International Journal of Scientific Engineering and Technology Research

ISSN 2319-8885
Vol.02,Issue.17,
November-2013,
Pages:2038-2048

www.semargroups.org,
www.ijsetr.com

DCT Modern Statics Approach based Blind Image Quality Assessment using A Natural Scene Statistics (NSS) Model

P. SANDHYA DEVA KUMAR[1], K. RAJASHEKAR[2]

[1]PG Scholar, Dept of ECE, University College of Engineering, JNTUK, Kakinada, A.P-INDIA, E-mail: palla.sanjay@yahoo.in.
[2]Asst Prof, Dept of ECE, University College of Engineering, JNTUK, Kakinada, A.P-INDIA,
E-mail: rajasekar487@gmail.com.

Abstract: We develop an efficient general-purpose blind/no-reference image quality assessment (IQA) algorithm using a natural scene statistics (NSS) model of discrete cosine transform (DCT) coefficients. The algorithm is computationally appealing, given the availability of platforms optimized for DCT computation. The approach relies on a simple Bayesian inference model to predict image quality scores given certain extracted features. The features are based on an NSS model of the image DCT coefficients. The estimated parameters of the model are utilized to form features that are indicative of perceptual quality. These features are used in a simple Bayesian inference approach to predict quality scores. The resulting algorithm, which we name BLIINDS-II, requires minimal training and adopts a simple probabilistic model for score prediction. Given the extracted features from a test image, the quality score that maximizes the probability of the empirically determined inference model is chosen as the predicted quality score of that image. When tested on the live IQA database, BLIINDS-II is shown to correlate highly with human judgments of quality, at a level that is competitive with the popular SSIM index.

Keywords: Discrete Cosine Transform (DCT), Generalized Gaussian Density, Natural Scene Statistics, No-Reference Image Quality Assessment.

I. INTRODUCTION

The Ubiquity of transmitted digital visual information in daily and professional life, and the broad range of applications that rely on it, such as personal digital assistants, high-definition televisions, internet video streaming, and video on demand, necessitate the means to evaluate the visual quality of this information. The various stages of the pipeline through which an image passes can introduce distortions to the image, beginning with its capture until its consumption by a viewer. The acquisition, digitization, compression, storage, transmission, and display processes all introduce modifications to the original image. These modifications, also termed distortions or impairments, may or may not be perceptually visible to human viewers. If visible, they exhibit varying levels of annoyance. Quantifying perceptually annoying distortions is an important process for improving the quality of service in applications such as those listed above. Since human raters are generally unavailable or too expensive in these applications, there is a significant need for objective image quality assessment (IQA) algorithms. Only recently did full-reference image quality assessment (FR-IQA) methods reach a satisfactory level of performance, as demonstrated by high correlations with human subjective judgments of visual quality. SSIM [2], MS-SSIM [3], VSNR [4], VIF index [5], and the divisive normalization-based indices in [6] and [7] are examples of successful FR-IQA algorithms. These methods require the availability of a reference signal against which to compare the test signal. In many applications, however, the reference signal is not available to perform a comparison against. This strictly limits the application domain of FR-IQA algorithms and points to the need for reliable blind/NR-IQA algorithms. However, no NR-IQA algorithm has been proven consistently reliable in performance [8]. While some FR-IQA algorithms are reliable enough to be deployed in standards, (e.g., the inclusion of the SSIM index in the H.264/MPEG4 Part 10 AVC reference software [9], [2]), generic NR-IQA algorithms have been regarded as having a long way to go before reaching similar useful levels of performance.

The problem of blindly assessing the visual quality of images, in the absence of a reference, and without assuming a single distortion type, requires dispensing with older ideas of quality such as fidelity, similarity, and metric comparison. Presently, NR-IQA algorithms generally follow one of three trends: 1) distortion-specific approaches. These employ a specific distortion model to drive an objective algorithm to predict a subjective quality score. These algorithms quantify one or more distortions

such as blackness [10], blur [11], [12], or ringing and score the image accordingly; 2) training based approaches: these train a model to predict the image quality score based on a number of features extracted from the image; and 3) natural scene statistics (NSS) approaches: these rely on the hypothesis that images of the natural world (i.e., distortion-free images) occupy a small subspace of the space of all possible images and seek to find a distance between the test image and the subspace of natural images. The first approach is distortion-specific, and hence to some degree, application-specific. It is important to understand, however, that while distortion modeling is important, it does not necessarily embody perceptual relevance (distortion annoyance), since such factors as masking and contrast sensitivity need to be considered. The second approach is only as reliable as the features used to train the learning model. Algorithms following this approach often use a large number of features without perceptually justifying each individual feature. The third approach relies on extensive statistical modeling and reliable generalization of visual content and the perception of it. DIIVINE, which is a recent wavelet-based algorithm, is a combination of the second and the third approaches. It uses a two-stage framework, where the distortion type is predicted first and then, based on this prediction, image quality is estimated. DIIVINE uses a support vector machine (SVM) to classify an image into a distortion class and support vector regression to predict quality scores. A large number of features are used for classification and for quality score prediction (88 features) to achieve high performance against human quality judgments.

As an alternative, we propose a fast single-stage framework that relies on a statistical model of local discrete cosine transform (DCT) coefficients. We derive an algorithm that we dub blind image integrity notator using DCT statistics (BLIINDS-II). The new BLIINDS-II index advances the ideas embodied in an earlier prototype (BLIINDS-I), which uses no statistical modeling and a different set of sample DCT statistics. BLIINDS-I was a reasonably successful experiment to determine whether DCT statistics could be used for blind IQA. BLIINDS-II fully unfolds this possibility and provides an improvement in both performance and in the use of an elegant and general statistical model. It uses a simple Bayesian approach to predict quality scores after a set of features is extracted from an image. For feature extraction, a generalized NSS based model of local DCT coefficients is estimated. The model parameters are used to design features suitable for perceptual image quality score prediction. The statistics of the DCT features vary in a natural and predictable manner as the image quality changes. The NSS features are used by the Bayesian probabilistic inference model to infer visual quality. We show that the method correlates highly with human subjective judgments of quality. We also interpret, analyze, and report how each feature in isolation correlates with human visual perception. The contributions of our approach are as follows.

