Blind Image Quality Assessment: A Natural Scene Statistics Approach in the DCT Domain

Michele A. Saad, *Student Member, IEEE*, Alan C. Bovik, *Fellow, IEEE*, and Christophe Charrier, *Member, IEEE*

***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.**

***Index Terms*— 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 [1], MS-SSIM [2], VSNR [3], VIF index [4], and the divisive normalization-based indices in [5] and [6] 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 [7]. 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 [8], [1]), 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 blockiness [9], blur [10], [11], or ringing [12] 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 [13]–[15]; 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 [16].

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

Manuscript received December 14, 2010; revised April 12, 2011; accepted March 5, 2012. Date of publication March 21, 2012; date of current version July 18, 2012. This work was supported in part by the Grant CSOSG ANR-08-SECU-007-04 and Intel and Cisco, Inc., under the VAWN Program. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Jesus Malo.

M. A. Saad and A. C. Bovik are with the Department of Electrical and Computer Engineering, University of Texas, Austin, TX 78701 USA (e-mail: michele.saad@utexas.edu; bovik@ece.utexas.edu).

C. Charrier is with the Department of Electrical and Computer Engineering, University of Caen, Caen 14000, France (e-mail: christophe.charrier@unicaen.fr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2012.2191563

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 [17]. 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) [18], 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 [19], [20].
2) BLIINDS-II is non-distortion-specific, while most NR-IQA 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 [21], [22]. 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/.

The rest of this paper is organized as follows. In Section II, we describe the DCT-domain NSS features and the motivation behind the choice of the features. In Section III, we show how each feature correlates with subjective *difference-mean-opinion-scores* (DMOS). In Section IV, we describe the generalized probabilistic prediction model. We present the results in Section V, and we conclude in Section VI.

II. Overview of the Method

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. The Kullback–Leibler divergence [21] as well as other distance measures have been used for this purpose, but no perceptual justification has been provided for its use.

Our approach relies on the IQA algorithm learning how the NSS model parameters vary across different perceptual levels of image distortion. The algorithm is trained using features derived directly from a generalized 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.

Unlike much of the prior work on image/video quality assessment (QA) [1], [2], [4], [23], [24], we make little direct use of specific perceptual models such as area V1 cortical

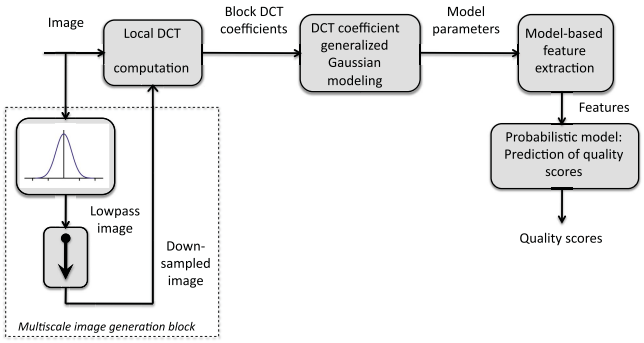

Fig. 1. High-level overview of the BLIINDS-II framework.

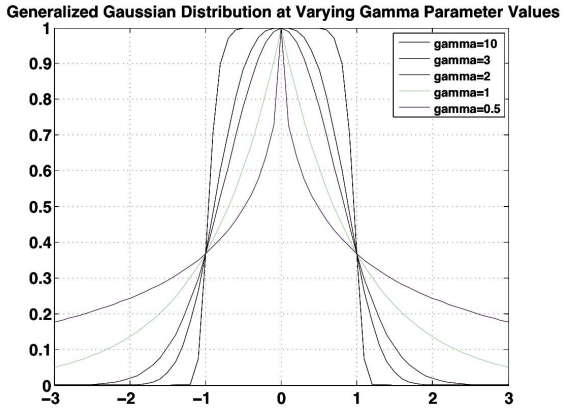

Fig. 2. Generalized Gaussian density for varying levels of the shape parameter γ.

decompositions [24], masking [1], [2], [4], [5], [6], [21], [25], and motion perception [24]. Yet we consider our approach as perceptually relevant since the NSS models reflect statistical properties of the world that drive perceptual functions of the HVS. This is a consequence of the belief that the HVS is adapted to the statistics of its visual natural environment. In other words, models of natural scenes embody characteristics of the HVS, which is hypothesized to be evolutionally adapted to models conforming to natural scenes [19], [20], [26]. HVS characteristics that are intrinsic to, or that can be incorporated into NSS models include: 1) visual sensitivity to structural information [1], [2]; 2) perceptual masking [19], [21], [22], [24]; 3) visual sensitivity to directional information [27], [28]; 4) multiscale spatial visual processing [4], [19], [24]; and 5) intolerance to flagrantly visible visual distortions [29]. In the following sections we explain how one or more of these HVS properties are embedded in the model.

The framework of the proposed approach is summarized in Fig. 1. 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) [19]. 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 subregions. A generalized Gaussian fit is obtained for each of the oriented DCT coefficient subregions. Another configuration for the DCT block partition is shown in Fig. 6. The partition reflects three radial frequency subbands in the DCT block. The upper, middle, and lower partitions correspond to the low-frequency, mid-frequency, and high-frequency DCT subbands, respectively. A generalized Gaussian fit is obtained for each of the radial DCT coefficient subregions as well.

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.

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.

A. Generalized Probabilistic Model

The Laplacian model has often been used to approximate the distribution of DCT image coefficients [30]. This model is characterized by a large concentration of values around zero and heavy tails. However, the introduction of distortions to the images changes the distribution of the coefficients, as shown in [4] and [31]. Such descriptive terms as *heavy tails*, *peakedness at zero*, and *skewness*, which have often been used to describe distributions, are intrinsically heuristic. In our prior work [18], we used such sample statistics (kurtosis, entropy, etc.,), without image modeling, to create a reasonably successful but preliminary blind IQA algorithm. We have refined our approach by modeling image features using a generalized Gaussian family of distributions which encompasses a wide range of observed behavior of distorted DCT coefficients. The generalized Gaussian model has recently been used as a feature in a NSS-based RR-IQA algorithm [21] and in a simple two-stage NR-IQA algorithm in [15].

