Blind Image Quality Assessment without Training on Human Opinion Scores

Anish Mittal, Rajiv Soundararajan, Gautam S. Muralidhar, Alan C. Bovik, and Joydeep Ghosh

The University of Texas at Austin Austin,Texas-78712,USA.

ABSTRACT

We propose a family of image quality assessment (IQA) models based on natural scene statistics (NSS), that can predict the subjective quality of a distorted image without reference to a corresponding distortionless image, and without any training results on human opinion scores of distorted images. These 'completely blind' models compete well with standard non-blind image quality indices in terms of subjective predictive performance when tested on the large publicly available 'LIVE' Image Quality database.[1]

1. INTRODUCTION

According to a survey in 2011, Americans captured 80 billion digital photographs and this number is increasing annually.[2] Numbers of photos posted daily on facebook exceed more than 250 million. Consumers are overwhelmed with the amount of digital visual content and finding ways to review and control of the content quality is becoming increasingly difficult. At the same time, the emergence of new handheld devices such as smart phones and tablets, and video streaming sites such as Hulu, Netflix, YouTube etc. is increasing the mobile video traffic and pushing the limited amount of spectrum to its limits. According to the CISCO Visual Networking Index (VNI)[3] forecast, mobile video traffic accounts for nearly 50% of mobile traffic, and by 2015, this percentage is expected to increase to more than 75%. Hence there is a great demand for fast and practical approaches for image and video quality algorithms that can match human judgments of visual quality.

Various quantitative models of visual quality have been proposed depending on the amount of information available. A QA model is called a full reference model if it has access to both the distorted image whose quality needs to be assessed and the corresponding exemplar reference image.[1] A QA model is called a reduced reference model if only partial information about the corresponding exemplar reference image is available to assess the quality of the distorted image.[4] Given human opinion scores on a database of distorted images, models can learn to predict human opinion via machine learning.[5–10] We call these kinds of approaches 'opinion aware'. Models having available at least one of the above forms of information have been shown to predict human judgements with high accuracy.[11]

Blind image quality assessment models that can operate without knowledge of distortion types, reference image, and human opinion scores are of great interest. Continuously obtaining access to human opinion scores is an expensive process. Thus, algorithms that do not rely on human scores can be very useful. Further, building a learning based model with access to distorted images might prove to be inadequate, since the space of all possible distortions an image might be afflicted with is very large.

We propose a family of such NSS driven blind IQA algorithms. A blind IQA model is 'opinion unaware' if it can successfully predict human subjective judgements of distorted images without the aid of human opinion scores and without the corresponding exemplar reference images. A blind IQA model is 'distortion unaware' if it can successfully predict human subjective judgements of distorted images without any samples of distorted images and without the corresponding exemplar reference images. Hence, a 'distortion unaware', 'opinion unaware' blind IQA model is one that can successfully predict human subjective judgements of distorted images without knowing the human opinion scores, sample of distorted images, and the corresponding exemplar reference images. However, we do allow the algorithm to learn from a general database of exemplar natural images that have not been subjected to distortion.

Further author information: (Send correspondence to Anish Mittal) : E-mail: mittal.anish@gmail.com

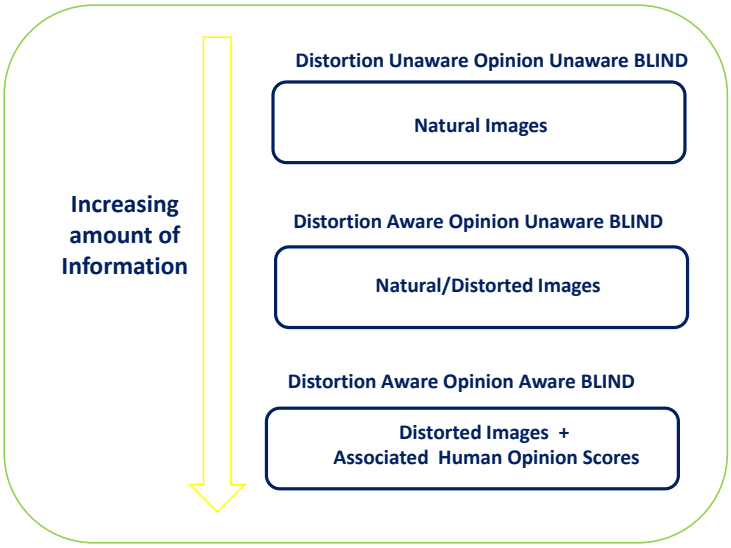

Figure 1. Blind image quality assessment models requiring different amounts of prior information.