1. BLIINDS-II inherits the advantages of the NSS approach to IQA. While the goal of IQA research is to produce algorithms that accord with visual perception of quality, one can to some degree avoid modeling poorly understood functions of the human visual system (HVS), by resorting to established models of the natural environment instead. This is motivated by the fact that HVS modeling and NSS modeling can be regarded as dual problems, owing to the widely accepted hypothesis that the HVS has evolutionally adapted to its surrounding visual natural environment.
2. BLIINDS-II is non-distortion-specific, while most NRIQA algorithms quantify a specific type of distortion; the features we use are derived independently of the type of image distortion and are effective across multiple distortion types.
3. We propose a novel model for the statistics of DCT coefficients. Previous work on reduced-reference (RR)-IQA has shown that local image wavelet coefficients are Laplacian-distributed and tend toward Gaussian-distributed when a divisive normalization transform is applied. Our experiments have shown that DCT coefficients have a symmetrical distribution in a manner that can be captured by a generalized Gaussian distribution (GGD) model.
4. Since the framework operates entirely in the DCT domain, one can exploit the availability of platforms devised for the fast computation of DCT transforms. Many image and video compression algorithms are based on block-based DCT transforms (JPEG, MPEG2, H263, and H264 which uses a variation of the DCT).
5. Minimal training is required under the simple Bayesian model.
6. Finally, the method correlates highly with human visual perception of quality and yields highly competitive performance. We provide a MATLAB implementation of BLIINDS-II, which can be downloaded from the Laboratory of Image and Video Engineering (LIVE) website at http://live.ece.utexas .edu/.

II. BLIND IMAGE QUALITY ASSESSMENT

Blind image quality assessment is useful to many image-related applications. In image interpolation, the original high-resolution image is often not available as the ground truth; therefore blind assessment of the quality of the interpolated image becomes necessary. The other example is intelligent memory management in digital cameras. We will refer to undistorted images captured by imaging devices that sense radiation from the visible spectrum as natural scenes, and statistical models built for undistorted natural scenes as NSS models. Deviations from NSS models, caused by the introduction of distortions to images, can be used to predict the perceptual quality of the image. The model-based NSS-IQA approach developed here is a process of feature extraction from the image, followed by statistical modeling of the extracted features. Purely NSS-based IQA approaches require the development of a distance measure between a given

distorted test image and the NSS model. This leads to the question of what constitutes appropriate and perceptually meaningful distance measures between distorted image features and NSS models.

An image entering the IQA "pipeline" is first subjected to local 2-D DCT coefficient computation. This stage of the pipeline consists of partitioning the image into equally sized nxn blocks, henceforth referred to as local image patches, then computing a local 2-D DCT on each of the blocks. The coefficient extraction is performed locally in the spatial domain in accordance with the HVS's property of local spatial visual processing (i.e., in accordance with the fact that the HVS processes the visual space locally). This DCT decomposition is accomplished across spatial scales. The second stage of the pipeline applies a generalized Gaussian density model to each block of DCT coefficients, as well as for specific partitions within each DCT block. We next briefly describe the DCT block partitions that are used. In order to capture directional information from the local image patches, the DCT block is partitioned directionally as shown in Fig. 8 into three oriented sub regions. A generalized Gaussian fit is obtained for each of the oriented DCT coefficient sub regions. Another configuration for the DCT block partition is shown in Fig. 6. The partition reflects three radial frequency sub bands in the DCT block. The upper, middle, and lower partitions correspond to the low-frequency, mid-frequency, and high-frequency DCT sub bands, respectively. A generalized Gaussian fit is obtained for each of the radial DCT coefficient sub regions as well. The third step of the pipeline computes functions of the derived generalized Gaussian model parameters. These are the. Generalized Gaussian density for varying levels of the shape parameter γ. Features used to predict image quality scores. In the following sections, we define and analyze each model-based feature, demonstrate how it changes with visual quality, and examine how well it correlates with human subjective judgments of quality. The fourth and final stage of the pipeline is a simple Bayesian model that predicts a quality score for the image. The Bayesian approach maximizes the probability that the image has a certain quality score given the model-based features extracted from the image. The posterior probability that the image has a certain quality score given the extracted features is modeled as a multidimensional GGD.

III. IMAGE REPRESENTATION: FEATURE SELECTION

A model is only as effective as the features it relies on to represent the data being modeled. In other words, its performance is a function of the representativeness of the features selected to represent the visual quality of the image being assessed. Consequently, the first issue we need to address is what type of features to extract from the image so as to capture as much of the visual quality of the image as possible. Proceeding towards this goal, we first note that humans perform the task of blind IQA quite well, without the need for a reference image to do so. This leads us to believe that the human visual system (HVS) is sensitive to a number of features that distinguish high visual quality images from distorted ones. Further it is believed that the HVS has evolved to adapt to the statistics of the projected natural world, and therefore embodies mechanisms that process images in accordance with these statistics.

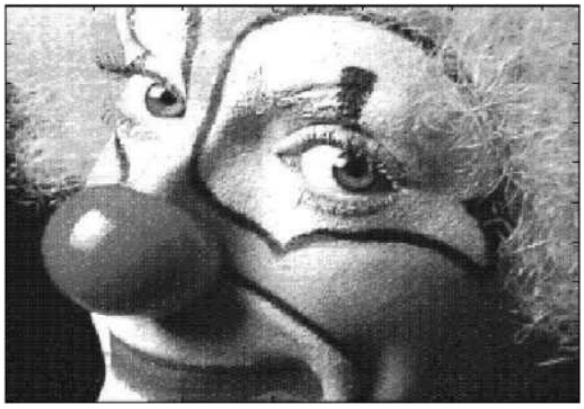

Fig.1. Low contrast image

Fig.2. High contrast image

Fig.3. Sharp object on blurred background.