The univariate generalized Gaussian density is given by

$$f(x|\alpha,\beta,\gamma) = \alpha e^{-(\beta|x-\mu|)^{\gamma}} \tag{1}$$

where μ is the mean, γ is the shape parameter, and α and β are the normalizing and scale parameters given by

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

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

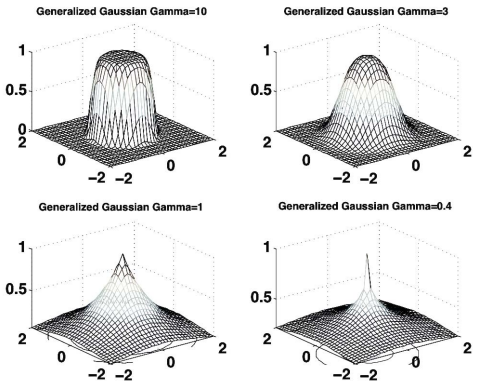

Fig. 3. 2-D generalized Gaussian density plotted for several values of the shape parameter γ.

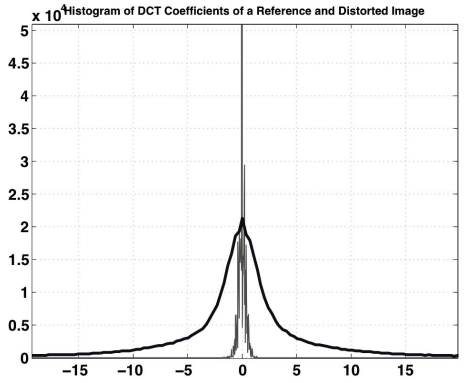

Fig. 4. Histograms of non-DC DCT coefficients. Sharp-peaked histogram: Distorted image histogram. Lower-peaked histogram: reference image histogram.

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

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt. \tag{4}$$

This family of distributions includes the Gaussian distribution ($\beta = 2$) and the Laplacian distribution ($\beta = 1$) [32], [33]. As $\beta \to \infty$, the distribution converges to a uniform distribution. Fig. 2 shows the GGD at varying levels of the shape parameter γ.

A variety of parameter estimation methods have been proposed for this model. We deploy the reliable method given in [34].

The multivariate version of the generalized Gaussian density is given by

$$f(x|\alpha, \beta, \gamma) = \alpha e^{-(\beta(\mathbf{x}-\mu)^{\mathbf{T}}\Sigma^{-\mathbf{1}}(\mathbf{x}-\mu))^{\gamma}} \tag{5}$$

where Σ is the covariance matrix of the multivariate random variable $\mathbf{x}$, and the remaining parameters are as defined in the univariate case. We use (5) to form a probabilistic prediction model in Section IV. Fig. 3 shows the 2-D GGDs for various values of the shape parameter γ.

Parameter estimation is treated similar to the univariate case once the quantity $(\mathbf{x}-\mu)^{\mathbf{T}}\Sigma^{-\mathbf{1}}(\mathbf{x}-\mu)$ is estimated from the sample data.

B. DCT Feature Domain

The performance of any IQA model is a function of the representativeness of the features that are used for quality score prediction. In other words, the prediction is only as good as the choice of features extracted. This motivates us to design features representative of human visual perception of quality. Consequently, given that it is broadly agreed upon that the HVS is adapted to the statistics of images of its natural environment, we design features motivated by natural scene characteristics. It has been shown that natural images exhibit strong spatial structural dependencies [1]. Consequently, we define features representative of image structure, and whose statistics are observed to change with image distortions. The structural information[1] in natural images may loosely be described as smoothness, texture, and edge information composed by local spatial frequencies that constructively and destructively interfere over scales to produce the spatial structure in natural scenes.

Visual images are subjected to local spatial frequency decompositions in the visual cortex [4], [19], [35]. Likewise, in our IQA model, feature extraction is performed in the local frequency (DCT) domain. The main motivation behind 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 [36]–[39], and fast algorithms for DCT computation [40], [41] can ease computation. For instance, DCTs can be computed efficiently by variable-change transforms from computationally efficient fast Fourier transform algorithms [42]. 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). Consequently, the model-based method could be applied to already-computed coefficients, resulting in even greater computational efficiency. Finally, and perhaps most importantly, it is possible to define simple and naturally defined model-based DCT features that capture perceptually relevant image and distortion characteristics in a natural and convenient manner.

We illustrate one instance of how the statistics of DCT coefficients changes as an image becomes distorted in Fig. 4, which shows the DCT coefficient histograms of a distortion-free image and a Gaussian blur distorted image (in Fig. 5), respectively. Fig. 4 shows an example of how the DCT coefficient histograms of distorted and pristine images may differ significantly. The differences in observed DCT coefficient distributions between distorted and nondistorted images (such as the difference observed in Fig. 4) are exploited in the design of features of visual quality score prediction.

Similar trends in the histogram statistics are observed over large databases of distorted images [18], [43], [44]. Among the observed differences in the histograms is the degree of peakedness at zero (blurred images are observed to have a higher histogram peak at zero), and variance (blurred images exhibit reduced variance). We utilize statistical differences

[1]Related to spatial correlation between pixel intensities.

Reference image (DMOS = 0) Blur distorted image (DMOS = 73.45)

Fig. 5. Images corresponding to the histograms of DCT coefficients in Fig. 4.

