

COMPUTER VISION AND IMAGE PROCESSING

PROJECT REPORT

Topic : NO-REFERENCE IMAGE QUALITY ASSESSMENT

Date: 26.01.2020

Prepared by: Murat AKOZLU

#### To: dr hab.inż. Krzysztof Okarma

**No-Reference Image Quality Assessment using Blur and Noise**

In recent years, digital camera is equipped in most of the mobile products like cellular phone, PDA and notebook computer. Image quality is the most important criteria to choose mobile products. In some cases, the benchmarks or reviews of products are based on subjective image quality test and thus are dependent on tester and environment. The subjective image quality assessment often misleads the decision for the image quality control parameters of Image Signal Processing (ISP) algorithm. The simple and widely used objective image quality metrics are Mean Square Error (MSE) and Peak Signal-to-Noise Ratio (PSNR). But both of them are known not to be well correlated with human perceptual visual quality and need the original reference image. However, it is not always possible to get the reference images to assess image quality. Human observers can easily recognize the distortion and degradation of image without referring to the original image. Therefore, there is absolutely necessary to develop objective quality assessment that correlates well with human perception without the reference image (No-Reference). In this paper, we propose a method for image quality assessment based on ratio and mean factors of edge blurriness and noise. The proposed quality assessment obtains excellent correlation with subjective image quality scores. There is high correlation between image quality factors and subjective quality scores. The rest of the paper is organized as follows: Section II discusses the related work showing the reason why objective image quality assessment is important and necessary. In section III, the new feature extraction algorithm is proposed. We present experimental results and correlation with subjective image quality assessment in section IV. Finally, Section V draws conclusions and provides future works.

**How it’s work**

In this paper, I prepared NR method which accounts only blur and noise. Although image quality is affected by many features like hue, edge, noise, and contrast, I assume that noise and blur are the most important factors on image quality degradation. The proposed work searches and quantifies the blur and noise as image quality factors.

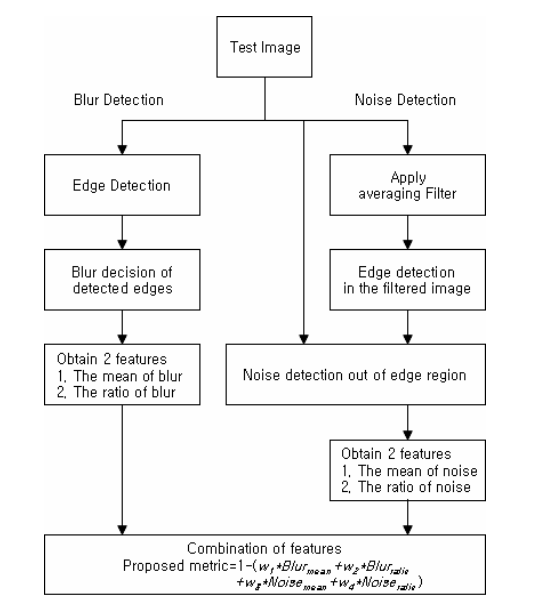


Figure 1-) Flow chart about estimation of Blur & Noise

Most of the digital cameras have the Image Signal Processor (ISP) to enhance the output of image sensor. One of the important functions of ISP is to remove noise: noise reduction. The strong noise reduction removes noise sufficiently but makes detail and texture blurred. In order to reduce much trial caused by trade-off between noise reduction and detail loss, the criteria for image quality control parameters are required. In case only one feature between blur and noise is considered for quality prediction, the results are to be insufficient for finding the optimized parameters of noise reduction. The proposed image quality metric meets the necessary criteria because our method analyses both blur and noise simultaneously. The proposed method calculates blur and noise in a spatial domain. Only the luminance parts of the images are used to estimate blurriness and noise. The blur is measured by simple numeric operations on pixel.

**Blur Estimation**

The blurriness is perceptually determined by human observers regardless of the type of blurring, for example, noise reduction, compression, motion blur, and out of focus. In the paper, we seek to find blur without any assumption about its formation. Blur estimation is divided into 2 stages: First is edge detection and second is blur decision. The blur in the paper is estimated by difference between the intensity of current pixel and average of neighbour pixels. The difference is then normalized by the average. Figure shows the blur estimation. If the intensity of center pixel is closer to the average intensity of both side pixels, the center pixel is supposed to be on blurred edge.

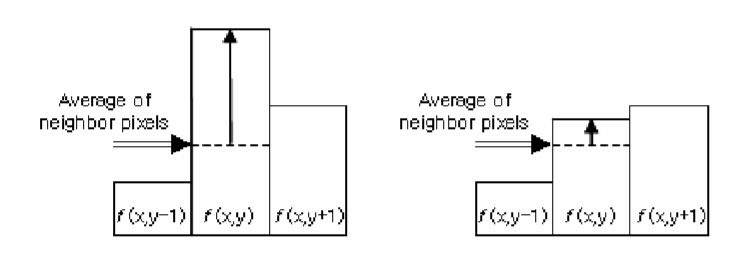
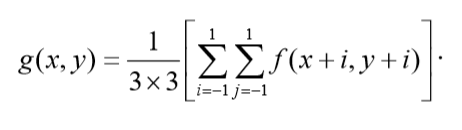


Figure 2

**Noise Estimation**

Since the noise along edges perceptually looks less apparent, we measure the noise out of the edge region. The edge detection can also be affected by noise. Hence, a pre-processing for noise filtering is needed prior to detecting the edge. In the paper, we apply an average filter to the noisy test image to remove the noise. The averaging filtered image g(x,y) is generated by this formula

**BIQI No-Reference Image Quality Assessment**

Objective blind/no-reference (NR) image quality assessment (IQA) refers to algorithms that seek to predict the quality of distorted images without any knowledge of pristine reference images and that correlate well with human perception of quality. Recently, the field of NR IQA has seen a significant rise in activity; however there is considerable room for improvement. This is largely due to the fact that NR IQA is an extremely difficult problem to solve. In fact, only recently has the field of the ‘easier’ full-reference (FR) IQA matured to produce algorithms that correlate well with human perception of quality.