Our feature selection process relies on the basic fundamental fact that natural images are highly structured (as was hypothesized), in the sense that their pixels exhibit strong dependencies, and these dependencies carry important information about the visual scene. Since it has

been noted that the visual system is highly adapted to its natural environment, and since natural images are highly structured, the HVS is thus assumed to be adapted to extracting structural information from the visual field. Towards this end, we choose to extract image features that represent image structure. In addition to structure, the HVS is highly sensitive to contrast. In general, high contrast in an image is a desirable property, accentuating image structure and making the image more visually appealing. For instance, a majority of viewers might prefer Fig. 1 over Fig. 2 due to the higher contrast in Fig. 1 relative to Fig. 2. Consequently, we choose to extract a feature representative of the contrast of the images, the quality of which is to be assessed. Additionally, the features we extract are also expected to account for image sharpness (without explicitly measuring blur distortion or any other specific distortion) and orientation anisotropies (properties of images that the HVS is also highly sensitive). We note however, that image sharpness for instance, is highly content dependent. For example, the background is blurred in Fig. 3, yet it is a desirable property of this specific image. This is why we do not seek to quantify sharpness or blur as is, but rather explore how the statistics of spatial frequency domain characteristics vary in natural and in distorted images. To do so, we employ the discrete cosine transform (DCT) to extract a number of features and model their statistics.

A. DCT-Based Contrast

Contrast is a basic perceptual attribute of an image. One may distinguish between global contrast measures and ones that are computed locally (and possibly pooled into one measure post local extraction). Several measures of contrast exist in the literature such as the simple global Michelson contrast and the more elaborate Peli's contrast. The Michelson contrast statistics did not correlate with human visual perception of quality in our experiments, while Peli's contrast is computationally too intensive. Instead we resort to computing contrast based on local DCT patches, and show that the statistics of these correlate with human visual perception of distortion. This also conforms to our DCT-only framework. In the simplest, single-scale implementation, the 2-D DCT is applied to 17×17 image patches centered at every pixel in the image. The local DCT contrast is then defined as the average of the ratio of the non-DC DCT coefficient magnitudes in the local patch normalized by the DC coefficient of that patch. The local contrast scores from all patches of the image are then pooled together by averaging the computed values to obtain a global image contrast value.

B. DCT-Based Structure Features

Structure features are derived locally from the local DCT frequency coefficients. We ignore the DC coefficient whose magnitude is usually much larger than the higher frequency DCT coefficients in the image patch. Ignoring the DC coefficient does not alter the local structural content of the image. We illustrate how the statistics of the higher frequency DCT coefficients change as an image becomes distorted in, which show the DCT coefficient histograms of a distortion free and a Gaussian blur distorted image, respectively. Similar trends in the histogram statistics are observed throughout the LIVE database of images, on which we perform our study. Among the observed differences in the histograms is the peakedness at zero, (the distorted images are observed to have a higher histogram peak at zero), and the variance along the support of the histogram. We seek to make use of statistical differences, such as the ones demonstrated above, to develop a NR-IQA index. To capture the statistical traits of the DCT histograms we compute its kurtosis, which quantifies the degree of its peakedness and tail weight, and is given by:

$$\kappa(x) = \frac{E(x-\mu)^4}{\sigma^4} \tag{1}$$

where μ is the mean of x, and σ is its standard deviation. The kurtosis of each DCT image patch, (the same 17×17 image patches from which we computed local DCT contrast) is computed, and the resulting values pooled together by averaging the lowest tenth percentile of the obtained values to obtain a global image kurtosis value.

IV. IMPLEMENTATION OF THE METHOD

Un-distorted images captured by imaging devices that sense radiation from the visible spectrum as Natural scenes, and statistical models built for undistorted natural scenes as NSS models. The model-based NSS-IQA approach developed here is a process of feature extraction from the image, followed by statistical modeling of the extracted features. Our approach relies on the IQA algorithm learning, how the NSS model parameters vary across different trained using features derived directly from a generalized

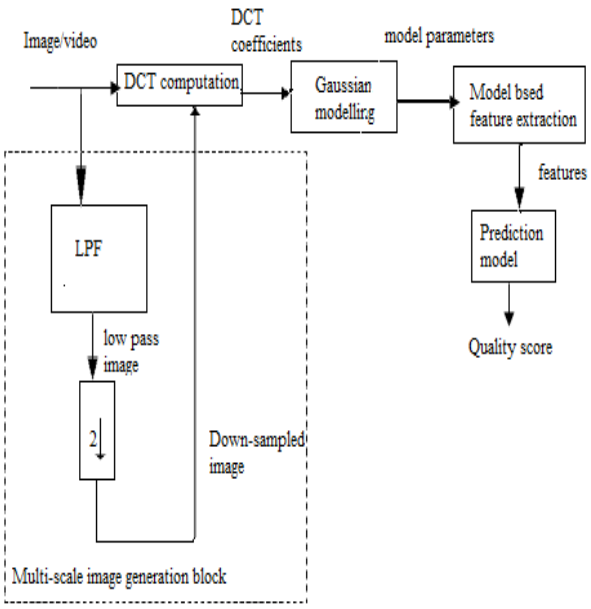

Fig.4 .Framework of BLIINDS-II Algorithm

parametric statistical model of natural image DCT coefficients against various perceptual levels of image distortion. The learning model is then used to predict perceptual image quality scores. NSS models reflect statistical properties of the world that drive perceptual functions of the HVS. Characteristics that can be incorporated into NSS models include: 1) visual sensitivity to structural information: 2) perceptual masking: 3) visual sensitivity to directional information.

A. DCT Feature Domain and Computation

The first stage of the pipeline deals about feature extraction in the DCT domain is the observation that the statistics of DCT coefficients change with the degree and type of image distortion. Another advantage is computational convenience: optimized DCT-specific platforms, and fast algorithms for DCT computation, can ease computation. Many image and video compression algorithms are based on block-based DCT transforms (JPEG, MPEG2, H263, and H264 that relies on a variation of the DCT). An image entering the IQA "pipeline" is first subjected to local 2-D DCT coefficient computation. This stage of the pipeline consists of partitioning the image into equally sized n x n blocks, henceforth referred to as local image patches, then computing a local 2-D DCT on each of the blocks. The coefficient extraction is performed locally in the spatial domain in accordance with the HVS's property of local spatial visual processing (i.e., in accordance with the fact that the HVS processes the visual space locally). This DCT decomposition is accomplished across spatial scales.