TABLE I

SROCC CORRELATIONS [SUBJECTIVE DMOS VERSUS DCT γ LOWEST 10th PERCENTILE AND 100th PERCENTILE, ζ HIGHEST 10th PERCENTILE AND 100th PERCENTILE, ENERGY SUBBAND RATIO HIGHEST 10th PERCENTILE AND 100th PERCENTILE, ORIENTED ζ VARIANCE POOLED ACCORDING TO THE HIGHEST 10th PERCENTILE AND 100th PERCENTILE (FEATURE EXTRACTION BASED ON 5×5 DCT BLOCKS)]

	γ		ζ		Subbands feature		Orientation feature	
LIVE Subset	10%	100%	10%	100%	10%	100%	10%	100%
JPEG2000	0.9214	0.7329	0.9334	0.9131	0.9313	0.8745	0.9175	0.9269
JPEG	0.7790	0.7295	0.8070	0.0446	0.9493	0.4601	0.8258	0.7662
WN	0.9580	0.9233	0.9582	0.9360	0.9754	0.9608	0.9524	0.9499
GBLUR	0.9009	0.3298	0.9245	0.8614	0.8850	0.5808	0.9277	0.9228
FASTFADING	0.8266	0.6282	0.8312	0.8410	0.8602	0.7558	0.8639	0.8656

such as these to develop an NR-IQA index. We describe each of the model-based features used and show how each correlates with human judgments of quality in the following.

III. MODEL-BASED DCT DOMAIN NSS FEATURES

We propose a parametric model to model the extracted local DCT coefficients. The parameters of the model are then utilized to extract features for perceptual quality score prediction. We extract a small number of model-based features (only four), as described next. Additionally, toward the end of this section we point out the challenge of blindly predicting visual quality across multiple distortions types, and we explain the importance of multiscale feature extraction.

A. Generalized Gaussian Model Shape Parameter

We deploy a generalized Gaussian model of the non-*DC* DCT coefficients from *nxn* blocks. The *DC* coefficient does not convey structural information about the block, including it neither increases nor decreases performance. The generalized Gaussian density in (1) is parameterized by mean μ, scale parameter β, and shape parameter γ. The shape parameter γ is a model-based feature that is computed over all blocks in the image.

The shape parameter quality feature is pooled in two ways. First, by computing the lowest 10th percentile average of the local block shape scores (γ) across the image. This kind of "percentile pooling" has been observed to result in improved correlations with subjective perception of quality [23], [45]. Percentile pooling is motivated by the observation that the "worst" distortions in an image heavily influence subjective impressions. We choose 10% as a round number to avoid the possibility of training. In addition, we compute

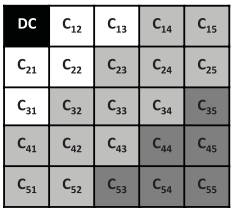

Fig. 6. DCT coefficients, three bands.

the 100th percentile average (ordinary sample mean) of the local γ scores across the image. Using both 10% and 100% pooling helps inform the predictor whether the distortions are uniformly annoying over space or exhibit isolated perceptually severe distortions.

We demonstrate the distortion prediction efficacy of the shape feature γ on a large database of distorted images. The LIVE IQA Database consists of five subset datasets, each of which consists of images distorted by five types of representative realistic distortions [JPEG2000 compression, JPEG compression, white noise, Gaussian blur, and fast-fading channel distortions (simulated by JPEG2000 distortion followed by bit errors)]. In Table I we report Spearman rank order correlation coefficient (SROCC) scores between the LIVE DMOS scores, and 10% and 100% pooled features, respectively. (The DCT blocks from which γ was estimated were chosen to be of dimension 5×5.)

Observe that the correlations are consistently higher when the lowest 10th percentile pooling strategy is adopted. This may be interpreted as further evidence that human sensitivity to image distortions is not a linear function of the distortion. For instance, humans tend to judge poor regions in an image more harshly than good ones, and hence penalize images with even a small number or area of poor regions more heavily

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}

Fig. 7. Matrix of DCT coefficients.

[29], [45]. Unlike perceptual masking, which has a quantitative explanation in terms of perceptual adaptive gain mechanisms [25], this effect is a behavioral one.

B. Coefficient of Frequency Variation

Let X be a random variable representing the histogrammed DCT coefficients. The next feature is the *coefficient of frequency variation* feature

$$\zeta = \frac{\sigma_{|X|}}{\mu_{|X|}} \tag{6}$$

which we shall show is equivalent to

$$\zeta = \sqrt{\frac{\Gamma(1/\gamma)\Gamma(3/\gamma)}{\Gamma^2(2/\gamma)} - 1} \tag{7}$$

where $\sigma_{|X|}$ and $\mu_{|X|}$ are the standard deviation and mean of the DCT coefficient magnitudes $|X|$, respectively.

If X has probability density function (1) and $\mu_X = 0$, then

$$\mu_{|X|} = \int_{-\infty}^{+\infty} |x| \alpha e^{-(\beta|x|)^\gamma} dx = \frac{2\alpha}{\beta^2 \gamma} \Gamma(\frac{2}{\gamma}) \tag{8}$$

where α and β are given by (2) and (3), respectively. Substituting for α and β yields

$$\frac{\Gamma(1/\gamma)\Gamma(3/\gamma)}{\Gamma^2(2/\gamma)} = \frac{\sigma^2}{\mu_{|X|}^2}. \tag{9}$$

Further

$$\sigma_{|X|}^2 = \sigma_X^2 - \mu_{|X|}^2 \tag{10}$$

so that

$$\zeta = \frac{\sigma_{|X|}}{\mu_{|X|}} = \sqrt{\frac{\Gamma(1/\gamma)\Gamma(3/\gamma)}{\Gamma^2(2/\gamma)} - 1} \tag{11}$$

and σ_X is the standard deviation of X.

The feature ζ is computed for all blocks in the image. The feature is pooled by averaging over the highest 10th percentile and over all (100th percentile) of the local block scores across the image. The motivation behind the percentile pooling strategy is similar to that for pooling of the shape parameter feature γ. As shown in Table I, the highest 10th percentile pooling correlates well with human subjectivity. As before, both pooling results (10% and 100%) are supplied to the predictor, since the difference between these is a compact but rich form of information regarding the distribution of severe scores.