If additional information in the form of exemplar images afflicted with a set of anticipated distortions is also supplied, but without associated human opinion scores, then we call this paradigm 'distortion aware' and 'opinion unaware' as the model is still blind to human opinion scores. A similar approach based on modeling the artifacts caused by distortions has been proposed elsewhere[12] though not extensible to 'distortion unaware' approaches in its present form. Opinion aware approaches are well explored.[5–10] We only focus on the challenge of developing opinion unaware blind approaches in the present work.

A taxonomy of approaches depending on the amount of information available is outlined in Fig. 1.

2. PROPOSED APPROACH

Humans have been exposed to natural scenes over the ages and as a result, the visual system has adapted itself to extract useful information from natural scenes. Conversely, natural scene statistic models have proved to be efficient tools for developing models of the responses of cortical neurons and for formulating models of visual processing.[13] By quantifying the 'unnaturalness' caused by a distortion process, these models have been successful ingredients of algorithm that can accurately predict human judgements of visual quality[11] .

We now describe our new 'completely blind' IQA models. Images are decomposed by an energy compacting filter bank and then divisive normalized, yielding responses well-modeled using NSS. Depending on the type of approach, either features computed from exemplar images alone, or from both exemplar and distorted images are used to create distributions over visual words. Quality prediction is accomplished by computing the Kullback-Leibler (KL) divergence between the visual word distribution of the distorted image and the signature visual word distribution of the space of exemplar images.

2.1 Quality Aware Features

We use perceptually relevant NSS features that are used in the Blind/Referenceless Image Spatial QUality Evaluator (BRISQUE) image quality index[5] to get a feature representation for every image patch. The principle behind the approach is that regularities in the statistical properties of natural images[14] are disturbed by the presence of distortions.[5,15] Quantifying such deviations has been shown to be useful for assessing the perceptual quality of images.[4,6,7,16,17]

BRISQUE uses classical spatial NSS model[14] that preprocesses the image by local mean removal followed by divisive normalization:

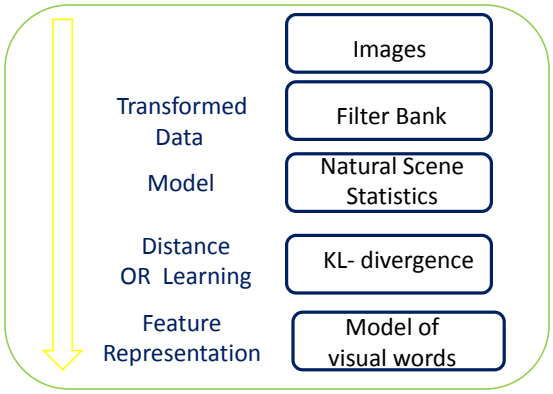

Figure 2. Flow diagram of 'completely blind' IQA model.

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

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

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

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

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. A GGD (Generalized Gaussian Model) distribution[18] 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:

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

where

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

and $\Gamma(\cdot)$ 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, σ^2), can be reliably estimated using the moment-matching based approach proposed in.[18]

The signs of the transformed image coefficients (1) have been observed to follow a fairly regular structure. However, distortions disturb this correlation structure.[5] 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 [5]: $\hat{I}(i,j)\hat{I}(i,j+1), \hat{I}(i,j)\hat{I}(i+1,j), \hat{I}(i,j)\hat{I}(i+1,j+1)$ 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):[19]