TABLE I
MATRIX OF DCT COEFFICIENTS

DC	C_{12}	C_{13}	C_{14}	C_{15}
C_{21}	C_{22}	C_{23}	C_{24}	C_{25}
C_{31}	C_{32}	C_{33}	C_{34}	C_{35}
C_{41}	C_{42}	C_{43}	C_{44}	C_{45}
C_{51}	C_{52}	C_{53}	C_{54}	C_{55}

B. Gaussian Modeling

The second stage of the pipeline applies a generalized Gaussian density model to each block of DCT coefficients, as well as for specific partitions within each DCT block. In order to capture directional information from the local image patches, the DCT block is partitioned directionally as into three oriented sub-regions. The partition reflects three radial frequency sub-bands in the DCT block. The upper, middle, and lower partitions correspond to the low-frequency, mid-frequency, and high-frequency DCT sub-bands, respectively. A generalized Gaussian fit is obtained for each of the radial DCT coefficient sub-regions as well. The introduction of distortions to the images changes the distribution of the coefficients.

TABLE II
ORIENTATION-I OF DCT COEFFICIENTS

DC	C_{12}	C_{13}	C_{14}	C_{15}
C_{21}	C_{22}	C_{23}	C_{24}	C_{25}
C_{31}	C_{32}	C_{33}	C_{34}	C_{35}
C_{41}	C_{42}	C_{43}	C_{44}	C_{45}
C_{51}	C_{52}	C_{53}	C_{54}	C_{55}

TABLE III: ORIENTATION-II OF DCT COEFFICIENTS

DC	C_{12}	C_{13}	C_{14}	C_{15}
C_{21}	C_{22}	C_{23}	C_{24}	C_{25}
C_{31}	C_{32}	C_{33}	C_{34}	C_{35}
C_{41}	C_{42}	C_{43}	C_{44}	C_{45}
C_{51}	C_{52}	C_{53}	C_{54}	C_{55}

Gaussian family of distributions encompasses a wide range of observed behavior of distorted DCT coefficients. The univariate generalized Gaussian density is given by,

$$f(x \mid \alpha, \beta, \gamma) = \alpha e^{-(\beta|X-\mu|)^{\gamma}} \tag{2}$$

where, μ is Mean, α is Normalizing parameter, β is Scale parameter, γ is Shape parameter and are given by,

$$\alpha = \frac{\beta\gamma}{2\Gamma(1/\gamma)} \tag{3}$$

$$\beta = \frac{1}{\sigma}\sqrt{\frac{\Gamma(3/\gamma)}{\Gamma(1/\gamma)}} \tag{4}$$

where, σ is the standard deviation, and Γ denotes the gamma function given by,

$$\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t}\, dt \tag{5}$$

C. Feature extraction

The third step of the pipeline computes functions of the derived generalized Gaussian model parameters. These are the features used to predict image quality scores. In the following sections, we define and analyze each model-based feature, demonstrate how it changes with visual quality, and examine how well it correlates with human subjective judgments of quality.

D. Prediction model

The fourth and final stage of the pipeline is a simple Bayesian model that predicts a quality score for the image.

The Bayesian approach maximizes the probability that the image has a certain quality score given the model-based features extracted from the image.

V. EXPERIMENTS AND RESULTS

BLIINDS-II was rigorously tested on the LIVE IQA database which contains 29 reference images, each impaired by many levels of five distortion types: JPEG2000, JPEG, white noise, Gaussian blur, and fast-fading channel distortions (simulated by JPEG2000 compression followed by channel bit errors.). The total number of distorted images (excluding the 29 reference images) is 779. The DCT computation was applied to 5×5 blocks with a 2-pixel overlap between the blocks. Multiple train–test sequences were run. In each, the image database was subdivided into distinct training and test sets (completely content separate). In each train–test sequence, 80% of the LIVE IQA database content was chosen for training, and the remaining 20% for testing. Specifically, each training set contained

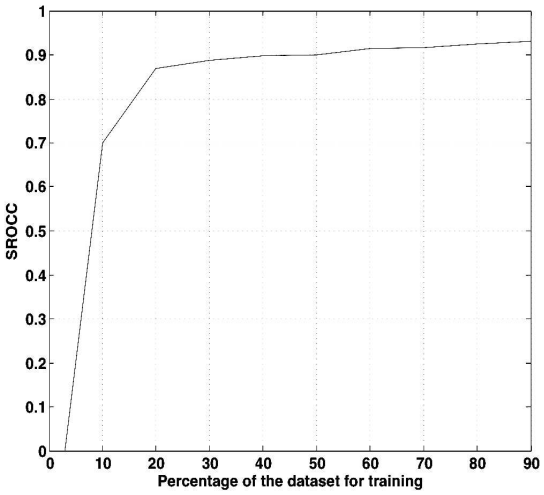

Fig.5. Plot of median SROCC between predicted and subjective DMOS scores (on all distortions) as a function of the percentage of the content used for training.

images derived from 23 reference images, while each test set contained the images derived from the remaining 6 reference images. One thousand randomly chosen training and test sets were obtained, and the prediction of the quality scores was run over the 1000 iterations. The model based-features were extracted over three scales. The total number of features per scale is 8 (4 features, 2 pooling methods/feature). These eight pooled features are: 1) the lowest 10th percentile of the shape parameter γ; 2) the mean of the shape parameter γ; 3) the highest 10th percentile of the coefficient of frequency variation ζ ; 4) the mean (100^{th} percentile) of the coefficient of frequency variation ζ; 5) the highest 10th percentile of the energy sub band ratio measure R_n; 6) the mean of the energy sub band ratio measure; 7) the highest 10th percentile of the orientation feature (which is the variance of ζ across the three orientations); and 8) the mean of the orientation feature.