Present day NR IQA algorithms generally assume that the distortion affecting the image is known. For example, there exist NR IQA algorithms that seek to assess the quality of JPEG/JPEG2000 compressed images or blurred images. Here, we propose a new two-step general-purpose framework for designing no-reference image quality indices based on natural scene statistic (NSS) models of images. The two steps are image distortion classification based on a measure of how the NSS are modified, followed by quality assessment, using an algorithm specific to the decided distortion. Once trained, an algorithm of the proposed framework does not require further foreknowledge of the distortion affecting the images to be assessed. The framework is modular in that it can be extended to any number of distortions.

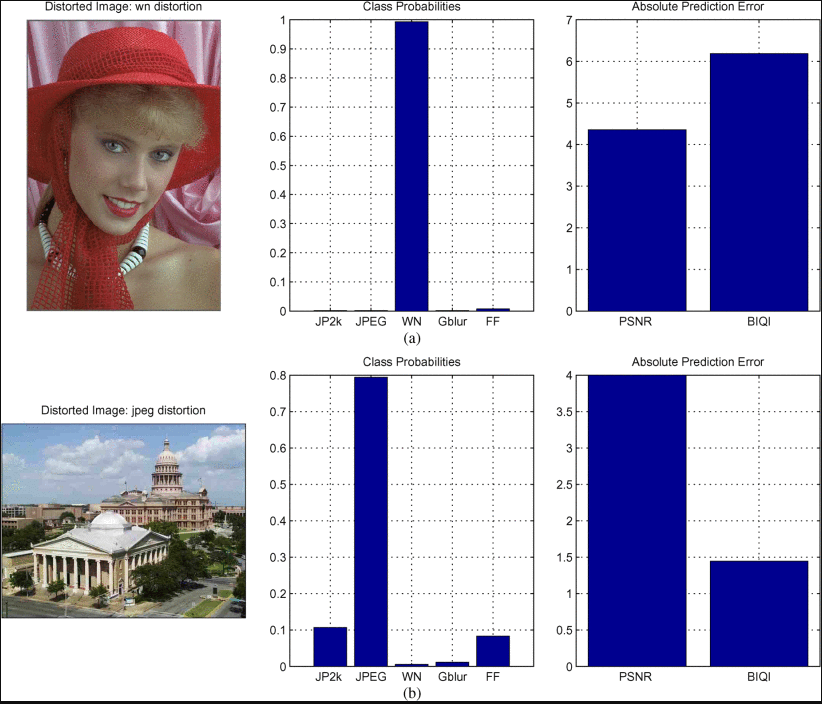
The works closest in concept to ours are those proposed in for video quality assessment (VQA). In both these cases, a combination of techniques were used to measure blockiness, blur, corner outliers and noise were combined into a quality score by a Minkowski sum. The authors used a set of pre-fixed thresholds as well as parameters obtained using a training set in order to produce these scores, which were then tested on a set of videos with a combination of the considered distortions. Our work differs from that in and in several ways. We do not explicitly seek to characterize the structure of blockiness and other distortions using local filters, but instead utilize concepts from NSS to produce an easily extensible approach to other distortions. We utilize a unique 2-stage strategy which identifies the likeliest distortion in the image and then quantifies this distortion using a NSS-based approach. Finally, we test our algorithm on the publicly available LIVE IQA dataset.

Thus, we present a new and unique framework for no-reference image quality index design. The aim is to produce algorithms that are truly ‘no-reference’—no information about the distortion affecting the image is contained within the distorted image or otherwise made available to the algorithm. In order to create such a no-reference algorithm, we utilize techniques from NSS.

It has been demonstrated that subband responses of natural scenes tend to follow a non-Gaussian (heavy-tailed) distribution, which can be parametrized. Given that this is true, a pertinent question is whether there exists such a general statistical description for distortions. Even though researchers have observed that distortions affect scene statistics, our aim is to assess whether such changes in NSS are *systematic* and *parameterizable*. Here, we show that not only do distortions affect NSS, they are also systematic and parameterizable.

A description of distorted image statistics (DIS) when obtained, can be used as a *distortion-specific signature* for classifying an image into a particular distortion category. Once such classification is achieved, it is as if the algorithm is aware of the distortion. The algorithm can then deploy a distortion-specific IQA algorithm. Although this may sound simple, there exist many subtleties that we explore in this paper. For example, our description here assumes that the distortion classes are disjoint—an assumption that is not always true. Here, we propose a novel scheme to handle such subtleties. Further, although there exist many distortion-specific IQA algorithms which we can use after the initial classification, we develop a general framework for IQA based on DIS. Indeed, this approach to QA does not measure distortion-specific indicators such as blocking, but provides a modular strategy that adapts itself to the distortion in question.

The first contribution of this work is the development of a novel two-stage modular framework for distorted image statistics based no-reference quality assessment. The modularity of the proposed framework implies that this approach is truly extensible in that any addition of distortion categories beyond those discussed here may be easily accomplished. The second contribution is an innovative strategy to assess quality when the distortion is known. A combination of these two contributions results in a no-reference image quality index. In this paper, we describe a a specific algorithm of our proposed framework—the blind image quality index (BIQI)—and test its performance by the construction of an example 5-distortion algorithm on the LIVE image database.



**CONCLUSION**

In this letter, we described a framework for constructing an objective no-reference/no-reference (NR) image quality assessment (IQA) measure. The framework is unique, since it assesses the quality of an image completely blind—i.e., without any knowledge of the source distortion. This is achieved by using distorted image statistics (DIS)—an extension of natural scene statistics for distorted images. In this paper, we discussed DIS, and demonstrated that each distortion has a unique signature which can be characterized by the use of DIS and used this signature to classify images into distortion categories. We also described how distortion-aware IQA may be undertaken using DIS. We then combined distortion-classification with distortion-aware IQA to produce a demonstration of the blind image quality index (BIQI) which is of value on its own. BIQI was tested on the LIVE image database and was shown to perform well in terms of correlation with human perception. Indeed, BIQI, an NR measure, performed competitively with (and in many cases beat) PSNR, an FR measure, across all distortions and in overall performance. A software release of BIQI has been made available at [9] to further research in the are of NR IQA.