In the coefficient of frequency variation ζ, the denominator $\mu_{|X|}$ measures the center of the DCT coefficient magnitude distribution, while $\sigma_{|X|}$ measures the spread or energy of the DCT coefficient magnitudes. The ratio ζ correlates well with visual impressions of quality as shown in Table I. The high correlation between ζ and subjective judgments of perceptual quality is an indication of the monotonicity between ζ and subjective DMOS. Since ζ is the ratio of the variance $\sigma_{|X|}$ to the mean $\mu_{|X|}$, the effect of an increase (or decrease) of $\sigma_{|X|}$ in the numerator is mediated by the decrease (or increase) of $\mu_{|X|}$ in the denominator of ζ. Indeed, two images may have similar perceptual quality even if their respective DCT coefficient magnitude energy ($\sigma_{|X|}$) is very different, depending on where the distribution of the coefficient magnitude energy is centered ($\mu_{|X|}$).

C. Energy Subband Ratio Measure

Image distortions often modify the local spectral signatures of an image in ways that make them dissimilar to the spectral signatures of pristine images. To measure this, we define a local DCT energy–subband ratio measure. Consider the 5×5 matrix shown in Fig. 7. Moving from the top-left corner of the matrix toward the bottom-right corner, the DCT coefficients represent increasingly higher radial spatial frequencies. Consequently, we define three frequency bands depicted by different levels of shading in Fig. 6. Let Ω_n denote the set of coefficients belonging to band n, where $n = 1, 2, 3$, (lower, middle, higher). Then define the average energy in frequency band n to be the model variance σ_n^2 corresponding to band n

$$E_n = \sigma_n^2. \tag{12}$$

This is found by fitting the DCT data histogram in each of the three spatial frequency bands to the generalized Gaussian model (1), and then using the σ_n^2 value from the fit. The ratio of the difference between the average energy in frequency band n and the average energy up to frequency band n, as well as the sum of these two quantities is then computed

$$R_n = \frac{\left| E_n - \frac{1}{n-1} \sum_{j<n} E_j \right|}{E_n + \frac{1}{n-1} \sum_{j<n} E_j}. \tag{13}$$

R_n is defined for $n = 2, 3$. A large ratio corresponds to a large disparity in the frequency energy between a frequency band and the average energy in bands of lower frequencies. This feature measures the relative distribution of energies in lower and higher bands, which can be affected by distortions. In the spatial domain, a large ratio roughly corresponds to uniform frequency (textural) content in the image patch. A low ratio, on the other hand, corresponds to a small frequency disparity between the feature band and the average energy in the lower bands. The mean of R_2 and R_3 is computed. This feature is computed for all blocks in the image. As before, the feature is pooled by computing the highest 10th percentile average and the 100th percentile average (ordinary mean) of the local block scores across all the image.

Table I we reports SROCC scores between the LIVE IQA Database DMOS scores and the pooled highest 10% and 100% averaged feature values, respectively, using 5×5 blocks. Observe that the correlation is consistently higher when the 10th percentile pooling strategy is adopted.

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}

Fig. 8. DCT coefficients collected along three orientations.

TABLE II
SROCC AND LCC CORRELATIONS OF EACH FEATURE AGAINST SUBJECTIVE DMOS ON THE ENTIRE LIVE IQA DATABASE

	SROCC		LCC	
Feature	10%	100%	10%	100%
Shape parameter γ	0.1167	0.1896	0.3830	0.4065
Coefficient of variation ζ	0.4173	0.1548	0.7285	0.6109
Energy subband ratio measure	0.3713	0.5495	0.3786	0.0629
Orientation feature	0.0236	0.012	0.1896	0.4065

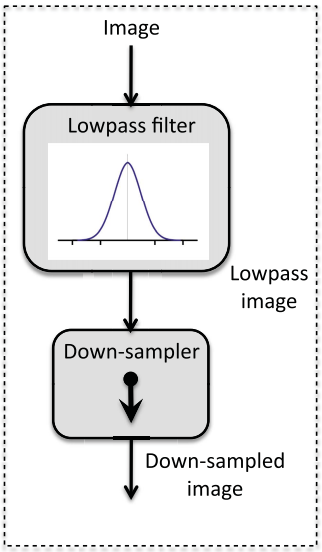

Fig. 9. Multiscale image generation.

D. Orientation Model-Based Feature

Image distortions often modify local orientation energy in an unnatural manner. The HVS, which is highly sensitive to local orientation energy [19] is likely to respond to these changes. To capture directional information in the image that may correlate with changes in human subjective impressions of quality, we model the block DCT coefficients along three orientations. We demonstrate how oriented DCT coefficients are captured in Fig. 8 below. The three differently shaded areas represent the DCT coefficients along three orientation bands. A generalized Gaussian model is fitted to the coefficients within each shaded region in the block, and ζ is obtained from the model histogram fits for each orientation. The variance of ζ is computed along each of the three orientations. The variance of ζ across the three orientations from all the blocks in the image is then pooled (highest 10th percentile and 100th percentile averages) to obtain two numbers per image. We report how this pooled feature correlates with subjective DMOS in Table I, again using 5×5 blocks.

Table I shows the SROCC for each feature pooled in two different ways against DMOS. Notice that the improvement in the percentile pooling (10% versus 100%) in JPEG and white noise is slighter than for the other distortion types. This can be attributed to the way in which distortions manifest in each of these subsets (JPEG2000, JPEG, white noise, Gaussian blur, and fast fading channel distortions). Percentile pooling is particularly helpful when the errors are localized in the image (i.e., occur at specific locations in the image), as opposed to occurring uniformly over the image. In JPEG, for instance, blocking can manifest over most of an image, whereas other distortions produce spatially sparser effects such as JPEG2000, which produces ringing near edges due to wavelet-based compression. Some distortions, such as the packet loss in the "fast fading" category of the LIVE IQA database, produce visible artifacts only at isolated spatial locations in the image.