TABLE IV
MEDIAN SROCC CORRELATIONS FOR 1000 ITERATIONS OF RANDOMLY CHOSEN TRAIN AND TEST SETS (SUBJECTIVE DMOS VERSUS PREDICTED DMOS) ON THE LIVE IQA DATABASE

LIVE subset	BIQI	DIIVINE	BLIINDS-II (SVM)	BLIINDS-II (Prob.)
JPEG2000	0.8557	0.9319	0.9285	0.9506
JPEG	0.7858	0.9483	0.9422	0.9419
White noise	0.9715	0.9821	0.9691	0.9783
GBlur	0.9107	0.9210	0.9231	0.9435
Fast fading	0.7625	0.8714	0.8893	0.8622
ALL	0.8190	0.9116	0.9306	0.9202

We report quality score prediction results for features extracted at one scale only (8 features), over two scales (16 features, 8 features per scale), and over three scales (24 features, 8 per scale). Linear correlation coefficient (LCC) scores (on a logistic fitted function of the predicted DMOS using BLIINDS-II and subjective DMOS scores) as well as SROCC scores between the predicted DMOS scores and the subjective DMOS scores of the LIVE IQA database are computed for each of the 1000 iterations. The comparison of prediction results for 1 scale, 2 scale, and 3 scale feature extraction. We found that no significant gain

TABLE V
MEDIAN LCC CORRELATIONS FOR 1000 ITERATIONS OF TRAIN AND TEST SETS (SUBJECTIVE DMOS VERSUS PREDICTED DMOS) ON THE LIVE IQA DATABASE

LIVE subset	BIQI	DIIVINE	BLIINDS-II (SVM)	BLIINDS-II (Prob.)
JPEG2000	0.8086	0.9220	0.9348	0.9630
JPEG	0.9011	0.9210	0.9676	0.9793
White noise	0.9538	0.9880	0.9799	0.9854
GBlur	0.8293	0.9230	0.9381	0.9481
Fast fading	0.7328	0.8680	0.8955	0.8636
ALL	0.8205	0.9170	0.9302	0.9232

TABLE VI
MEDIAN SROCC AND LCC CORRELATIONS FOR 1000 ITERATIONS OF RANDOMLY CHOSEN TRAIN AND TEST SETS (SUBJECTIVE DMOS VERSUS PREDICTED DMOS) ON THE LIVE IQA DATABASE

	SROCC		LCC	
LIVE subset	SSIM	PSNR	SSIM	PSNR
JPEG2000	0.9496	0.8658	0.9401	0.8640
JPEG	0.9664	0.8889	0.9416	0.8860
White noise	0.9644	0.9791	0.9791	0.9788
GBlur	0.9315	0.7887	0.8910	0.7823
Fast fading	0.9415	0.8986	0.9428	0.8876
ALL	0.9180	0.8669	0.9003	0.8630

in performance was obtained beyond the third scale of feature extraction. To show that the approach is not heavily dependent on the training set, we performed the following analysis. We varied the percentage of the train/test splits from 90% of the content used for training and (the remaining) 10% used for testing, to only 10% of the content used for training and (the remaining) 90% for testing. The remainder of this section: 1) compares BLIINDS-II with other full-reference and no-reference approaches; 2) studies its robustness against various distortion types; 3) addresses database independence; and 4) analyzes its computational complexity.

A. Statistical Comparison with Full-Reference and No-Reference Approaches

For comparison purposes, we also trained a radial basis function-kernel regression SVM, based on the implementation in, and performed quality prediction utilizing this more complex model as well. We also compared BLIINDS-II to the recent SVM-based NR-IQA algorithms BIQI and DIIVINE, the full-reference PSNR, and the state-of-the art FR-IQA SSIM index. The SROCC and LCC results2 are shown in Tables IV–VI. Tables IV and V compare the SROCC and the LCC results between the four NR-IQA methods (recent SVM-based NR-IQA algorithms BIQI and DIIVINE, BLIINDS-II with the SVM prediction model, and BLIINDS-II with the probabilistic prediction model), respectively. Table VI

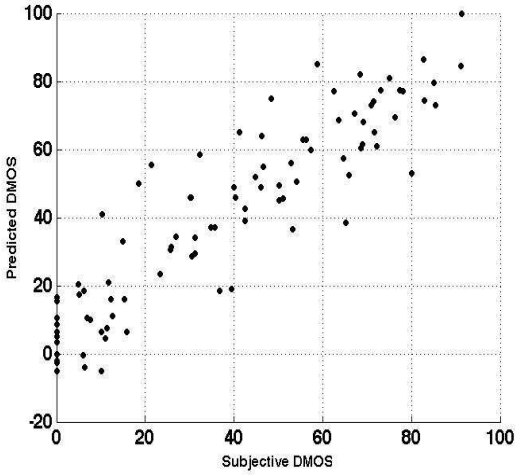

Fig.6. Predicted versus subjective DMOS on the JPEG2000 database subset.

reports the SROCC and LCC results of PSNR and SSIM (the implementation in 3), both of which are full-reference algorithms that require the presence of a reference image to perform quality score prediction on a test image. The two prediction models (probabilistic and SVM) used in BLIINDS-II perform very similarly, with slightly higher correlation for the probabilistic prediction model on the individual distortion subsets (JPEG2000, JPEG, white noise, Gaussian blur, and fast-fading channel distortions) than on the entire dataset. The SVM prediction model only slightly outperforms the simple probabilistic prediction on the entire LIVE IQA database. With either prediction model, BLIINDSII outperforms BIQI and the full-reference PSNR measure. BLIINDS-II also slightly outperforms DIIVINE (on all distortions mixed together) and approaches the performance of the reliable full-reference SSIM index.

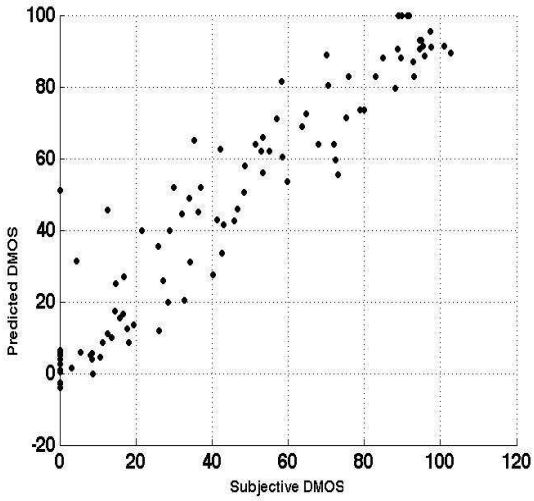

Fig.7. Predicted versus subjective DMOS on the JPEG database subset.

