# MAKING IMAGE QUALITY ASSESSMENT ROBUST

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

Department of Electrical and Computer Engineering
The University of Texas at Austin
Austin, TX, USA

## ABSTRACT

We develop a robust framework for natural scene statistic (NSS) model based blind image quality assessment (IQA). The robustified IQA model utilizes a robust statistics approach based on *L*-moments. Such robust statistics based approaches are effective when natural or distorted images deviate from assumed statistical models, and achieves better prediction performance on distorted images relative to human subjective judgments. We also show how robustifying the model makes IQA approach resilient against deviation in model assumptions, small variations in the distortions and amount of data the model is trained on.

***Index Terms***— Image quality assessment, BRISQUE, spatial domain, robust statistics, *L*-moments

## 1. INTRODUCTION

Various quantitative measures of image quality have been proposed under which different amounts of *a priori* information is assumed to be available [1]. Generally, no reference or blind IQA models seek to predict the quality of distorted images using learned natural and distorted scene knowledge but without access to reference images [2–7].

We only consider NSS based no reference (NR) IQA models here. NR IQA algorithms are becoming well explored [2–7]. These algorithms learn to predict image quality from knowledge of anticipated distortions. Therefore, it is of interest to study whether these algorithms are robust to deviation in assumed statistical models and variations in the distortions they are trained on, such as modified codecs for compression, non Gaussian blur, or divergence from assumed noise characteristics.

To study this issue, we subject a top performing NR IQA index [2] to a variety of distortion perturbations. We also introduce a new robust NR IQA algorithm that uses robust statistical methods to estimate QA model parameters. We find that the robustified approach, which uses sample *L*-moments to estimate model parameters, is indeed robust to model and distortion perturbations and hence is better able to respond to

variations in NSS and to be more sensitive to distorted scene statistics (DSS), resulting in 'statistically' better correlation with human judgments of visual quality. Robustification process makes the approach generic enough to handle a wide variety of unanticipated distortions which might occur in real world scenarios. Devices such as wireless routers, smart phones and hand held devices could be programmed using such approaches to supply visual quality feedback to enable perceptual optimization in communication and surveillance networks.

We also conduct experiments to simulate the scenarios where only a small number of distorted images with associated DMOS scores are available for training the model and show that robust approaches fare better than model based approaches. Given that the procedure of obtaining human opinion scores is very inconvenient, time and effort expensive, robust approaches provide a clear advantage.

Robust statistical approaches have been applied to standard problems in image processing like local image smoothing [8], image reconstruction [9], blur classification [10], robust image denoising [11] but we are the first ones to explore their potential for image quality asssessment applications.

## 2. PROPOSED APPROACH

We use the recently introduced NSS based Blind/Referenceless Image Spatial QUality Evaluator (BRISQUE) proposed in [2] as a test IQA model to study its robustness against image model assumptions and small variations in distortions. We use the BRISQUE model because of its excellent predictive performance, its computational simplicity, and the easy way in which robust methods can be introduced. We modify BRISQUE using *L*-moment based parameter estimators to robustify the IQA model.

### 2.1. BRISQUE Features

The basic assumption underlying the BRISQUE approach is that statistical regularities of natural images are disturbed when distortions are introduced [2]. By assaying distortions based on how the NSS are changed, this model can be used
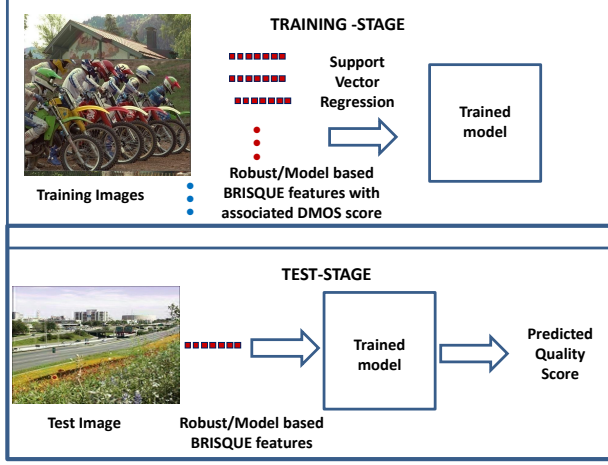
**Fig. 1**. shows the SVM based learning framework for quality inference using model based/robust BRISQUE features.

to create algorithms that predict subjective image quality accurately. Similar ideas have developed in [3, 4, 12–14].

