control. Given a luminance image I, we compute:

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

where, $i \in 1, 2 \dots M$, $j \in 1, 2 \dots N$ are the spatial indices, M, N are the image height and width respectively, $C = 1$ is a constant that is used to prevent instabilities from occurring when the denominator tends to zero (eg., in the case of an image patch corresponding to the plain sky) and

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

$$\sigma(i,j) = \sqrt{\sum_{k=-K}^{K} \sum_{l=-L}^{L} w_{k,l}(Y_{k,l}(i,j) - \mu(i,j))^2} \quad (3)$$

and $w = \{w_{k,l} | k = -K, \dots, K, l = -L, \dots L\}$ is a circularly-symmetric Gaussian filter sampled out to 3 standard deviations. In our implementation, $K = L = 3$. Variation in these values between $2 - 4$ did not have much effect on performance. In our algorithm, such normalized luminance coefficients (eq. 1) are referred to as mean subtracted contrast normalized (MSCN) coefficients, and are used for all further processing.

Our model is based on the premise that MSCN coefficients have characteristic statistical properties that are perturbed in the presence of distortion [7] and that modeling these coefficients makes it possible to predict the type and perceptual severity of distortions [9]. Fig. 1 illustrates how MSCN coefficient distributions change with distortion type. An important observation is how pristine images exhibit Gaussian-like statistical properties as shown by Ruderman [11]. By contrast, each distortion modifies natural scene statistics in a unique way. For instance, blocking and blurring artifacts produced by JPEG induce Laplacian-like statistics as compared with grainy artifacts caused by white noise, which increases the variance of the distribution.

To quantify the statistics of natural images and how they change with distortion, we utilize an AGGD (Asymmetric Generalized Gaussian Model) distribution to fit the MSCN statistics from pristine as well as distorted images. For each image, we estimate 2 parameters $(\alpha, (\sigma_l^2 + \sigma_r^2)/2)$ from the AGGD fit of MSCN coefficients, forming our first set of features.
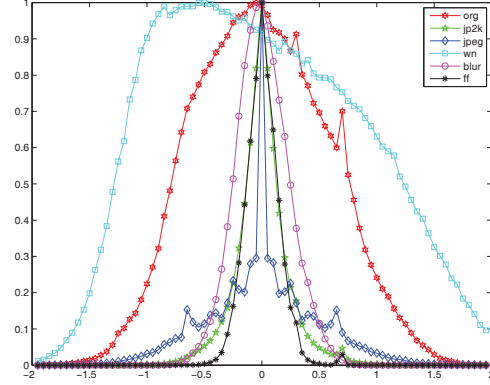
Fig. 1. Histogram of MSCN coefficients for a natural undistorted image and various distorted versions. Distortions from the LIVE IQA database [15] – JPEG2000 (Jp2K) and JPEG compression, additive white Gaussian noise (WN), Gaussian blur (blur), and a Rayleigh fast-fading channel simulation (FF).

While MSCN coefficients are definitely more homogenous, the signs of the coefficients exhibit a regular structure [16]. In order to quantify this dependency between neighbors, we model directional relationships via pairwise products between adjacent MSCN coefficients at a distance of 1 pixel, along four orientations – horizontal (Hz), vertical (Vt), maindiagonal (Dg_1) and secondary-diagonal (Dg_2).

$$Hz(i,j) = MSCN(i,j)MSCN(i,(j+1)modN) \quad (4)$$

$$Vt(i,j) = MSCN(i,j)MSCN((i+1)modM,j) \quad (5)$$

$$Dg_1(i,j) = MSCN(i,j)MSCN((i+1)modM,(j+1)modN) \quad (6)$$

$$Dg_2(i,j) = MSCN(i,j)MSCN((i+1)modM,(j-1)modN) \quad (7)$$

$\forall i \in 1, 2 \dots M \quad and \quad \forall j \in 1, 2 \dots N$.

Fig. 2 shows how the paired products along each of the four orientations vary in the presence of distortion for the reference image and each of its distorted versions.