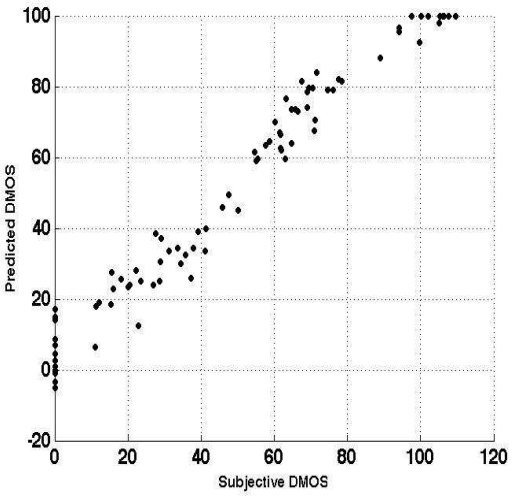

Fig.8. Predicted versus subjective DMOS on the white noise database subset.

Scatter plots (for each of the distortion sets as well as for the entire LIVE IQA Database) of the predicted DMOS using BLIINDS-II versus subjective DMOS on the test sets are shown in Figs.6–11. These exhibit nice properties: a nearly linear relationship against DMOS, tight clustering, and a roughly uniform density along each axis. To visualize the statistical significance of the comparison, we show box plots of the distribution of the SROCC and LCC values for each of the 1000 experimental trials. The plots are shown in Figs. 12 and 13, respectively. We report the standard deviation of the SROCC and LCC results on the 1000 trials for each algorithm in Table VII. Obviously, the lower the standard deviation with a higher median SROCC, the better the performance. The plots show that

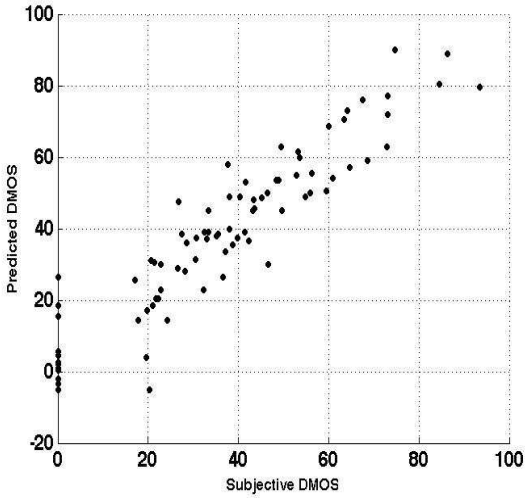

Fig.9. Predicted versus subjective DMOS on the Gaussian blur database subset.

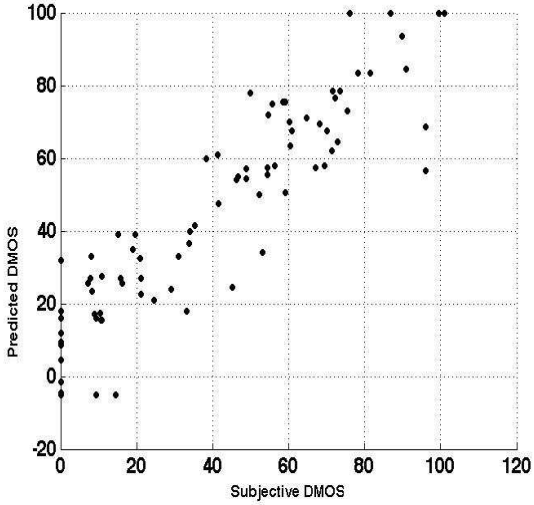

Fig.10. Predicted versus subjective DMOS on the fast-fading channel distortions database subset.

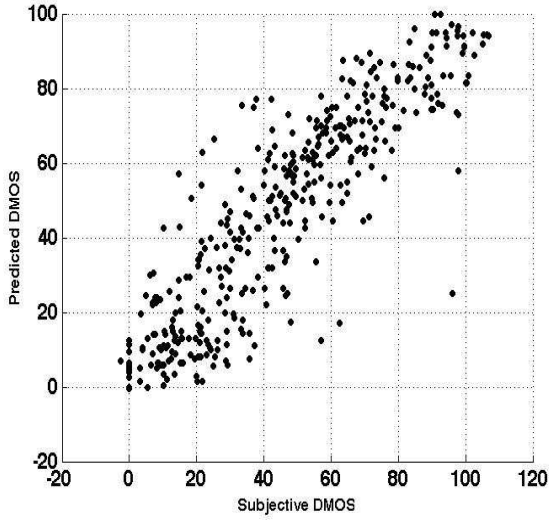

Fig.11. Predicted versus subjective DMOS on the entire LIVE IQA database.

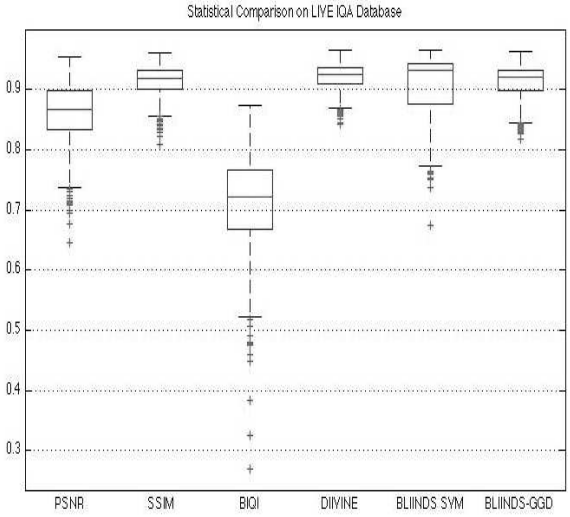

Fig.12. Box plot of SROCC distributions of the algorithms over 1000 trials for algorithm comparison on the LIVE IQA database.

SSIM, DIIVINE, and BLIINDS-II are not statistically significantly different in performance (knowing they are designed for different application domains).