BRISQUE uses classical spatial NSS model [15] that pre-processes the image by local mean removal followed by divisive normalization:

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

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

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

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

estimate the local mean and contrast, respectively, where $w = \{w_{k,l} | k = -K, \ldots, K, l = -L, \ldots L\}$ is a 2D circularly-symmetric Gaussian weighting function sampled out to 3 standard deviations ($K = L = 3$) and rescaled to unit volume. A GGD (Generalized Gaussian Model) distribution [16] is utilized as a model of the empirical distribution of the MSCN coefficients of natural images and how they change with distortion. The generalized Gaussian distribution (GGD) with zero mean is given by:

$$f(x; \alpha, \sigma^2) = \frac{\alpha}{2\beta\Gamma(1/\alpha)} \exp\left(-\left(\frac{|x|}{\beta}\right)^\alpha\right) \quad (4)$$

where

$$\beta = \sigma\sqrt{\frac{\Gamma(1/\alpha)}{\Gamma(3/\alpha)}} \quad (5)$$

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

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

The parameters of the GGD ($\alpha$, $\sigma^2$), can be reliably estimated using the moment-matching based approach proposed in [16]. The signs of the transformed image coefficients (1) have been observed to follow a fairly regular structure. However, distortions disturb this correlation structure [2]. 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 [2]: $\hat{I}(i,j)\hat{I}(i, j+1), \hat{I}(i,j)\hat{I}(i+1, j), \hat{I}(i,j)\hat{I}(i+1, j+1)$ and $\hat{I}(i,j)\hat{I}(i+1, j-1)$. The products of neighboring coefficients are well-modeled as following a zero mode asymmetric generalized Gaussian distribution (AGGD) [17]:

$$f(x; \gamma, \sigma_l^2, \sigma_r^2) = \begin{cases} \frac{\gamma}{(\beta_l + \beta_r)\Gamma\left(\frac{1}{\gamma}\right)} \exp\left(-\left(\frac{-x}{\beta_l}\right)^\gamma\right) & \forall x \leq 0 \\ \frac{\gamma}{(\beta_l + \beta_r)\Gamma\left(\frac{1}{\gamma}\right)} \exp\left(-\left(\frac{x}{\beta_r}\right)^\gamma\right) & \forall x \geq 0. \end{cases} \quad (7)$$

.

where

$$\beta_l = \sigma_l \sqrt{\frac{\Gamma\left(\frac{1}{\gamma}\right)}{\Gamma\left(\frac{3}{\gamma}\right)}} \quad (8)$$

$$\beta_r = \sigma_r \sqrt{\frac{\Gamma\left(\frac{1}{\gamma}\right)}{\Gamma\left(\frac{3}{\gamma}\right)}} \quad (9)$$

The parameters of the AGGD ($\gamma, \sigma_l^2, \sigma_r^2$) can be efficiently estimated using the moment-matching based approach in [17]. Mean of the distribution is also used as a feature:

$$\eta = (\beta_r - \beta_l)\frac{\Gamma(\frac{2}{\gamma})}{\Gamma(\frac{1}{\gamma})}. \quad (10)$$

16 parameters are arrived by computing estimates ($\gamma, \sigma_l^2, \sigma_r^2, \eta$) along the four orientations, yielding 18 overall features. 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.

## 2.2. A Robust NR IQA index

*L*-moments [18–20] are useful for robustly estimating the parameters of distributions. These estimators are relatively unaffected by small departures from model assumptions. We have found that they can be used to make BRISQUE less sensitive to empirical variations in NSS statistics and consequently, more responsive to variations in DSS. Such variations can occur when encountering a slightly different codec,

non Gaussian blur, or unanticipated noise characteristics, or even different distortions.

The $L$-moments of a sample $X_i$, $i=1, \ldots N$ utilize probability weighted moments [21] of the order statistics [22, 23] $X_{(i)}$, $i=1, \ldots N$ of the sample:

$$b_0 = \frac{\sum_{i=1}^{N} X_{(i)}}{N} \qquad (11)$$

$$b_r = \frac{\sum_{i=r+1}^{N} \frac{(i-1)(i-2)\ldots(i-r)}{(n-1)(n-2)\ldots(n-r)} X_{(i)}}{N} \qquad (12)$$