**BRISQUE No-Reference Image Quality Assessment**

We propose a natural scene statistic-based distortion-generic blind/no-reference (NR) image quality assessment (IQA) model that operates in the spatial domain. The new model, dubbed blind/referenceless image spatial quality evaluator (BRISQUE) does not compute distortion-specific features, such as ringing, blur, or blocking, but instead uses scene statistics of locally normalized luminance coefficients to quantify possible losses of “naturalness” in the image due to the presence of distortions, thereby leading to a holistic measure of quality. The underlying features used derive from the empirical distribution of locally normalized luminances and products of locally normalized luminances under a spatial natural scene statistic model. No transformation to another coordinate frame (DCT, wavelet, etc.) is required, distinguishing it from prior NR IQA approaches. Despite its simplicity, we are able to show that BRISQUE is statistically better than the full-reference peak signal-to-noise ratio and the structural similarity index, and is highly competitive with respect to all present-day distortion-generic NR IQA algorithms. BRISQUE has very low computational complexity, making it well suited for real time applications. BRISQUE features may be used for distortion-identification as well. To illustrate a new practical application of BRISQUE, we describe how a nonblind image denoising algorithm can be augmented with BRISQUE in order to perform blind image denoising. Results show that BRISQUE augmentation leads to performance improvements over state-of-the-art methods.

**No-Reference Image Quality Assessment Algorithms on Printed Images**

1.Introduction

Despite rapid developments in electronic media, most people still prefer reading text printed on paper rather than reproduced on electronic displays. Printed media remain more suitable for delivering local news than electronic media, and the packaging industry increasingly relies on the production of visually pleasing and personalized packages by using digital printing. In addition, an increasing number of images are captured each year, and despite the fact that the digitization has created novel ways to share and distribute images, printed images still have their users. For example, the amount of bound photobooks has grown rapidly during the recent years. These, among other reasons, are why paper and other fiber-based products still play an important role in communication, and printed products, such as books, newspapers, and packages, are an important part of daily life. When a customer decides to purchase a printed book or magazine, one of the key factors is print and image quality. Rather than using technical measurements, humans do not evaluate the quality of print and images based on physical parameters, but rather based on personal preferences and what they see as pleasurable.

The problem of how humans perceive the quality of a reproduced image is of interest to researchers in many fields, including optics and material physics, image processing (compression and transfer), printing and media technology, and psychology. The problem is particularly difficult for printed media, since solving it requires understanding the paper and ink physics, viewing parameters, optics, and elements of human visual perception. No measure of visual print quality can be defined without ambiguity, because it is ultimately a subjective opinion of an “end-user” observing the result. As a consequence, visual evaluations have been traditionally conducted using groups of human observers, but recent developments in perceptual models and machine vision have made it possible to develop automatic methods of print quality evaluation. The use of machine vision founded on reliable perceptual and print models promises to make it possible to replace humans in laborious off-line evaluations. In addition, such computational methods suggest the potential for quality-optimized on-line measurements during printing.

Image quality assessment (IQA) models can be divided into three categories: full-reference (FR), reduced-reference (RR), and no-reference (NR) methods. In FR methods, a reference image with presumed ideal quality is available, whereas in RR methods only a small amount of information describing the reference image is given as input. NR methods operate in the absence of any reference image. Currently, FR methods are the main approach for evaluating and comparing the quality of digital images, especially compressed ones. The digital representations of the original and compressed images are in correspondence, i.e., there exist no-spatial transformations between the images, and the compression should retain at least photometric equivalence. Therefore, FR measures can be computed in a straightforward manner by computing “distance metrics,” and the actual problem is to define an appropriate metric for the task. NR-IQA is the most difficult task, and the majority of the proposed methods are designed for a single-distortion type and can be considered as domain specific.

The FR-IQA has been shown to be a good approach to predict the image quality when the reference image exists. With a carefully designed measurement framework, it is possible to apply an FR approach to printed images and relatively high correlations with subjective evaluation results can be achieved, as was shown by Eerola et al. However, several problems exist when the quality of printed images is evaluated by FR methods.

The first obvious weakness is the fact that the FR methods are suitable only when a digital reference image exists. This is not always the case with printed images. Even more notable problems arise from the basic assumption of the FR approach that the reference image is of ideal quality and can be used as a basis for quality evaluation. For quality assessment (QA) of compressed images, this assumption is perhaps justified; a good image compression method reduces the size of the image in such a manner that the visual appearance of the image changes as little as possible, i.e., the evaluated (compressed) image is visually similar to the reference (original) image. For printed images, however, it is not clear that such an assumption applies. First of all, the original image is in a very different form than the printed image that is being evaluated, making its use as a reference image not only difficult, but also rather questionable. It is not clear how the difference between a printed photograph and a digital image should be measured. Second, visual quality may be impaired when the original image is transferred onto paper even if no printing artifacts appear, since different aspects of image quality take different degrees of importance on different media. For example, gloss is not a property of a digital image, but has a remarkable effect on the perceived quality of a printed image. Even in the hypothetical ideal situation, where the original image is of “perfect quality” (whatever that is) and the printer or paper do not cause any visible artifacts, quality may still be compromised after transferring onto paper due to the different natures of the media. Third, while making subjective evaluations of printed samples, showing a digital reference to human observers is not a simple matter, since simultaneous viewing of digital and printed images does not allow an observer to adapt both white points, whereas the memory viewing technique does not allow one to directly compare the images. Often a digital reference image is not shown to the observers, and they are forced to make decisions without knowing what the printed image was supposed to look like. RR-QA algorithms suffer from similar problems since a reference image is still needed.