B. Robustness against Distortion Types

A limitation of algorithms that require training is that they are applicable to the set of distortions present in the training phase of the algorithm, i.e., they suffer the limitation of regression techniques. BLIINDS-II was shown to correlate highly with human subjective judgments of quality on images distorted by several common types of distortions available in the LIVE IQA database, namely JPEG, JPEG2000, blur, additive while Gaussian noise, and fast-fading channel distortions. It is however, possible for a trained IQA algorithm to encounter distortions for which it has not been trained. BLIINDS-II can be safely applied to images affected by distortion types that have been included in the training phase of the algorithm (JPEG2000, JPEG, white noise, and blur). Of course, we cannot claim that the algorithm will perform as well on distortions it has not encountered since the algorithm requires training. However, to study how robust the performance of BLIINDSII is when assessing distortions it has not encountered before, we performed the following experiment. We trained the algorithm on all but one distortion. Specifically, we left out the JPEG2000 distorted images from the training set, and mixed the other distortions together in the training set. We then tested on the JPEG2000 subset. We repeated this for each of the other distortion subsets (JPEG, white noise, Gaussian blur, and fast fading channel distortions). To make the problem even more difficult, we split the database according to both content (80% for training, 20% for testing). The resulting SROCCs on each of the distortion types (which were not used for training, but were left out for testing) and which are completely content independent from the training sets are shown in Table VIII. The results are the median SROCC obtained over 1000 iterations of random

train/test splits. Notice that, despite the split in train and test sets, the SROCC correlations obtained were still high on all distortion categories except white noise which is very different from the other distortions.

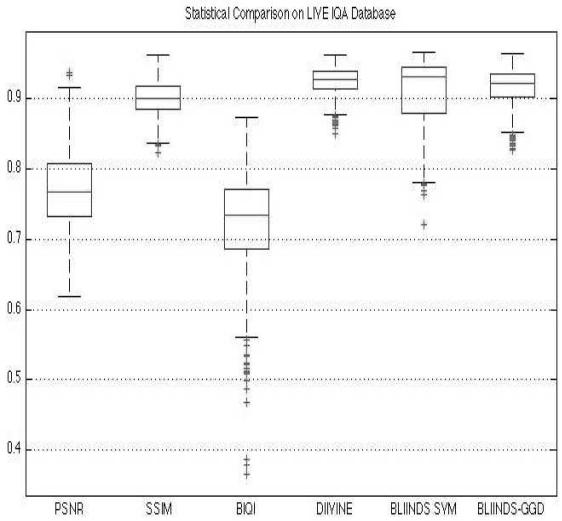

Fig.13. Box plot of LCC distributions of the algorithms over 1000 trials for algorithm comparison on the LIVE IQA database.

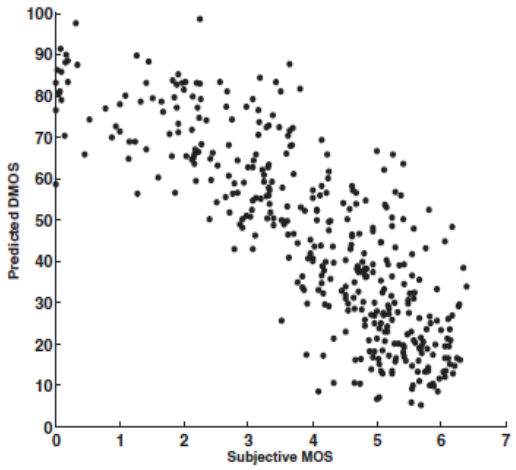

Fig.14. Predicted DMOS versus subjective MOS on the TID2008 database.

C. Database Independence

To study whether the algorithm is database dependent, we tested BLIINDS-II (and the top performing full-reference SSIM index) on a portion of the TID2008 image database. The database contains a large number of distortions, many of which pertain to color distortions (which is not dealt with in this paper). We tested on the mixture of commonly occurring distortions present in the TID2008 database: JPEG2000, JPEG, Gaussian noise, and blur. We trained BLIINDS-II on the LIVE IQA database, and tested it on the same distortions in the TID2008 database. We report SROCC results in Table IX. The SROCC of BLIINDS-II dropped because of the differences in the simulated distortions present in the databases. However, the correlations are still consistently high. A scatter plot of the predicted MOS scores on the TID2008 database as a function of the subjective MOS scores of the database are shown in Fig.14.

TABLE VII
STANDARD DEVIATION OF SROCC AND LCC CORRELATIONS FOR 1000 ITERATIONS OF RANDOMLY CHOSEN TRAIN AND TEST SETS (SUBJECTIVE DMOS VERSUS PREDICTED DMOS) ON THE LIVE IQA DATABASE

SROCC STD						LCC STD					
PSNR	SSIM	BIQI	DIIVINE	BLIINDS-II SVM	BLIINDS-II Prob. model	PSNR	SSIM	BIQI	DIIVINE	BLIINDS-II SVM	BLIINDS-II Prob. model
0.0491	0.0231	0.0745	0.0600	0.0497	0.0279	0.0560	0.1417	0.0676	0.0201	0.0454	0.0279

TABLE VIII
MEDIAN SROCC CORRELATIONS ON EACH OF THE LIVE IQA DATABASE DISTORTION SUBSETS LEFT OUT OF THE TRAINING PHASE AND USED FOR TESTING, AND USING 80%/20% TRAIN/ TEST SPLITS OVER 1000 ITERATIONS. THIS DEMONSTRATES THE ALGORITHM'S ROBUSTNESS RELATIVE TO DISTORTIONS IT HAS NOT BEEN TRAINED ON, AS WELL AS LACK OF ROBUSTNESS IF THE "UNTRAINED" DISTORTION IS VERY DIFFERENT FROM THOSE IT WAS TRAINED ON

JPEG2000	JPEG	White noise	GBlur	Fast fading
0.9034	0.8971	0.1000	0.8514	0.8573

TABLE IX
SROCC RESULTS OBTAINED BY TRAINING ON THE LIVE IQA DATABASE AND TESTING ON TID2008

	PSNR	SSIM	BLIINDS-II (SVM)	BLIINDS-II (Prob.)
JPEG2000	0.8250	0.9603	0.9157	0.9147
JPEG	0.8760	0.9354	0.8901	0.8889
White noise	0.9230	0.8168	0.6600	0.6956
GBlur	0.9342	0.9544	0.8500	0.8572
All	0.8700	0.9016	0.8442	0.8542