The $L$-moments are then expressed as linear combinations of the probability weighted moments:

$$l_1 = b_0 \qquad (13)$$

$$l_2 = 2b_1 - b_0 \qquad (14)$$

$$l_3 = 6b_2 - 6b_1 + b_0 \qquad (15)$$

$$l_4 = 20b_3 - 30b_2 + 12b_1 - b_0 \qquad (16)$$

We robustify the BRISQUE features using $L$-moment based statistics. $L$-moments are closely related to $L$-estimators extensively used in robust image filtering theory [22, 23]. The second and fourth $L$-moments are computed from the pointwise statistics of MSCN coefficients corresponding to $(\sigma^2, \alpha)$. Similarly, for each pairwise product of adjacent MSCN coefficients along four orientations, the first $L$-moment, second $L$-moment using only negative products of pairwise MSCN coeffients, second $L$-moment using only positive products of pairwise MSCN coefficients and fourth $L$-moment corresponding to $(\eta, \sigma_l^2, \sigma_r^2, \gamma)$ are estimated. Positive and negative $L$-moments are handled separately since the AGGD model is generally asymmetric. This procedure is performed over 2 scales, as is done in BRISQUE.

### 2.3. Image Quality Inference

A support vector machine regressor (SVR) [24] was used to learn a mapping from feature space to quality scores where two different regressors were trained using model based and robust features respectively. SVRs previously have been used with success for image quality assessment [2, 3]. We used the LIBSVM package [25] to implement the SVR; the radial basis function (RBF) kernel was used for regression. Fig. 1 shows the SVM based learning framework for quality inference using model based/robust BRISQUE features.

### 3. EXPERIMENTS

### 3.1. Robustness to Model Assumption

We used the LIVE IQA database [26] to study whether NSS model based BRISQUE features can be robustified. The database consists of 29 reference images together with 779 distorted images spanning five different kinds of distortions

- JPEG and JPEG2000 (JP2K) compression, additive white Gaussian noise (WN), Gaussian blur (blur) and a Rayleigh fast fading channel distortion (FF). Each image has an associated difference mean opinion score (DMOS) indicative of human judgments of visual quality.

As mentioned earlier, the BRISQUE approach and the robust modification to image quality assessment are learning based. Hence the database needs to be split into training and test sets so that a mapping from features to quality scores can be learned. To achieve this, the LIVE database was divided into 2 randomly chosen subsets with 80% training and 20% test sets ensuring that no overlap existed between train and test content. To make the reported results independent of the selection of training content, this procedure of train-test split was repeated 1000 times and we report the median of the performance scores. The performance measures used are Spearman's rank ordered correlation coefficient (SROCC) and Pearson linear correlation coefficient (LCC). The results are tabulated in Tables 1 and 2. Clearly, robustification leads to better correlation with human judgments for all distortions other than Gaussian blur. Robust statistics appear to work well when the distortion statistics deviate from model assumptions.

### 3.2. Statistical Significance and Hypothesis Testing

Although there exist differences in the median correlations between model based and robust BRISQUE algorithms (see Table 1), these differences may not be statistically relevant. To evaluate the same, we performed hypothesis testing based on the $t$-test [27] on the SROCC values obtained from the 1000 train-test trials. The null hypothesis is that 'the mean correlation for the (row) algorithm is equal to mean correlation for the (column) algorithm with a confidence of 95%'. The alternate hypothesis is that the mean correlation of row is greater than or lesser than the mean correlation of the column. 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), i.e., we could not reject the null hypothesis at the 95% confidence level. We found that robust BRISQUE is statistically superior to model based BRISQUE which suggests that robust BRISQUE is more resilient to deviation in model assumptions.

### 3.3. Robustness to amount of training data

This section addresses the question: How much resilience can robust approaches provide when only a small amount of data is available for training the framework. To conduct this experiment, only 10% of the LIVE database was used for training and 90% for testing, again ensuring that there is no overlap between train and test content. Similar to previous experiments, this procedure of train-test split was repeated 1000

| Approach | JPEG 2000 | JPEG | White Noise | Gaussian Blur | Fast fading | Overall |
|---|---|---|---|---|---|---|
| Model Based | 0.9098 | 0.9594 | 0.9786 | 0.9517 | 0.8832 | 0.9390 |
| Robust | 0.9297 | 0.9629 | 0.9818 | 0.9221 | 0.8885 | 0.9448 |