For the aforementioned reasons, NR-QA methods are of high interest. However, NR-QA is a much more difficult task than FR-QA, and until recently, there did not exist any general NR-QA methods. All methods were either application specific or measured only a specific kind of distortion such as blur or noise. While these methods have a role in QA, no such method alone can predict the perceived quality of an image. During the last few years, significant developments have led to the creation of generic NR-QA algorithms. Thus, the objective of this study is to determine the efficacy of these new NR models for predicting the subjective quality of printed images by statistically evaluating their performances against subjective mean opinion scores (MOSs) obtained from psychometric experiments on printed samples. Since NR models have been developed for images of natural scenes, this study focuses only on the important application of printed photograph quality analysis while the quality of printed text and graphics is not considered.

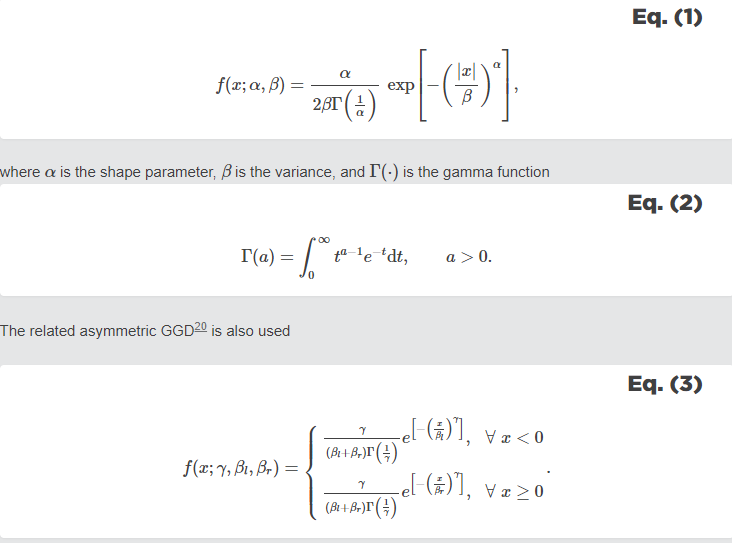
The paper is organized as follows. Section 2 introduces the existing generic NR-IQA algorithms that are statistically evaluated in this study. Section 3 presents the data, a method to apply NR-IQA algorithms to printed images, and the results. The results are discussed in Sec. 4, and the conclusion is drawn in Sec. 5.

**2.NR-IQA**

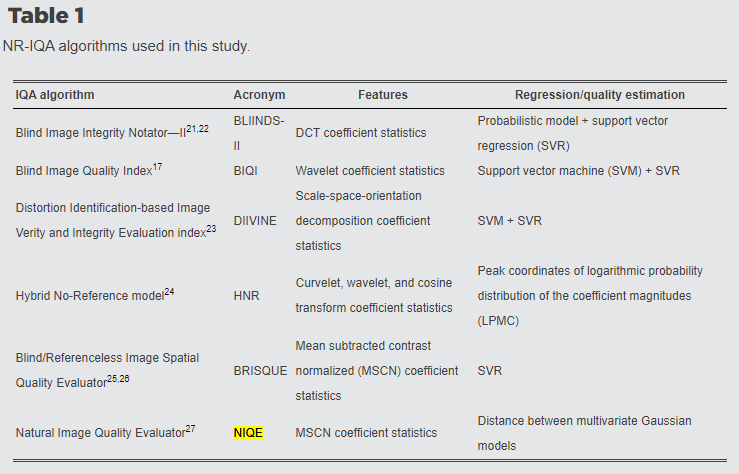
Most existing NR-IQA algorithms fall in one of the following categories: distortion-specific IQA algorithms, training-based IQA algorithms, and natural scene statistics (NSS)-based IQA algorithms. The first category is composed of methods that try to model the distortion such as blur or blockiness. These methods are application specific and are not in the scope of this work. However, it should be mentioned that the distortion-specific methods have also been developed for printed images. The second category contains methods and models that use image-based features and require training on appropriate data. These methods are highly dependent on the quality of the selected features and often require a large amount of training data, i.e.,

distorted images with subjective data that is laborious to collect. Generic image features to describe the image quality are difficult to establish, which limits the application domain. The third category contains methods that are based on the assumption that the pristine natural images form a subset of images that have different statistical properties than the distorted images. NSS methods have turned out to be a very promising approach, and most existing NR-IQA algorithms with good performance more or less rely on NSS. These methods may also require training, but the amount of training is greatly reduced relative to training-based IQA algorithms.

Ideally, NSS features are invariant to image content, but are sensitive to distortions. Such features can be used to assess the image quality by estimating the degree of distortion or the distance of the distorted image to the pristine (natural) image, regardless of the content of the image. Several promising NSS approaches have been proposed in the literature. A typical NSS-based IQA algorithm starts with a multiscale image transform, such as the discrete cosine16 or wavelet transform, but spatial NSS methods also exist. The computed transform coefficients have statistical properties that vary based on the presence of distortions. For example, the distribution of wavelet coefficients computed from natural images usually has a sharp peak near zero and long and smooth tails. Most image distortions break this regularity, which makes the shape of the coefficient distribution a good feature for IQA. The coefficient distributions are often parametrized using the (zero mean) general Gaussian distribution (GGD), which has been found to capture the broad spectrum of possible distribution shapes. The GGD is defined as;

The distribution parameters (αα, ββ) or (γγ, βlβl, βrβr) are then used as features to either classify images based on the decided distortion, then apply a distortion-specific IQA algorithm or estimate the quality directly using regression techniques. Most of the methods are trained on data containing MOSs or difference MOSs (DMOSs) of the images.

The NR-IQA algorithms selected for this study and their basic information are listed in Table 1. All the selected methods are based on NSS. There are also promising training-based methods, such as the learning-based blind image quality (LBIQ) measure and the visual codebook-based image quality (CBIQ) measure, but these were excluded from the study due to their strong dependence on training data. As discussed later in Sec. 3.3, due to the laborious nature of preparing and subjectively evaluating printed samples, data volume remains a problem and limits the selection of the NR-IQA algorithms that can be used.



The Blind Image Quality Index (BIQI) makes use of NSS features based on wavelet coefficients to first classify an image between different distortion types and to estimate distortion-specific quality scores. A support vector machine (SVM) is used for classification, and a support vector regression (SVR) is used for a distortion-specific quality score estimation. The final quality score is computed as a weighted sum of the distortion-specific quality scores. The weights are the probability estimates provided by the SVM used in the distortion classification stage.