D. Algorithm Complexity

Let m × k be the image dimension, and let n × n be the dimension of the blocks from which the model-based features are extracted (in our algorithm n = 5). Then the computational complexity of the algorithm is of the order of $m \times k/n^2 \times n^2 \log n = m \times k \times \log n$. The computational complexity is determined by computation of the DCT transforms and of parameter estimation of the generalized

Gaussian model. Fast algorithms exist for DCT computation. These are of the order O (n^2logn), where n is the dimension of the block (i.e., the block is n × n). Parameter estimation of the generalized Gaussian is of the order of computing moments of the data within each block (O (n^2)), and of numerically estimating the shape parameter γ. From empirical data of natural scenes, it is observed that $0 < \gamma < K$. We set K = 10, since γ was observed to be <<10. The interval [0, K] was partitioned in steps of size ϵ, and the parameter γ was determined by solving an inverse function by numerically sweeping the interval [0, K] in increments of size ϵ. The complexity of such an operation is on the order O (log (1/ϵ)). ϵ was chosen to be 0.001, and hence log (1/ϵ) << min (m, k). The algorithm is also highly parallelizable because one can perform computations on the image blocks in parallel. A further computational advantage can be attained by bypassing DCT computation when DCT coefficients are readily available from an encoder. We envision that the BLIINDS-II approach may also be extendable to scenarios involving DCT-like transforms such as the H.264 integer transforms.

VI. CONCLUSION

We have described a natural scene statistic model-based approach to the no-reference/blind IQA problem. The new NR-IQA model uses a small number of computationally convenient DCT-domain features. The BLIINDS-II algorithm can be easily trained to achieve excellent predictive performance using a simple probabilistic prediction model. The method correlates highly with human visual judgments of quality. BLIINDS-II and the recent no-reference DIIVINE have similar prediction performance results. Both algorithms have limitations. The main limitation of these types of "learning based" algorithms is that they require training to learn the prediction parameters(i.e., they suffer regression limitations). Consequently, if these algorithms are trained on a subset (of all possible) image distortions, then these algorithms are expected to perform well on the distortions they have encountered during training, or on distortions that affect images in a similar manner to the ones encountered during training. We leave the design of effective no-reference methods that are completely non reliant on training as challenging future work. There are significant design differences between DIIVINE and BLIINDS-II. DIIVINE uses a dense complex representation of images in the wavelet domain and extracts a large number of features to train two stages of the algorithm: 1) a nonlinear SVM training for classification and 2) a nonlinear SRV training for regression within each class. Given M assumed distortions, DIIVINE requires M distortion-specific quality assessment engines to be trained and applied. Hence, DIIVINE does not directly accomplish multi-distortion QA. Instead, it computes a probability-weighted linear combination of single-distortion QA predictions. BLIINDS-II adopts a much simpler representation. It uses a lower dimensional feature space and a simpler single-stage (Bayesian prediction based) framework operating in a more sparsely sampled DCT domain. There is only one QA engine in BLIINDS-II that does the multi-distortion QA. Thus, the feature-DMOS relationship is much simpler in BLIINDS-II.

In addition, the DIIVINE index and the BLIINDS index target essentially different application domains. The DIIVINE index, by a two-stage strategy, enables the identification of distortions afflicting the image. This is not only valuable for accomplishing directed quality assessment but also for identifying image distortions to be repaired. This is accomplished at considerable computational expense using a much larger feature set and sophisticated learning mechanisms. By comparison, the BLIINDS index is designed to achieve the speed and performance required by a quality assessment algorithm operating in a high speed video network. It accomplishes this via a simple one-stage QA process using a small number of NSS features that are easily computed from a small (sub sampled) number of fast DCT coefficients, using a very simple probabilistic classifier. In the future, we envision NSS-based QA algorithms that use spatiotemporal features for video-QA and NSS-depth features for stereo-QA. These may efficiently operate in the DCT domain like BLIINDS-II.

VII. REFERENCES

[1] Michele A. Saad, Student Member, IEEE, Alan C. Bovik, Fellow, IEEE, and Christophe Charrier, Member, IEEE, "Blind Image Quality Assessment: A Natural Scene Statistics Approach in the DCT Domain", IEEE Transactions on Image Processing, Vol. 21, No. 8, August 2012.

[2] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," IEEE Trans. Image Process., vol. 13, no. 4, pp. 600–612, Apr. 2004.

[3] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity image quality assessment," in Proc. 37th Asilomar Conf. Signals Syst. Comput., Nov. 2003, pp. 1398–1402.

[4] D. M. Chandler and S. S. Hemami, "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," IEEE Trans. Image Process., vol. 16, no. 9, pp. 2284–2298, Sep. 2007.

[5] H. R. Sheikh, A. C. Bovik, and G. de Veciana, "Image information and visual quality," IEEE Trans. Image Process., vol. 15, no. 2, pp. 430–444, Feb. 2006.

[6] P. C. Teo and D. J. Heeger, "Perceptual image distortions," Proc. SPIE, vol. 2179, pp. 127–141, Feb. 1994.

[7] V. Laparra, J. Munoz-Mari, and J. Malo, "Divisive normalization image quality metric revisited," J. Opt. Soc. Amer., vol. 27, no. 4, pp. 852–864, Apr. 2010.

[8] E. Cohen and Y. Yitzhaky, "No-reference assessment of blur and noise impacts on image quality," Signal Image Video Process. vol. 4, no. 3, pp. 289–302, 2010.

[9] A. M. Tourapis, A. Leontaris, K. Suhring, and G. Sullivan, "H.264/14496-10 AVC reference software manual," in Proc. 31st Meeting Joint Video Team, Jul. 2009, pp. 1–90.

[10] Z. Wang, A. C. Bovik, and B. L. Evans, "Blind measurement of blocking artifacts in images," in Proc. IEEE Int. Conf. Image Process., Sep. 2000, pp. 981–984.

[11] Z. M. P. Sazzad, Y. Kawayoke, and Y. Horita, "No-reference image quality assessment for jpeg2000 based on spatial features," Signal Process. Image Commun. vol. 23, no. 4, pp. 257–268, Apr. 2008.

[12] X. Zhu and P. Milanfar, "A no-reference sharpness metric sensitive to blur and noise," Quality Multim. Exp. Int. Workshop, San Diego, CA, Jul. 2009, pp. 64–69.