The paired product distributions are also parameterized using the AGGD which effectively captures the asymmetry in their distributions as observed in Fig. 2. Hence, apart from the parameters of the AGGD $(\alpha, \sigma_l^2, \sigma_r^2)$, we also compute the mean of the paired product distribution to account for it. For an AGGD, the mean is given by:

$$\mu = (\beta_r - \beta_l)\frac{\gamma(\frac{2}{\alpha})}{\gamma(\frac{1}{\alpha})} \quad (8)$$

Thus for each paired product, $(\alpha, \sigma_l^2, \sigma_r^2, \mu)$ are computed, and these 16 parameters (4 parameters/orientation $\times$ 4 orientations) form another set of features. Owing to the inherent multiscale nature of images and distortions affecting
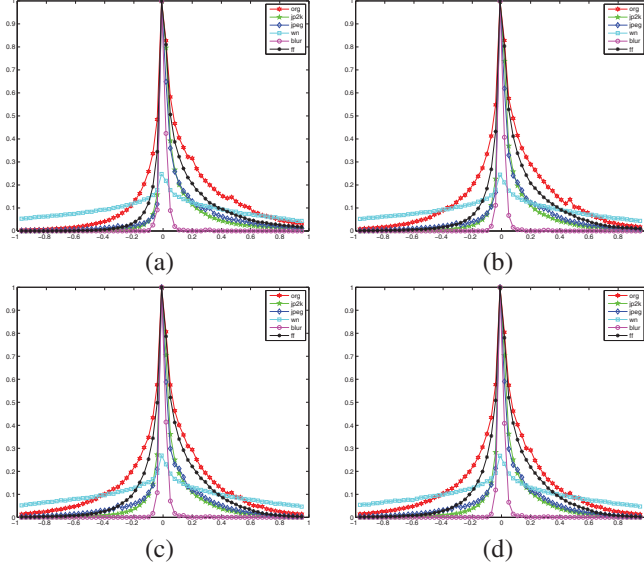
Fig. 2. Histograms of paired-products of MSCN coefficients for a natural undistorted image and its various distorted versions. (a) Horizontal, (b) Vertical, (c) Main-diagonal, (d) Secondary-diagonal. Distortions from the LIVE IQA database [15] – JPEG2000 (Jp2k) and JPEG compression, additive white Gaussian noise (WN), Gaussian blur (blur), and a Rayleigh fast-fading channel simulation (FF).

image structure across scales, we compute BRISQUE features across multiple scales. Further, it has been shown by researchers that multiscale QA algorithms perform well in terms of correlation with human perception [21, 12]. We observed that increasing the number of scales beyond 2 did not improve the performance. Thus, we have a total of 36 features – 18 at each scale.

3. ASYMMETRIC GENERALIZED GAUSSIAN MODEL

The AGGD with zero mode is given by:

$$f_X(x; \alpha, \sigma_l^2, \sigma_r^2) = \begin{cases} \frac{\alpha}{(\beta_l + \beta_r)\gamma(\frac{1}{\alpha})} \exp(-(\frac{-x}{\beta_l})^\alpha) & x < 0 \\ \frac{\alpha}{(\beta_l + \beta_r)\gamma(\frac{1}{\alpha})} \exp(-(\frac{x}{\beta_r})^\alpha) & x \geq 0 \end{cases}$$

(9)

where α is a shape parameter and σ_l and σ_r are scale parameters

$$\beta_l = \sigma_l \sqrt{\frac{\gamma(\frac{1}{\alpha})}{\gamma(\frac{3}{\alpha})}}$$

(10)

$$\beta_r = \sigma_r \sqrt{\frac{\gamma(\frac{1}{\alpha})}{\gamma(\frac{3}{\alpha})}}$$

(11)

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

$$\gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt \quad a > 0$$

(12)

α governs 'shape' of the distribution and σ_l^2 and σ_r^2 determine the left and right variances (about the mode) respectively. The AGGD subsumes the generalized Gaussian distribution (GGD) [14] when left and right variance are equal. Researchers have utilized this asymmetric distribution to model skewed and heavy-tailed distributions for the analysis of texture [6]. The moment based method proposed in [6] was used to estimate the parameters of the AGGD ($\alpha, \sigma_l^2, \sigma_r^2$).

4. TWO-STAGE QUALITY ASSESSMENT FRAMEWORK

Once we have these features, a 2 stage framework similar to DIVIINE [9] is used for quality inference. First, the computed features are used to train a probabilistic support vector classification (SVC) to find the probability of occurence of each distortion in the image, followed by evaluation of distortion specific image quality using trained support vector regression (SVR) functions for each distortion. Once we have quality from each regressor, we compute dot product of the distortion probability vector and the distortion quality vector. We have used the LIBSVM package [2] to implement the SVM and SVR and the radial bias function (RBF) kernel for both classification and regression.

The motivation behind this framework is that each distortion afflicts natural images in a different way. Therefore, a different set of features are important for every distortion. We use a generic feature set to represent the 'naturalness' of images. Although the features do not model any specific artifacts such as blur, ringing or blocking, they still capture the characteristics of diverse distortions making it possible to train a classifier which can map deviations of the model parameters to the type of distortion the image is afflicted with.