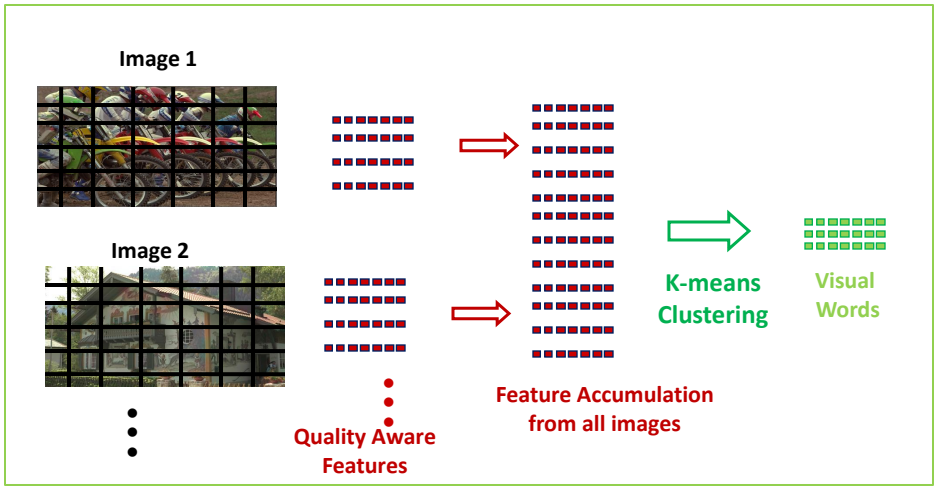

Figure 3. Visual word formation

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

.

where

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

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

The parameters of the AGGD $(\gamma, \sigma_l^2, \sigma_r^2)$ can be efficiently estimated using the moment-matching based approach in.[19] Mean of the distribution is also used as a feature:

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

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 Visual Word Vocabulary

We adopt the approach of[12] to create a visual word vocabulary. A similar approach has also been used for object detection by Sivic *et al.*,[20] the key difference being that appearance based features are used in,[20] while we use perceptually relevant quality based features. Similarly, an IQA algorithm is designed in,[9] where visual words are formed using Gabor based local appearance descriptors which are trained on human scores. However, unlike,[9] the algorithms we developed are 'human opinion' and 'reference image' free.

In the 'distortion unaware' 'opinion unaware' paradigm, we do use a collection of 'exemplar' distortionless images. In order to create the visual vocabulary from these images, we form 'exemplar' visual words by clustering

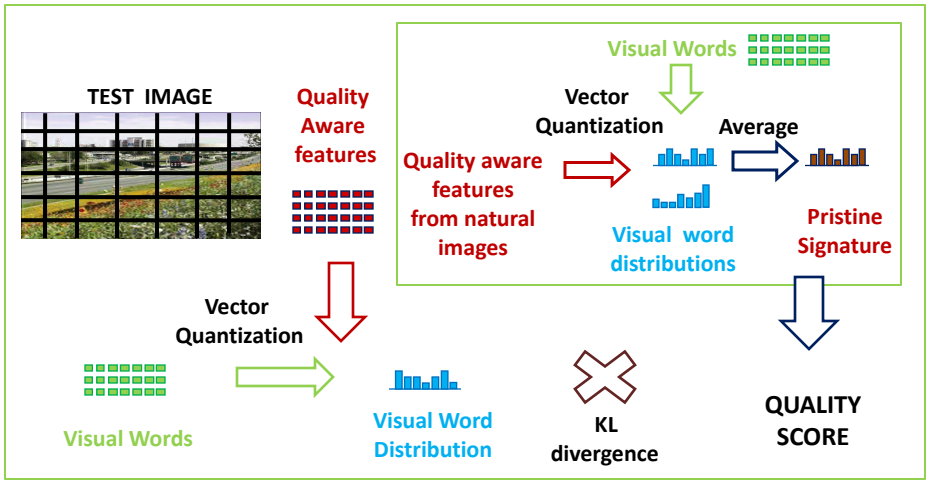

Figure 4. Image quality inference

the BRISQUE features computed from multiple patches in each image belonging to the collection. In this work, we set the number of visual words to 1000 although it is possible that this number could be significantly reduced. The performance variation with the number of visual words has not been studied here and is a subject of future work.

In the case of 'distortion aware' 'opinion unaware' QA, a sample of images afflicted with anticipated distortions are also known. However, there is no knowledge of either the labels of individual distorted images or human opinion scores. Distortions with varying degrees of severity are introduced to the exemplar images and a separate collection of distorted images is constructed. Then, 'distortion aware' visual words are learned by clustering the BRISQUE features computed from multiple patches in each image belonging to an assortment of exemplar and distorted images. The number of visual words is again set to 1000.