The Distortion Identification-based Image Verity and Integrity Evaluation (DIIVINE) index is an extension of BIQI with a richer set of NSS-based features. Instead of the wavelet transform, a scale-space-orientation decomposition is used.

BLIINDS-II is an extension of the Blind Image Integrity Notator (BLIINDS). It uses NSS features based on the local discrete cosine transform coefficients and a simple probabilistic model to predict the quality.

Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) differs from the methods above because it uses only spatial features, albeit over multiple scales. There is no need to map the data to a different coordinate domain, such as the wavelet or DCT domain. The NSS used is based on locally normalized luminances [mean subtracted contrast normalized (MSCN) coefficients].

BRISQUE was further developed leading to the very generic, training-free Natural Image Quality Evaluator (NIQE). It uses similar NSS features, but instead of establishing the quality value directly from features using an SVR, the NSS features are modeled as multivariate Gaussian. The main advantage of NIQE is that, unlike other methods, it does not require training data with subjective human evaluations. Instead, the model is constructed from features drawn from a corpus of undistorted natural images, making it the first truly distortion-independent NR-IQA algorithm.

The hybrid NR (HNR) model is based on curvelet, wavelet, and discrete cosine transform coefficient statistics, and the quality is predicted using the peak coordinates of the logarithmic probability distribution of the coefficient magnitudes (LPMC).

**2.1.Previous Comparisons of NR-IQA Algorithms**

Most of the original references provide the results achieved on the well-known LIVE database. Therefore, reliable conclusions can be made on the performance on the specific distortions present in the LIVE database, i.e., JPEG and JPEG2000 compressions, additive white Gaussian noise, Gaussian blur, and a Rayleigh fast-fading channel distortion. The results show that BRISQUE outperforms the other NR methods with a 0.94 linear correlation against the subjective scores, but LBIQ, BLIINDS-II, DIIVINE, and NIQE also attain good results with better than 0.9 correlation over all distortion types.

**3.Results**

The descreening was performed using six different cut-off wavelengths: 0.05,0.10,…,0.30mm, and seven scale factors: 0.08,0.10,…,0.20 corresponding to 100,125…,250dpi. The NR-IQA algorithms were applied with every cut-off wavelength-scale pair to find the optimal parameter values for each method.

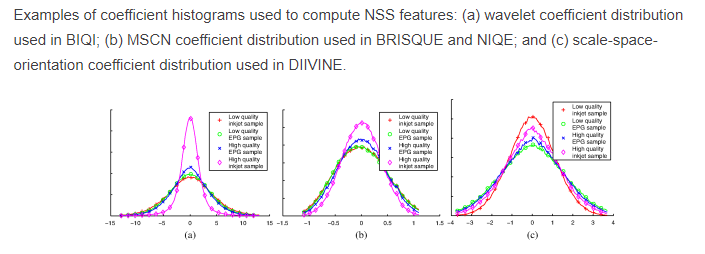
Figure 2 presents the examples of coefficient histograms that were used to compute NSS features. For visualization purposes, the histograms are presented for four samples: low-quality inkjet sample (the sample with lowest MOS in Test Set A), low-quality EPG sample, high-quality EPG sample, and high-quality inkjet sample. Although the two test sets were never combined, the quality variation between the selected samples is high. Most people would prefer the high-quality inkjet sample as the best one, followed by the high quality EPG sample and low-

Figure 2

quality EPG sample, the low-quality inkjet sample being the worst one. As can be seen from Fig. 2(a), the wavelet coefficient distributions used by BIQI seem to have a lower-standard deviation for high-quality samples and a higher one for low-quality samples, which indicates their potential in print QA. The same also applies for the MSCN coefficients used in BRISQUE and NIQE. However, the scale-space-orientation coefficients used in DIIVINE seem to be more suitable for distinguishing printing methods from each other than predicting the quality of printed images.

Figures 3 and 4 present the results of the best NR-IQA algorithm with the optimal cut-off wavelength and scale factor for each image content. As can be seen, the correlations are very high. However, selecting the optimal parameter values for each method produces overly optimistic results. Since the image content is often unknown beforehand in a practical application, the parameters should be fixed in such a manner that the NR-IQA algorithm is not sensitive to image content. Figure 5 shows the optimal parameter values for each method for different image contents, while Table 2 presents the corresponding correlations [linear correlation coefficient (LCC) and Spearman’s rank correlation coefficient (SRCC)] between the MOS and algorithm scores. Also, the parameter values that maximize the mean correlation over all image contents and the corresponding correlation coefficients are presented.