We performed a leave-one-out cross-validation analysis of each of the four features using the prediction model described in Section IV. Our results showed that the prediction accuracy drops if any one of the four features is left out. This analysis was done to ensure that the set of features is nonredundant.

E. Blind Quality Assessment Across Distortion Types

Table I shows that each of the features correlates highly with human visual perception of quality when applied to some, but not all, individual distortion types (JPEG2000, JPEG, white noise, Gaussian blur, fast fading channel distortions). A major challenge arises when one assumes no knowledge of the type of distortion affecting an image. It then becomes necessary to combine complementary features that collectively perform well at predicting image quality blindly, over a wide range of distortion types. In Table II, we demonstrate the complementarity of the features in terms of correlation with DMOS on the entire LIVE IQA database of images (with all distortion types mixed together). The low correlations between each individual feature and subjective DMOS across all distortion types points up the need to combine the complementary features in a manner that reliably enables distortion-agnostic quality score prediction. The manner in which we combine the features to predict blind image quality scores is discussed in Section IV.

F. Multiscale Feature Extraction

It is well understood that images are naturally multiscale [4], [20], and that the early visual system involves decompositions over scales [19]. Successful FR-IQA algorithms have exploited this fact to create natural multiscale measurements of image quality [2], [4]. Toward this end, we implement the BLIINDS-II concept over multiple scales in a simple way. Specifically, the NSS-based DCT features are extracted from 5×5, overlapping blocks in the image. The feature extraction is repeated after low-pass filtering the image and subsampling it by a factor of 2 as shown in Fig. 9. Prior to downsampling, the image is filtered by a rotationally symmetric discrete 3×3 Gaussian filter kernel depicted in Fig. 10. At each scale, the

0.0113	0.0838	0.0113
0.0838	0.6193	0.0838
0.0113	0.0838	0.0113

Fig. 10. Gaussian kernel used prior to downsampling.

overlap between neighboring blocks is two pixels. This defines a multiscale feature extraction approach. Multiscale feature extraction and processing generally improves performance when dealing with changes in the image resolution, distance from the image display to the observer, or variations in the acuity of the observer's visual system. In BLIINDS-II, feature extraction over multiple scales makes it possible to capture variations in the degree of distortion over scales.

IV. Prediction Model

We have found that a simple probabilistic predictive model is adequate for training the features used in BLIINDS-II. The prediction model is the only element of BLIINDS-II that carries over from BLIINDS-I. The efficacy of this simple predictor demonstrates the effectiveness of the NSS-based features used by BLIINDS-II to predict image quality. Let $X_i = [x_1, x_2, ..., x_m]$ be the vector of features extracted from the image, where i is the index of the image being assessed, and m be the number of pooled features that are extracted. Additionally, let DMOS_i be the subjective DMOS associated with the image i. We model the distribution of the pair (X_i, DMOS_i).

The probabilistic model is trained on a subset of the LIVE IQA database, which includes DMOS scores, to determine the parameters of the probabilistic model by distribution fitting. The multivariate GGD model in (5) is used to model the data. Parameter estimation only requires the mean and covariance of the empirical data from the test set. The probabilistic model $P(X, \text{DMOS})$ is applied by fitting (5) to the empirical data of the training set. Specifically, once the quantity $(x-\mu)^T \Sigma^{-1} (x-\mu)$ is estimated from the sample data, parameter estimation of the GGD model in (5) is performed using the fast method in [34]. A full analysis of the method is found in [34]. The distribution fitting ($P(X, \text{DMOS})$) on the training data is only a fast intermediate step toward DMOS prediction. The end goal is not to fit the sample data of the training set as accurately as possible to the prediction model. Instead, the aim is to achieve high correlations between predicted and subjective DMOS using this prediction model. We show in Section V that a large number of images are not needed to train the model in order to predict DMOS accurately. The training and test sets are completely content-independent, in the sense that no two images of the same scene are present in both sets. The probabilistic model is then used to perform prediction by maximizing the quantity $P(\text{DMOS}_i | X_i)$. This is equivalent to maximizing the joint distribution $P(X, \text{DMOS})$ of X and DMOS since $P(X, \text{DMOS}) = P(\text{DMOS}|X)\,p(X)$.

TABLE III
Median SROCC Correlations for 1000 Iterations of Randomly Chosen Train and Test Sets (Subjective DMOS Versus Predicted DMOS). Comparison for Multiple Scales of Feature Extraction

LIVE subset	BLIINDS-II One scale	BLIINDS-II Two scales	BLIINDS-II Three scales
JPEG2000	0.9313	0.9533	0.9506
JPEG	0.9294	0.9403	0.9419
White noise	0.9753	0.9772	0.9783
GBlur	0.9417	0.9509	0.9435
Fast fading	0.88555	0.8657	0.8622
ALL	0.8973	0.8980	0.9202

TABLE IV
Median LCC Correlations for 1000 Iterations of Randomly Chosen Train and Test Sets (Subjective DMOS Versus Predicted DMOS). Comparison for Multiple Scales of Feature Extraction

LIVE subset	BLIINDS-II One scale	BLIINDS-II Two scales	BLIINDS-II Three scales
JPEG2000	0.9550	0.9571	0.9630
JPEG	0.9664	0.9781	0.9793
White noise	0.9804	0.9833	0.9854
GBlur	0.9300	0.9450	0.9481
Fast fading	0.8500	0.8701	0.8636
ALL	0.8919	0.9091	0.9232

Since a user might be interested in applying BLIINDS-II to images suffering distortions other than those used here, the question arises regarding how large the training set would need to be to produce accurate DMOS predictions on the new distortions. Generally, this question is difficult or impossible to answer since it would likely rely on the type and number of the distortions as well as their ranges of perceptual severity or visibility. From a purely "surface fitting" perspective, in order to accurately fit the GGD model to sample data one would need the number of points in the training set to be on the order of the number of unknown parameters in the probabilistic model [34]. In our case, this constraint is easily satisfied.