5. PERFORMANCE EVALUATION

The performance of quality assessment algorithms is gauged by correlation with human perception, where human opinions on visual quality are usually obtained from large-scale human studies, using a large number of distorted (and possibly reference) signals that are rated by human observers. The mean opinion scores (MOS) or differential mean opinion score (DMOS) is obtained by averaging across humans for each of the visual signals in the study. Resulting MOS/DMOS is representative of the perceptual quality of the visual signal. The goal of quality assessment (QA) algorithms is to emulate quality scores for these signals so that they correlate well with human opinion on quality (MOS/DMOS).

	JP2k	JPEG	WN	Blur	FF	All
PSNR	0.8947	0.9191	0.9536	0.8357	0.8964	0.8873
SSIM	0.9461	0.9547	0.9679	0.9321	0.9393	0.9201
MSSSIM	0.9632	0.9773	0.9786	0.9607	0.9464	0.9550
Anisotropy	*0.2931*	*0.1519*	*0.7286*	*0.6929*	*0.5929*	*0.3614*
BLIINDS-II	*0.9270*	*0.9376*	*0.9565*	*0.8902*	*0.8409*	*0.9123*
DIIVINE	*0.9114*	*0.9213*	*0.9818*	*0.9386*	*0.8732*	*0.9265*
BRISQUE	*0.8999*	*0.9467*	*0.9849*	*0.9435*	*0.8861*	*0.9314*

Table 1. Median spearman rank ordered correlation coefficient (SROCC) across 1000 train-test combination on the LIVE IQA database. *Italics* indicate no-reference algorithms.

	JP2k	JPEG	WN	Blur	FF	All
PSNR	0.9044	0.9375	0.8981	0.8490	0.9119	0.8866
SSIM	0.9575	0.9630	0.9887	0.9395	0.9644	0.9110
MSSSIM	0.9795	0.9357	0.9919	0.9762	0.9689	0.9529
Anisotropy	*0.2235*	*0.1351*	*0.5855*	*0.5631*	*0.5290*	*0.2332*
BLIINDS-II	*0.9329*	*0.9386*	*0.9674*	*0.8944*	*0.8665*	*0.9145*
DIIVINE	*0.9225*	*0.9363*	*0.9869*	*0.9383*	*0.8938*	*0.9283*
BRISQUE	*0.9090*	*0.9551*	*0.9903*	*0.9498*	*0.9148*	*0.9377*

Table 2. Median pearson's linear correlation coefficient (PLCC) across 1000 train-test combination on the LIVE IQA database. *Italics* indicate no-reference algorithms.

	JP2k	JPEG	WN	Blur	FF	All
Accuracy (%)	82.9	88.9	100.0	96.7	83.3	88.6

Table 3. Median classification accuracy across 1000 train-test trials.

5.1. Correlation with Human Perception

We used the LIVE IQA database [15] to test the performance of our algorithm. It consists of 29 reference images with 779 distorted images which span five different distortion categories - JPEG and JPEG 2000 compression, white noise, Gaussian blur and a Rayleigh fast fading channel distortion where each distorted image has an associated difference mean opinion score (DMOS). A classifer is trained to predict the occurence of each distortion and a regression function is learned corresponding to each distortion. Quality is computed using this information as explained earlier.

We divide the dataset into 80% training and 20% testing ensuring that there is no overlap between train and test content. To make sure that the reported results do not depend on the selection of spatial content used to train the framework, we randomly pick 80% of the spatial content for training and the remaining 20% to test and repeated this standard cross validation procedure 1000 times on the LIVE database. We report the median of the scores across iterations. [1].

The performance measures used are Spearman's rank ordered correlation coefficient (SROCC) and Pearson's linear correlation coefficient (PLCC) between the predicted algorithm scores and DMOS. Before computing PLCC, algorithm scores are mapped to DMOS using a logistic non-linearity as described in [15]. A value close to 1 for SROCC and PLCC indicate good performance in terms of correlation with human opinion. These performance indices are tabulated in Tables 1 and 2 respectively.

We tabulated the performance of the full-reference indices: peak-signal-to-noise ratio (PSNR) and structural similarity index (SSIM) [20] and several no-reference algorithms: Anisotropy based NR IQA [5], BLind Image Notator using DCT Statistics (BLIINDS) - II [12] and the Distortion Identification-based Image Verity and INtegrity Evaluation (DIIVINE) index [9] for comparison.

As seen from Tables 1 and 2, BRISQUE performs quite well in terms of correlation with human perception, beating present-day full-reference and no-reference image quality assessment indices.

5.2. Classification Accuracy

Although a probabilistic framework is used for distortion identification, we also show how our features are useful for explicit distortion-identification [8]. We report the median classification accuracy of the classifier for each of the distortions in the LIVE database, as well as across all distortions in Table 3.