To create the visual word vocabulary, each image is divided into overlapping patches of size 64×64, with an overlap of 8×8 between neighboring patches. Local BRISQUE features are then computed from each patch. No significant difference was observed in performance when the patch size was changed to 32×32, with an overlap of 8×8 between neighboring patches. The k-means clustering algorithm with the squared euclidean distance metric was used to cluster the features. The procedure for creating the visual words is also depicted in Fig. 3.

2.3 Image Quality Inference

The visual word vocabulary is now used to vector quantize every patch in all of the 'exemplar' images in the collection. The empirical probability distributions of visual word occurrences, characteristic of natural images, is obtained for every 'exemplar' image in the collection. We average the empirical distributions thus obtained over all 'exemplar' images to obtain a signature probability distribution of natural exemplar images. For a given distorted image, we perform vector quantization similarly and compute its empirical distribution over visual words. The distance of this distribution from the signature distribution of natural images is interpreted as a measure of 'unnaturalness' or distortion in the image. We use the KL divergence as the distance measure to infer the quality of the distorted image. For the 'distortion unaware' 'opinion unaware' model, the visual word vocabulary formed by clustering patches from only exemplar pristine images is used, while for the distortion aware, opinion unaware model, the visual word vocabulary formed by clustering patches from an assortment exemplar pristine images and distorted images is used. Note that the signature probability distribution has a length of 1000 for both 'distortion unaware' 'opinion unaware' blind and 'distortion aware' 'opinion unaware' blind. The algorithm is summarized in Fig. 4.

Approach	JPEG 2000	JPEG	White Noise	Gaussian Blur	Fast fading	Overall
PSNR	0.8951	0.8812	0.9853	0.7812	0.8904	0.8754
MS-SSIM	0.9627	0.9815	0.9733	0.9542	0.9471	0.9513
Distortion Unaware Opinion Unaware blind	0.7797	0.8483	0.9400	0.7941	0.7969	0.8111
Distortion Aware Opinion Unaware blind	0.8825	0.9083	0.9703	0.9098	0.8245	0.8861

Table 1. Spearman rank ordered correlation coefficient of Distortion Aware Opinion Unaware blind, Distortion Unaware Opinion Unaware blind, PSNR, MS-SSIM approaches with average human opinion scores on LIVE IQA database.

Approach	JPEG 2000	JPEG	White Noise	Gaussian Blur	Fast fading	Overall
PSNR	0.8995	0.8899	0.9861	0.7837	0.8897	0.8723
MS-SSIM	0.9686	0.9829	0.9844	0.9565	0.9466	0.9489
Distortion Unaware Opinion Unaware blind	0.7837	0.8729	0.9481	0.8070	0.8074	0.7945
Distortion Aware Opinion Unaware blind	0.8894	0.9238	0.9598	0.9098	0.8276	0.8790

Table 2. Pearson linear correlation coefficient of Distortion Unaware Opinion Unaware blind, Distortion Aware Opinion Unaware blind, PSNR, and MS-SSIM approaches with average human opinion scores on LIVE IQA database.

3. EXPERIMENTS

We conducted experiments on the LIVE IQA database,[1] which contains 29 reference images and 5 distortion types - JPEG, JPEG 2000 (JP2K), Blur, White Noise and Fast Fading (FF) with a total of 779 distorted images. We also used a set of 500 exemplar images of varied sizes ranging from 480×320 to 1280×720 taken from copyright free Flickr data and from the Berkeley image segmentation database,[21] to construct the visual word vocabularies and to obtain the signature 'exemplar' visual word distribution. Similar kinds of distortions as anticipated in the test set - JPEG 2000, JPEG, White Noise, Gaussian Blur, and Fast fading channel errors were introduced in each image at varying degrees of severity resulting in 500 distorted images in each distortion category totaling to 2500 distorted images.

Table 1 and 2 show the Spearman rank ordered correlation and Pearson linear correlation of the image QA indices we developed with difference mean opinion scores (DMOS) on the LIVE database respectively. Both 'distortion unaware' and 'distortion aware' 'opinion unaware' models compete well with standard non-blind metrics when compared against full reference QA indices such as PSNR and MS-SSIM.[22] Observe that the 'distortion aware' 'opinion unaware' algorithm is as good as PSNR. We did not compare the performance with 'opinion aware' approaches due to the absence of human scores for our training dataset.