V. Experiments and Results

BLIINDS-II was rigorously tested on the LIVE IQA database [43] 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

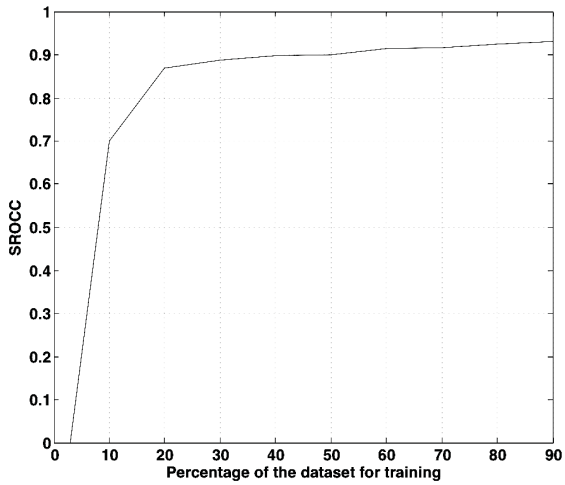

Fig. 11. Plot of median SROCC between predicted and subjective DMOS scores (on all distortions) as a function of the percentage of the content used for training.

TABLE V
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

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 (100th percentile) of the coefficient of frequency variation ζ; 5) the highest 10th percentile of the energy subband ratio measure R_n; 6) the mean of the energy subband 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.

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

TABLE VI
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 VII
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

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 is shown in Tables III and IV. We found that no significant gain 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 SROCC (between predicted and subjective DMOS score on all distortions in the database mixed together) was observed to increase with the size of the training set, but the drop in the correlations when the training set was reduced in size was not significant. The results are shown in Fig. 11. Notice that an SROCC of 0.85 is obtained when using only 30% of the content for training, and that the knee of the curve occurs at roughly 20%. This shows that our reported results are not tainted by overtraining or overfitting to the training data.

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

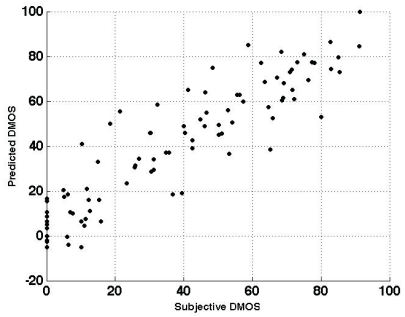

Fig. 12. Predicted versus subjective DMOS on the JPEG2000 database subset.

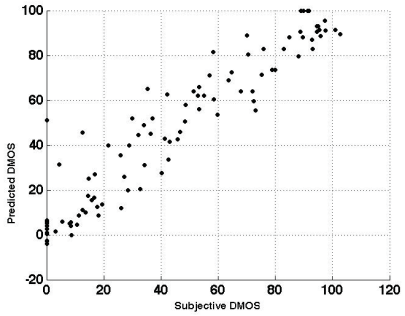

Fig. 13. Predicted versus subjective DMOS on the JPEG database subset.

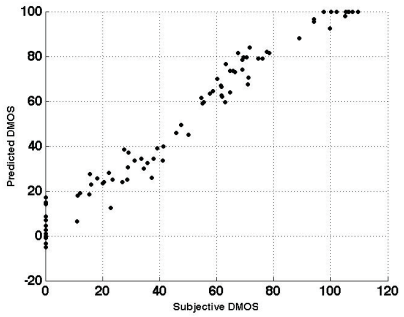

Fig. 14. Predicted versus subjective DMOS on the white noise database subset.

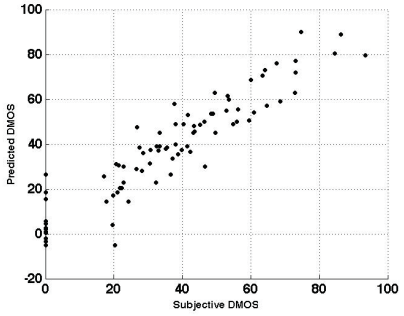

Fig. 15. Predicted versus subjective DMOS on the Gaussian blur database subset.

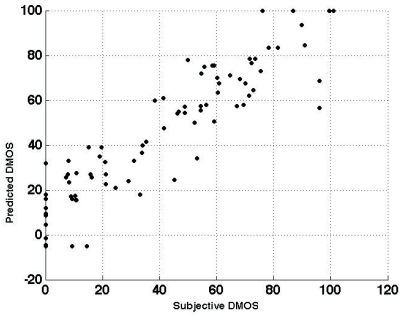

Fig. 16. Predicted versus subjective DMOS on the fast-fading channel distortions database subset.

in [46], 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 [15] and DIIVINE [17], the *full-reference* PSNR, and the state-of-the-art FR-IQA SSIM index.

The SROCC and LCC results[2] are shown in Tables V–VII. Tables V and VI 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 VII reports the SROCC and LCC results of PSNR and SSIM (the implementation in [47][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, BLIINDS-II outperforms BIQI [15] 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.

[2]Algorithms were trained and tested on individual distortions and on all distortions mixed together.

[3]Using default C_1 and C_2 parameters.

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. 12–17. 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. 18 and 19, respectively. We report the standard deviation of the SROCC and LCC results on the 1000 trials for each algorithm in Table VIII. Obviously, the lower the standard deviation with a higher median SROCC, the better the performance. The plots show that SSIM, DIIVINE, and BLIINDS-II are not statistically significantly different in performance (knowing they are designed for different application domains).

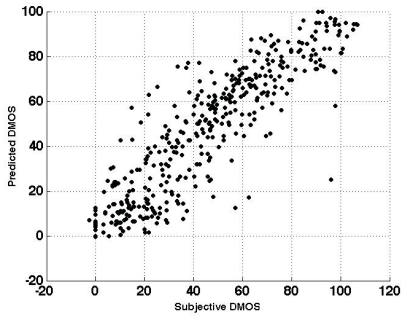

Fig. 17. Predicted versus subjective DMOS on the entire LIVE IQA database.