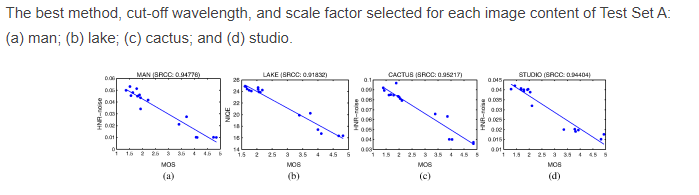


Figure 3

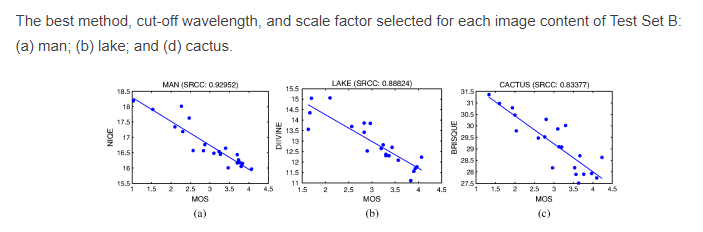
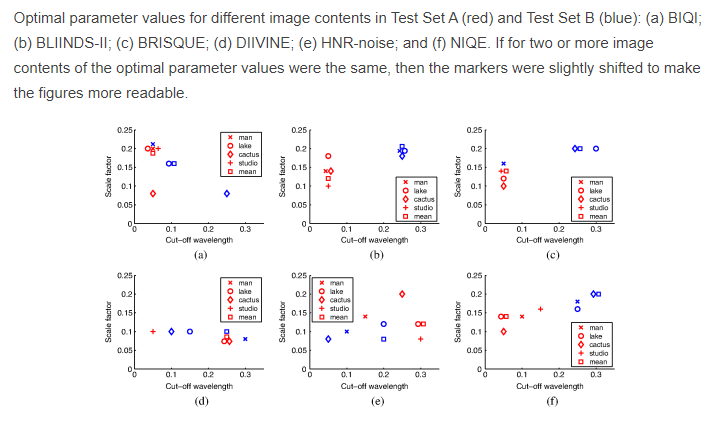
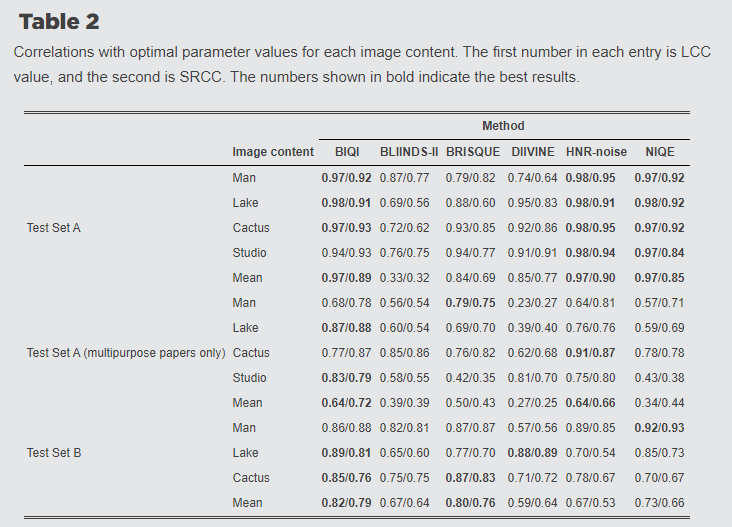


Figure 4

Figure 5



As it can be seen from Fig. 3, Test Set A contains two relatively distinct clusters of different qualities: high-quality photopapers and lower-quality multipurpose papers. Any NR-IQA algorithm that distinguishes the two clusters and places them into the right order gets a high LCC value, while a method that fails to correct select the better cluster gets a high negative correlation coefficient. Moreover, the overall correlation coefficient increases if an algorithm places the cluster further away even if the correlations inside the clusters do not change. Therefore, based on the LCCs obtained using Test Set A, one can only determine whether a method works adequately at a coarse level or not, but it does not reveal the relative performances of the method studied in detail. Hence, when selecting the optimal combination parameter values, SRCC appears to be a more suitable measure.

To further study the performances of the various NR-IQA models on Test Set A using LCC, the samples representing the two clusters were divided into two and an additional test was carried out using the challenging lower-quality multipurpose papers. The result is shown in Fig. 6 and Table 2, and as expected, the correlations are much lower than on the full Test Set A. However, although the quality variation inside the set is very low, clear correlations between the MOS values and the algorithm scores can be observed.

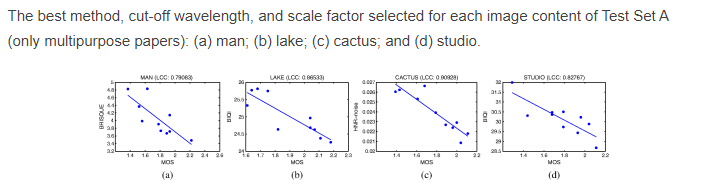


Figure 6

Figures 7 and 8 show the results with optimal parameter values (mean SRCC over all image contents). Based on the figures, it is clear that the image content significantly affects the algorithm scores. Although the within-content correlations are high, different image quality score values for different contents cause a low overall correlation over all contents. However, it should be noted that the subjective evaluation results were separately scaled to the interval 1 to 5 for each content, and the MOS values are not directly comparable between the image contents. Therefore, combining the plots without preprocessing is not a well-grounded approach.

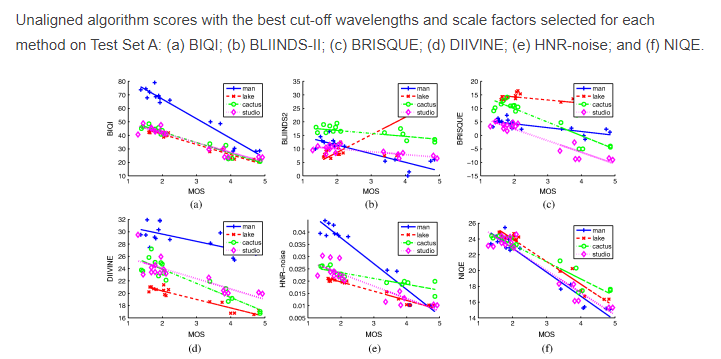


Figure 7

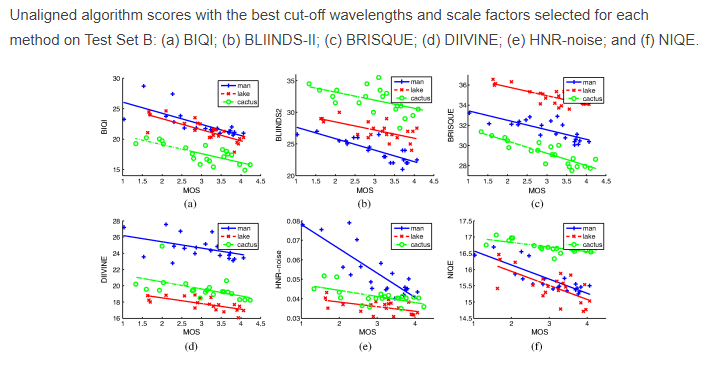


Figure 8

Correlations between the MOS and the aligned algorithm scores from NR-IQA algorithms and from selected FR-IQA algorithms.5 The numbers shown in bold indicate the best results.