6. CONCLUSION

We proposed a fast, simple spatial domain based natural scene statistics based model for no reference image quality assessment and demonstrated that not only does it give superior performance compared to existing NR IQA approaches, but it also performs better than the full-reference peak signal-to-noise-ratio (PSNR) and the structural similarity index (SSIM). BRISQUE may be easily extended beyond the distortions considered here making it suitable for a general-purpose blind IQA problems. Furthermore, BRISQUE is computationally quite efficient making it an attractive option for use in practical applications such as image restortation and quality evaluation in networks.

7. REFERENCES

[1] A.C. Bovik. Perceptual image processing: Seeing the future. *IEEE*, 98(11):1799–1803, 2010.

[2] C.C. Chang and C.J. Lin. LIBSVM: A library for support vector machines. 2001.

---

[1]Performance was evaluated using the realigned DMOS scores as recommended in [15] only on the distorted images, as in [15] . Further, note that due to the random 1000 trials, there may be a slight discrepancy between results reported here and elsewhere, however, these differences in correlations would not be statistically significant, and are simply an artifact of the random train-test sampling.

[3] J. Chen, Y. Zhang, L. Liang, S. Ma, R. Wang, and W. Gao. A no-reference blocking artifacts metric using selective gradient and plainness measures. *Proceedings of Pacific Rim Conference on Multimedia: Advances in Multimedia Information Processing*, pages 894–897, 2008.

[4] R. Ferzli and L.J. Karam. A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB). *IEEE Transactions on Image Processing*, 18(4):717–728, 2009.

[5] S. Gabarda and G. Cristóbal. Blind image quality assessment through anisotropy. *Journal of Optical Society of America*, 24(12):42–51, 2007.

[6] N.E. Lasmar, Y. Stitou, and Y. Berthoumieu. Multiscale skewed heavy tailed model for texture analysis. In *International Conference on Image Processing*, pages 2281–2284. IEEE, 2009.

[7] A.K. Moorthy and A.C. Bovik. Statistics of natural image distortions. In *International Conference on Acoustics Speech and Signal Processing*, pages 962–965.

[8] A.K. Moorthy and A.C. Bovik. A two-stage framework for blind image quality assessment. In *International Conference on Image Processing*, pages 2481–2484, 2010.

[9] A.K. Moorthy and A.C. Bovik. Blind image quality assessment: From natural scene statistics to perceptual quality. *IEEE Transactions on Image Processing*, 2011.

[10] N.D. Narvekar and L.J. Karam. A no-reference perceptual image sharpness metric based on a cumulative probability of blur detection. In *International Workshop on Quality of Multimedia Experience*, pages 87–91, 2009.

[11] D.L. Ruderman. The statistics of natural images. *Network computation in neural systems*, 5(4):517–548, 1994.

[12] M. Saad, A.C. Bovik, and C. Charrier. Model-Based Blind Image Quality Assessment Using Natural DCT Statistics . *IEEE Transactions on Image Processing*, 2011.

[13] N.G. Sadaka, L.J. Karam, R. Ferzli, and G.P. Abousleman. A no-reference perceptual image sharpness metric based on saliency-weighted foveal pooling. In *International Conference on Image Processing*, pages 369–372, 2008.

[14] K. Sharifi and A. Leon-Garcia. Estimation of shape parameter for generalized Gaussian distributions in subband decompositions of video. *IEEE Transactions on Circuits and Systems for Video Technology*, 5(1):52–56, 1995.

[15] H.R. Sheikh, M.F. Sabir, and A.C. Bovik. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on Image Processing*, 15(11):3440–3451, 2006.

[16] E P Simoncelli. Statistical modeling of photographic images. In Alan Bovik, editor, *Handbook of Image and Video Processing*, chapter 4.7, pages 431–441. Academic Press, May 2005. 2nd edition.

[17] A. Srivastava, A.B. Lee, E.P. Simoncelli, and S.C. Zhu. On advances in statistical modeling of natural images. *Journal of mathematical imaging and vision*, 18(1):17–33, 2003.

[18] S. Suthaharan. No-reference visually significant blocking artifact metric for natural scene images. *Journal of Signal Processing*, 89(8):1647 – 1652, 2009.

[19] S. Varadarajan and L.J. Karam. An improved perception-based no-reference objective image sharpness metric using iterative edge refinement. In *International Conference on Image Processing*, pages 401–404, 2008.

[20] Z. Wang, A.C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[21] Z. Wang, E. P. Simoncelli, and A.C. Bovik. Multiscale structural similarity for image quality assessment. In *Asilomar Conference on Signals, Systems and Computers*, volume 2, pages 1398–1402, 2003.