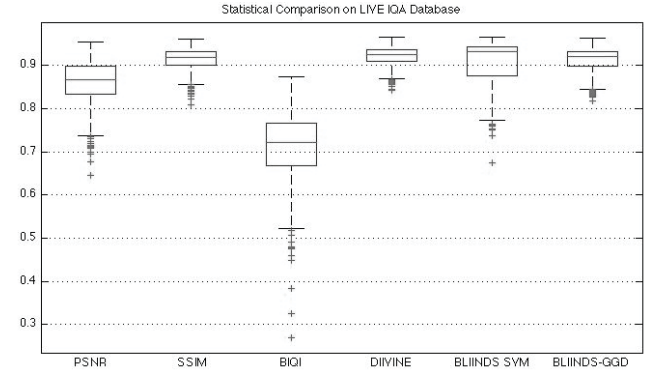

Fig. 18. Box plot of SROCC distributions of the algorithms over 1000 trials for algorithm comparison on the LIVE IQA database.

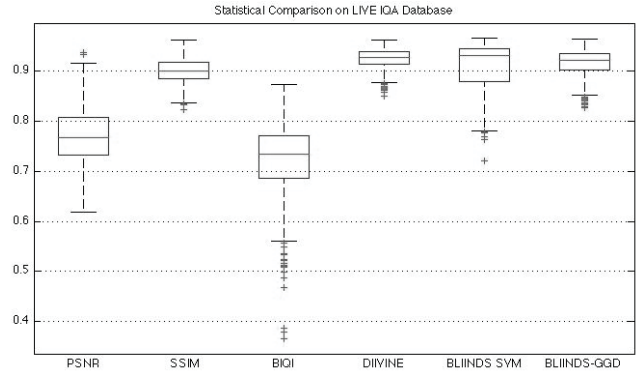

Fig. 19. Box plot of LCC distributions of the algorithms over 1000 trials for algorithm comparison on the LIVE IQA database.

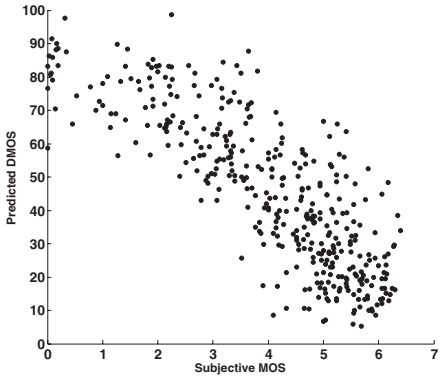

Fig. 20. Predicted DMOS versus subjective MOS on the TID2008 database.

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 BLIINDS-II 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 IX. 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.

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 [44]. 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 X. 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. 20.

TABLE VIII
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 IX
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 X
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 $m \times k$ be the image dimension, and let $n \times n$ 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 $m \times k/n^2 \times n^2 logn = m \times k \times logn$. 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 $O(n^2 logn)$ [48], where n is the dimension of the block (i.e., the block is $n \times n$). 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 [34]. The complexity of such an operation is on the order $O(log(1/\epsilon))$. ϵ was chosen to be 0.001, and hence $log(1/\epsilon) << min(m, k)$.

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 nonreliant 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 multidistortion 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-distorion 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 iden-

tifying 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 (subsampled) 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.

REFERENCES

[1] 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.

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

[3] 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.

[4] 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.

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

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] X. Feng and J. P. Allebach, "Measurement of ringing artifacts in JPEG images," *Proc. SPIE*, vol. 6076, pp. 74–83, Jan. 2006.

[13] M. Jung, D. Léger, and M. Gazalet, "Univariant assessment of the quality of images," *J. Elect. Imag.*, vol. 11, no. 3, pp. 354–364, Jul. 2002.

[14] C. Charrier, G. Lebrun, and O. Lezoray, "A machine learning-based color image quality metric," in *Proc. 3rd Euro. Conf. Color Graphics Imag. Vision*, Jun. 2006, pp. 251–256.

[15] A. K. Moorthy and A. C. Bovik, "A two-step framework for constructing blind image quality indices," *IEEE Signal Process. Lett.*, vol. 17, no. 5, pp. 513–516, May 2010.

[16] T. Brandao and M. P. Queluz, "No-reference image quality assessment based on DCT-domain statistics," *Signal Process.*, vol. 88, no. 4, pp. 822–833, Apr. 2008.

[17] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3350–3364, Dec. 2011.

[18] M. A. Saad, A. C. Bovik, and C. Charrier, "A DCT statistics-based blind image quality index," *IEEE Signal Process. Lett.*, vol. 17, no. 6, pp. 583–586, Jun. 2010.

[19] R. Blake and R. Sekuler, *Perception*, 5th ed. New York: McGraw Hill, 2006.

[20] W. S. Geisler, "Visual perception and the statistical properties of natural scenes," *Annu. Rev. Psychol.*, vol. 59, pp. 167–192, Jan. 2008.

[21] Q. Li and Z. Wang, "Reduced-reference image quality assessment using divisive-normalization-based image representation," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 2, pp. 202–211, Apr. 2009.

[22] J. Malo and V. Laparra, "Psychophysically tuned divisive normalization approximately factorizes the PDF of natural images," *Neural Comput.*, vol. 22, no. 12, pp. 3179–3206, 2010.

[23] A. K. Moorthy and A. C. Bovik, "Visual importance pooling for image quality assessment," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 2, pp. 193–201, Apr. 2009.

[24] K. Seshadrinathan and A. C. Bovik, "Motion tuned spatio-temporal quality assessment of natural videos," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 335–350, Feb. 2010.

[25] G. E. Legge and J. M. Foley, "Contrast masking in human vision," *J. Opt. Soc. Amer.*, vol. 70, no. 12, pp. 1458–1471, 1980.