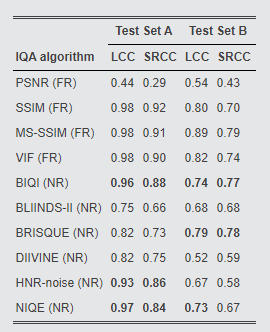


Table 3

The statistical significance of the previous results was studied using the variance test. It expresses the trust in the superiority or inferiority of one QA algorithm over another based on performance measures. The test is based on the assumption that the residuals (difference between MOS and the IQA algorithm score linearly fitted to MOS) are normally distributed. The normality of the residuals was tested using the Lilliefors test at a 5% significance level, and the residuals were shown to follow a normal distribution for all methods in Test Set A and for all methods except DIIVINE in Test Set B. Moreover, since DIIVINE had the lowest LCC in Test Set B, the non-normality has no significant effect on our conclusions. The F-test was used to test whether the variances of the residuals of two QA algorithms are identical, i.e., the QA algorithm residuals are randomly drawn from the same distribution. The null hypothesis is that the residuals of both QA algorithms come from the same distribution and are statistically indistinguishable with 90% confidence. The significance test results for the aligned algorithm scores are shown in Tables 4 and 5 for both test sets and for all possible pairings of QA algorithms.

F-test results for Test Set A: 0 means that the QA algorithms are statistically indistinguishable from each other, 1 means that the IQA algorithm for the row is statistically better than the IQA algorithm in the column, and −1 means that the IQA algorithm in the row is statistically worse than the IQA algorithm for the column.

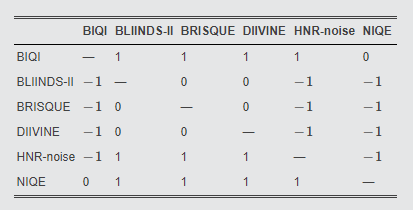


Table 4

F-test results for Test Set B.

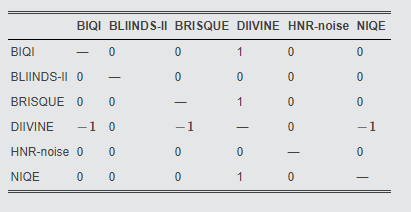


Table 5

**4.Discussion**

As described earlier, the training of IQA algorithms was performed using a separate set of distorted or pristine digital images, and the printed and scanned images were used only for testing. Therefore, the training and testing data were very different from each other and the training data did not contain the printing distortions. It is understandable that this reduced the performance of the learned models and might favor certain IQA algorithms. However, as the results showed, most IQA algorithms were still able to achieve high correlations with MOS, suggesting that the distortions in the training data were close to the distortions in the testing data. Moreover, since the ultimate goal is to develop an NR-IQA algorithm that needs to be trained only once, after which it should be possible to apply it to any appropriate image quality application, it is worthwhile to also evalate the algorithms’ dependency on the quality of the training data. This justifies the use of different training and testing data as it reveals not only the performance of the methods, but also their sensitivity to imperfect training data.

On both test sets, most of the NR methods outperformed PSNR and the best methods were shown to produce almost as good results as state-of-the-art FR-IQA algorithms. As mentioned earlier, Test Set A contains two relatively distinct clusters, making SRCC a more reliable metric for comparing the performance of IQA algorithms. Based on the SRCC results, the best methods are BIQI, NIQE, and HNR-noise and these are also statistically significantly better than the rest of the methods. Test Set B does not contain similar distinguishable clusters, and therefore, LCC also provides useful information. Unlike SRCC, it also measures the linear dependence, i.e., whether a change in the quality at the high-quality end of the scale corresponds to a similar change at the low-quality end. Based on the LCC results, the best methods are BIQI, BRISQUE, and NIQE. The SRCC results support this conclusion. However, it should be noted that no statistically significant differences in performance were found between these methods and BLIINDS-II or HNR-noise.

The notably lower performance of DIIVINE compared with BIQI on both test sets is a surprising result, since DIIVINE is an extension of BIQI and it outperformed BIQI in experiments made using the LIVE database. The results shown in Fig. 2 suggest that the scale-space-orientation decomposition used by DIIVINE is more sensitive to the different printing methods and, therefore, a less suitable approach for the QAs of printed images. Moreover, since it is the more complex of the two (88 NSS features in DIIVINE compared with 18 in BIQI), there is a higher chance of over-tuning, making it more vulnerable to the large differences between training and testing data.

As can be seen in Figs. 7 and 8, image content has a noticeable effect on the algorithm scores. The main reasons for this are the fact that the NR-IQA algorithms were trained using a separate database containing different image contents, and the scaling of the MOS values for each type of image content was performed separately. Scaling the algorithm scores similarly to the MOS values does not really solve the problem, since we do not know if the effect of image content will be eliminated even if the above problems did not exist. Therefore, the results presented in Figs. 9 and 10 can be seen as an upper limit of method performance and to achieve such results, a manual effort is needed. However, if the parameters a and b that define the scaling of the algorithm scores could be predicted based on the image content, then the results presented in Figs. 9 and 10 could be achieved using a fully automatic method.

BIQI and NIQE perform well on both the test sets. While BIQI achieves slightly higher correlations, NIQE contains one significant benefit: it requires only pristine images to be trained. This enables training with a much larger number of images, making it less vulnerable to new image content.

**5.Conclusion**

We applied the leading general-purpose NR-IQA algorithms to printed images and evaluated the performance of several state-of-the-art NR-IQA algorithms on an extensive set of printed photographs. We found that the BIQI and NIQE algorithms outperformed the other QA algorithms. Of these two, NIQE was less sensitive to image content, making it the most promising current method for NR-printed IQA.