**Table 1**. SROCC of model based and robust approaches with average human opinion scores on LIVE IQA database.

| Approach | JPEG 2000 | JPEG | White Noise | Gaussian Blur | Fast fading | Overall |
|---|---|---|---|---|---|---|
| Model based | 0.9297 | 0.9734 | 0.9881 | 0.9577 | 0.9200 | 0.9438 |
| Robust | 0.9469 | 0.9777 | 0.9903 | 0.9386 | 0.9295 | 0.9500 |

**Table 2**. LCC of model based and robust approaches with average human opinion scores on LIVE IQA database.

| Distortion type | Model based | Robust |
|---|---|---|
| Additive Gaussian noise | 0.8142 | 0.8275 |
| Additive noise in color components | 0.7477 | 0.7693 |
| High frequency noise | 0.6749 | 0.7268 |
| Impulse noise | 0.6016 | 0.6498 |
| Quantization noise | 0.7147 | 0.7274 |
| Gaussian blur | 0.8760 | 0.8484 |
| Image denoising | 0.6431 | 0.6935 |
| JPEG compression | 0.9030 | 0.9284 |
| JPEG2000 compression | 0.9042 | 0.9030 |
| Overall | 0.7225 | 0.7087 |

**Table 4**. SROCC of model based and robust approaches with average human opinion scores on TID database using the model obtained by training using images from LIVE database.

| Distortion type | Model Based | Robust |
|---|---|---|
| Additive Gaussian noise | 0.8042 | 0.8283 |
| Additive noise in color components | 0.7755 | 0.7847 |
| High frequency noise | 0.7066 | 0.7200 |
| Impulse noise | 0.6004 | 0.6521 |
| Quantization noise | 0.7144 | 0.7523 |
| Gaussian blur | 0.8765 | 0.8589 |
| Image denoising | 0.7366 | 0.7480 |
| JPEG compression | 0.9382 | 0.9622 |
| JPEG2000 compression | 0.9070 | 0.9077 |
| Overall | 0.7178 | 0.7460 |

**Table 5**. LCC of model based and robust approaches with average human opinion scores on TID database using the model obtained by training using images from LIVE database.

times and the median of the SROCC performance scores was computed as shown in Table 3. Clearly, robust approaches fare better in this scenario. This is a very useful result in the spirit of time and money expenses required to obtain human opinion scores.

### 3.4. Robustness to Distortion Variations

In this set of experiments, we show that robustification can lead to better correlation with human judgements when the IQA algorithm is tested on different but related distortions than what it was trained on. We trained our model based and robust approaches using the entire LIVE IQA database [26] and then tested them on the distorted images in the TID database [28]. The TID database consists of 25 reference images and 1700 distorted images over 17 distortion categories. We only tested on distortions related to those in the LIVE database - additive Gaussian noise, additive noise in color components, high frequency noise, impulse noise, quantization noise, Gaussian blur, image denoising, JPEG compression and JPEG2000 compression. Also, since our models are based on NSS, we tested our approach only on the natural images in TID. The results of applying BRISQUE on TID are tabulated using SROCC and LCC in Tables 4 and 5. The results indicate that robustification improves correlations with human judgments for all distortions other than Gaussian blur. A likely explanation for the slightly reduced performance on Gaussian blur is that the DSS closely follow the model assumptions, whereas the robust model tends to work well when

the DSS deviate from them. Also, the performance improvement is greater for distortions which are not present in the LIVE database.

### 4. CONCLUSION AND FUTURE WORK

We proposed a method of robustification of model based approaches to image quality assessment using $L$-moments. Better correlation with human judgments of image quality are achieved when distortions deviate from natural scene statistics models. However, model based approaches should be used when it is known what distribution is induced by distortion. We also showed that robust methods perform better when only small amount of training data is available for learning the framework. Also, robust methods can make an IQA model robust to variations in the distortions.

Future work will involve a deeper robust analysis, using classical measures, such as efficiency, of robust methods against model based approaches.

## References

[1] Z. Wang and A. C. Bovik, "Reduced-and no-reference image quality assessment," *IEEE Sig Process Mag*, vol. 28, pp. 29–40, 2011.

[2] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, 2012.

| Approach | JPEG 2000 | JPEG | White Noise | Gaussian Blur | Fast fading | Overall |
|---|---|---|---|---|---|---|
| Model Based | 0.5719 | 0.8133 | 0.9376 | 0.7709 | 0.7411 | 0.7621 |
| Robust | 0.7684 | 0.8896 | 0.9518 | 0.8102 | 0.7567 | 0.8365 |

**Table 3**. SROCC of model based and robust approaches with average human opinion scores on LIVE IQA database using 10-90 train test split.