4. CONCLUSION AND FUTURE WORK

We introduced two new paradigms for blind IQA, 'distortion aware/unaware' 'opinion unaware' QA. We designed NSS based algorithms by measuring the distance between the visual word distributions of the given distorted image and the space of exemplar images.

While the paper presents an example of how to approach these blind IQA problems, there is scope for improving certain aspects of the algorithm. In particular, the creation of visual word distributions from the features is a lossy process. Measuring the distance between the distorted image and the space of exemplar images without such lossy processing could help improve performance.

REFERENCES

[1] Sheikh, H. R., Sabir, M. F., and Bovik, A. C., "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans Image Process* **15**, 3440–3451 (2006).

[2] "Image obsessed," *National Geographic* **221**, 35 (2012).

[3] "CISCO VNI Report." http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360_ns827_Networking_Solutions_White_Paper.html.

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

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

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

[7] Saad, M. A., Bovik, A. C., and Charrier, C., "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE Trans. Image Process.* **21**(8), 3339–3352 (2012).

[8] Shen, J., Li, Q., and Erlebacher, G., "Hybrid no-reference natural image quality assessment of noisy, blurry, JPEG2000, and JPEG images," *IEEE Trans Image Process* **20**, 2089–2098 (2011).
[9] Ye, P. and Doerman, D., "No-reference image quality assessment based on visual codebook," in [*IEEE Int'l Conf Image Process*], (2011).
[10] Tang, H., Joshi, N., and Kapoor, A., "Learning a blind measure of perceptual image quality," in [*IEEE Conf Comput Vision Pattern Recog*], (June 2011).
[11] Wang, Z. and Bovik, A., "Reduced-and no-reference image quality assessment," *IEEE Signal Process Mag* **28**, 29–40 (2011).
[12] Mittal, A., Muralidhar, G. S., Ghosh, J., and Bovik, A. C., "Blind image quality assessment without human training using latent quality factors," in [*IEEE Signal Process Lett*], **19**, 75–78 (2011).
[13] Simoncelli, E. and Olshausen, B., "Natural image statistics and neural representation," *Ann Review Neurosci* **24**, 1193–1216 (2001).
[14] Ruderman, D. L., "The statistics of natural images," *Network computation in neural systems* **5**(4), 517–548 (1994).
[15] Moorthy, A. K. and Bovik, A. C., "Statistics of natural image distortions," in [*IEEE International Conference on Acoustics Speech and Signal Processing*], 962–965, IEEE (2010).
[16] Sheikh, H. R. and Bovik, A. C., "Image information and visual quality," *IEEE Trans Image Process* **15**, 430 – 444 (2006).
[17] Soundararajan, R. and Bovik, A. C., "RRED Indices: Reduced Reference Entropic Differencing Framework for Image Quality Assessment," in [*International Conference on Acoustics, Speech and Signal Processing*], (2011).
[18] Sharifi, K. and Leon-Garcia, A., "Estimation of shape parameter for generalized Gaussian distributions in subband decompositions of video," *IEEE Trans. Circ. and Syst. Vid. Tech.* **5**(1), 52–56 (1995).
[19] Lasmar, N. E., Stitou, Y., and Berthoumieu, Y., "Multiscale skewed heavy tailed model for texture analysis," in [*IEEE Internat'l Conf Image Process*], 2281–2284 (2009).
[20] Sivic, J., Russell, B. C., Efros, A. A., Zisserman, A., and Freeman, W. T., "Discovering objects and their location in images," in [*IEEE Int'l Conf Comput Vision*], 370–377 (2005).
[21] Martin, D., Fowlkes, C., Tal, D., and Malik, J., "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in [*IEEE Int'l Conf Comput Vision*], **2**, 416–423 (2001).
[22] Wang, Z., Simoncelli, E. P., and Bovik, A. C., "Multi-scale structural similarity for image quality assessment," in [*Proceed Asilomar Conf Signals, Syst. Comput.*], (2003).