**REFERENCES**

1. J. ImaiM. Omodani, “Reasons why we prefer reading on paper rather than displays: studies for seeking paper-like readability on electronic paper,” J. Imaging Sci. Technol., 52 (5), 1 –5 (2008). https://doi.org/http://dx.doi.org/10.2352/J.ImagingSci.Technol.(2008)52:5(051002) JIMTE6 1062-3701 Google Scholar

2. Prophoto, “Trends in the Photo and Imaging Market—Photokina,” (2012) http://www.prophoto-online.de/img/ftp/broschueren/Trends-in-the-photo-and-imaging-market-photokina-2012.pdf June 2014). Google Scholar

3. P. G. Engeldrum, “A theory of image quality: the image quality circle,” J. Imaging Sci. Technol., 48 (5), 446 –456 (2004). JIMTE6 1062-3701 Google Scholar

4. H. R. SheikhM. F. SabirA. C. Bovik, “A statistical evaluation of recent full reference image quality assessment algorithms,” IEEE Trans. Image Process., 15 (11), 3440 –3451 (2006). https://doi.org/http://dx.doi.org/10.1109/TIP.2006.881959 IIPRE4 1057-7149 Google Scholar

5. T. Eerolaet al., “Full reference printed image quality: measurement framework and statistical evaluation,” J. Imaging Sci. Technol., 54 (1), 1 –13 (2010). https://doi.org/http://dx.doi.org/10.2352/J.ImagingSci.Technol.2010.54.1.010201 JIMTE6 1062-3701 Google Scholar

6. U. Rajashekaret al., “Performance evaluation of mail-scanning cameras,” J. Electron. Imaging, 19 (2), 023008 (2010). https://doi.org/http://dx.doi.org/10.1117/1.3421975 JEIME5 1017-9909 Google Scholar

7. K. M. BraunM. D. FairchildP. J. Alessi, “Viewing techniques for cross-media image comparisons,” Color Res. Appl., 21 (1), 6 –17 (1996). https://doi.org/http://dx.doi.org/10.1002/(ISSN)1520-6378 CREADU 0361-2317 Google Scholar

8. C. Liet al., “No-reference blur index using blur comparisons,” Electron. Lett., 47 (17), 962 –963 (2011). https://doi.org/http://dx.doi.org/10.1049/el.2011.0921 ELLEAK 0013-5194 Google Scholar

9. P. V. VuD. M. Chandler, “A fast wavelet-based algorithm for global and local image sharpness estimation,” IEEE Signal Process. Lett., 19 (7), 423 –426 (2012). https://doi.org/http://dx.doi.org/10.1109/LSP.2012.2199980 IESPEJ 1070-9908 Google Scholar

10. C. T. VuT. D. PhanD. M. Chandler, “S3: a spectral and spatial measure of local perceived sharpness in natural images,” IEEE Trans. Image Process, 21 (3), 934 –945 (2012). https://doi.org/http://dx.doi.org/10.1109/TIP.2011.2169974 IIPRE4 1057-7149 Google Scholar

11. Z. WangH. R. SheikhA. C. Bovik, “No-reference perceptual quality assessment of JPEG compressed images,” in Proc. of the Int. Conf. on Image Processing, 477 –480 (2002). https://doi.org/http://dx.doi.org/10.1109/ICIP.2002.1038064 Google Scholar

12. T. Eerolaet al., “Bayesian network model of overall print quality: construction and structural optimisation,” Pattern Recogn. Lett., 32 (11), 1558 –1566 (2011). https://doi.org/http://dx.doi.org/10.1016/j.patrec.2011.04.006 PRLEDG 0167-8655 Google Scholar

13. H. TangN. JoshiA. Kapoor, “Learning a blind measure of perceptual image quality,” in Proc. of the IEEE Computer Vision and Pattern Recognition, 305 –312 (2011). https://doi.org/http://dx.doi.org/10.1109/CVPR.2011.5995446 Google Scholar

14. C. LiA. C. BovikX. Wu, “Blind image quality assessment using a general regression neural network,” IEEE Trans. Neural Networks, 22 (5), 793 –799 (2011). https://doi.org/http://dx.doi.org/10.1109/TNN.2011.2120620 ITNNEP 1045-9227 Google Scholar

15. P. YeD. Doermann, “No-reference image quality assessment using visual codebooks,” IEEE Trans. Image Process., 21 (7), 3129 –3138 (2012). https://doi.org/http://dx.doi.org/10.1109/TIP.2012.2190086 IIPRE4 1057-7149 Google Scholar

16. M. A. SaadA. C. BovikC. Charrier, “A DCT statistics-based blind image quality index,” IEEE Signal Process. Lett., 17 (6), 583 –586 (2010). https://doi.org/http://dx.doi.org/10.1109/LSP.2010.2045550 IESPEJ 1070-9908 Google Scholar

17. A. K. MoorthyA. C. Bovik, “A two-step framework for constructing blind image quality indices,” IEEE Signal Process. Lett., 17 (5), 513 –516 (2010). https://doi.org/http://dx.doi.org/10.1109/LSP.2010.2043888 IESPEJ 1070-9908 Google Scholar

18. Z. Wang, A.C. Bovik, and L. Lu, “Why is image quality assessment so difficult?”, Proc. IEEE Inter. Conference Acoustics, Speech, and Signal Processing(ICASSP-2002),Vol.4, pp. 3313-3316, Orlando, FL, 13-17 May 2002.

19. Video Quality Experts Group (VQEG), <http://www.vqeg.org>.

20. M. Pinson and S. Wolf, “Comparing subjective video quality testing methodologies”, Proceedings of the SPIE, Vol. 5150, pp. 573-582, 2003.

21. Z.M. Parvez Sauzzad, Y. Kawayoke, and Y. Horita, “No reference image quality assessment for JPEG2000 based on spatial features”, Signal Process: Image Communication 23 (2008) pp.257-268.

22. R. Venkatesh Babu, S. Suresh, and Andrew Perkis, “No-reference JPEG-image quality assessment using GAP-RBF”, Signal Processing 87 (2007) pp.1493-1503 23. Z. Wang, H.R. Sheikh, and A.C. Bovik, “No-reference perceptual quality assessment of JPEG compressed images”, Proceedings of the ICIP,02, vol. 1, pp. 477-480, 2002

24. H.R. Sheikh, Z. Wang, L. Cormack and A.C. Bovik, "LIVE Image Quality Assessment Database Release 2", <http://live.ece.utexas.edu/research/quality>.

25. <https://www.sciencedirect.com>

26. <https://github.com>