[3] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From scene statistics to perceptual quality," *IEEE Trans Image Process*, vol. 20, pp. 3350–3364, 2011.

[4] M. A. 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.

[5] 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*, vol. 20, pp. 2089–2098, 2011.

[6] P. Ye and D. Doerman, "No-reference image quality assessment based on visual codebook," in *IEEE Int'l Conf Image Process*, 2011.

[7] H. Tang, N. Joshi, and A. Kapoor, "Learning a blind measure of perceptual image quality," in *IEEE Conf Comput Vision Pattern Recog*, June 2011.

[8] P. J. Besl, J. B. Birch, and L. T. Watson, "Robust window operators," *Mach. Vis. and Appl.*, vol. 2, no. 4, pp. 179–191, 1989.

[9] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, no. 6, pp. 721–741, 1984.

[10] D. S. Chen and B. G. Schunck, "Robust statistical methods for building classification procedures," in *Proc. Int. Workshop on Robust Computer Vision*, 1990, pp. 72–85.

[11] M. Black, G. Sapiro, D. Marimont, and D. Heeger, "Robust anisotropic diffusion," *IEEE Trans. Image Process.*, vol. 7, no. 3, pp. 421–432, 1998.

[12] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans Image Process*, vol. 15, pp. 430 – 444, 2006.

[13] Z. Wang, G. Wu, H. R. Sheikh, E. Simoncelli, E. Yang, and A. Bovik, "Quality-aware images," *IEEE Trans Image Process*, vol. 15, pp. 1680–1689, 2006.

[14] R. Soundararajan and A. C. Bovik, "RRED indices: Reduced reference entropic differencing for image quality assessment," *IEEE Trans Image Process*, 2012 (to appear).

[15] D. L. Ruderman, "The statistics of natural images," *Network computat. in neural sys.*, vol. 5, no. 4, pp. 517–548, 1994.

[16] K. Sharifi and A. Leon-Garcia, "Estimation of shape parameter for generalized Gaussian distributions in subband decompositions of video," *IEEE Trans. Circ. and Syst. Vid. Tech.*, vol. 5, no. 1, pp. 52–56, 1995.

[17] N. E. Lasmar, Y. Stitou, and Y. Berthoumieu, "Multiscale skewed heavy tailed model for texture analysis," in *IEEE Internat'l Conf Image Process*, 2009, pp. 2281–2284.

[18] J. R. M. Hosking, "L-moments: Analysis and estimation of distributions using linear combinations of order statistics," *Jour. Royal Stat. Soc.*, pp. 105–124, 1990.

[19] ——, "Moments or *L*-moments? an example comparing two measures of distributional shape," *The American Stat.*, vol. 46, pp. 186–189, 1992.

[20] ——, "On the characterization of distributions by their *L*-moments," *Jour. Stat. Plan Inf.*, pp. 193–198, 2006.

[21] J. A. Greenwood, J. M. Landwehr, N. C. Matalas, and J. R. Wallis, "Probability weighted moments: Definition and relation to parameters of several distributions expressable in inverse form," *Water Resources Res.*, vol. 15, pp. 1049–1054, 1979.

[22] A. C. Bovik, T. S. Huang, and D. C. Munson, "A generalization of median filtering using linear combinations of order statistics," *IEEE Trans. Acous., Speech, and Sig. Proc.*, vol. ASSP-31, pp. 1342–1350, 1983.

[23] H. G. Longbotham and A. C. Bovik, "Theory of order statistic filters and their relationship to linear FIR filters," *IEEE Trans. Acous., Speech, and Sig. Proc.*, vol. ASSP-37, pp. 275–287, 1989.

[24] B. Schölkopf, A. J. Smola, R. C. Williamson, and P. L. Bartlett, "New support vector algorithms," *Neural Comput.*, vol. 12, no. 5, pp. 1207–1245, 2000.

[25] C. C. Chang and C. J. Lin, "LIBSVM: a library for support vector machines," 2001.

[26] 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, pp. 3440–3451, 2006.

[27] D. Sheskin, *Handbook of parametric and nonparametric statistical procedures*, 2004.

[28] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti, "TID2008-A database for evaluation of full-reference visual quality assessment metrics," *Adv. Modern Radioelect.*, vol. 10, pp. 30–45, 2009.