[26] H. B. Barlow, "Redundancy reduction revisited," *Netw. Comput. Neural Syst.*, vol. 12, no. 3, pp. 241–253, 2001.

[27] S. Gabarda and G. Cristobal, "Blind image quality assessment through anisotroy," *J. Opt. Soc. Amer.*, vol. 24, no. 12, pp. B42–B51, Dec. 2007.

[28] W. S. Geisler and J. S. Perry, "Contour statistics in natural images: Grouping across occlusions," *Visual Neurosci.*, vol. 26, no. 1, pp. 109–121, 2009.

[29] A. K. Moorthy and A. C. Bovik, "Perceptually significant spatial pooling strategies for image quality assessment," *Proc. SPIE Human Vis. Electron. Imag.*, vol. 7240, pp. 724012-1-–724012-11, Jan. 2009.

[30] A. Srivastava, A. B. Lee, E. P. Simoncelli, and S. C. Zhu, "On advances in statistical modeling of natural images," *J. Math. Imag. Vision*, vol. 18, no. 1, pp. 17–33, 2003.

[31] H. R. Sheikh, A. C. Bovik, and G. De Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2117–2128, Dec. 2005.

[32] A. C. Bovik, T. S. Huang, and D. C. Munson, "A generalization of median filtering using linear combinations of order statistics," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 31, no. 6, pp. 1342–1350, Dec. 1983.

[33] H. G. Longbotham and A. C. Bovik, "Theory of order statistic filters and their relationship to linear FIR filters," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 37, no. 2, pp. 275–287, Feb. 1989.

[34] K. Sharifi and A. Leon-Garcia, "Estimation of shape parameter for generalized gaussian distributions in subband decompositions of video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, no. 1, pp. 52–56, Feb. 1995.

[35] A. C. Bovik, M. Clark, and W. S. Geisler, "Multichannel texture analysis using localized spatial filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 1, pp. 55–73, Jan. 1990.

[36] Y. Shain, A. Akerib, and R. Adar, "Associative architecture for fast DCT," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, May 1998, pp. 3109–3112.

[37] J. Huan, M. Parris, J. Lee, and R. F. DeMara, "Scalable FPGA-based architecture for DCT computation using dynamic partial reconfiguration," *ACM Trans. Embedded Comput. Syst.*, vol. 9, no. 1, pp. 1–18, Oct. 2009.

[38] S. Balam and D. Schonfeld, "Associative processors for video coding applications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 2, pp. 241–250, Feb. 2006.

[39] P. Duhamel, C. Guillemot, and J. C. Carlach, "A DCT chip based on a new structured and computationally efficient DCT algorithm," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 1990, pp. 77–80.

[40] N. I. Cho and S. U. Lee, "Fast algorithm and implementation of 2-D discrete cosine transform," *IEEE Trans. Circuits Syst.*, vol. 38, no. 3, pp. 297–305, Mar. 1991.

[41] M. Haque, "A 2-D fast cosine transform," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 33, no. 6, pp. 1532–1539, Dec. 1985.

[42] N. Bozinovic and J. Konrad, "Motion analysis in 3D DCT domain and its application to video coding," *Signal Process. Image Commun.*, vol. 20, no. 6, pp. 510–528, Jul. 2005.

[43] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik. *LIVE Image Quality Assessment Database Release 2* [Online]. Available: http://live.ece.utexas.edu/research/quality

[44] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti, "TID 2008 a database for evaluation of full-reference visual quality assessment metrics," *Adv. Modern Radioelectron.*, vol. 10, pp. 30–45, 2009.

[45] M. H. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," *IEEE Trans. Broadcast.*, vol. 10, no. 3, pp. 312–322, Sep. 2004.

[46] A. Karatzoglou, A. Smola, K. Hornik, and A. Zeileis, "Kernlab an S4 package for kernel methods in R," *J. Statist. Software*, vol. 11, no. 9, pp. 1–20, 2004.
[47] Z. Wang, A. C. Bovik, and H. R. Sheikh (2004). *Image Quality Assessment: From Error Visibility to Structural Similarity.* [Online]. Available: https://ece.uwaterloo.ca/z70wang/research/ssim/
[48] W. H. Chen, C. H. Smith, and S. Fralick, "A fast computational algorithm for discrete cosine transform," *IEEE Trans. Commnu.*, vol. 25, no. 9, pp. 1004–1009, Sep. 1977.

Michele A. Saad (S'07) received the B.E. degree in computer and communications engineering from the American University of Beirut, Lebanon, in 2007, and the M.S. degree in electrical and computer engineering from the University of Texas, Austin, in 2009, where she is currently pursuing the Ph.D. degree in electrical and computer engineering.

Her current research interests include statistical modeling of images and videos, motion perception, design of perceptual image and video quality assessment algorithms, and statistical data analysis.

Ms. Saad was the recipient of the Microelectronics and Computer Development Fellowship from the University of Texas from 2007 to 2009. She is a member of the Laboratory of Image and Video Engineering and the Wireless Networking and Communications Group, University of Texas.

Alan C. Bovik (F'96) is the Curry/Cullen Trust Endowed Chair Professor with the University of Texas, Austin, where he is the Director of the Laboratory for Image and Video Engineering. He is also a Faculty Member with the Department of Electrical and Computer Engineering and the Center for Perceptual Systems, Institute for Neuroscience. He is a Professional Engineer of the State of Texas and is a frequent consultant to legal, industrial, and academic institutions. He has published over 600 technical articles in these areas and holds two U.S. patents. His several books include the recent companion volumes, *The Essential Guides to Image and Video Processing* (Academic Press, 2009). His current research interests include image and video processing, computational vision, and visual perception.

Prof. Bovik was the recipient of the SPIE/IS&T Imaging Scientist of the Year for 2011, as well as a number of major awards from the IEEE Signal Processing Society, including the Best Paper Award in 2009, the Education Award in 2007, the Technical Achievement Award in 2005, and the Meritorious Service Award in 1998. He received the Hocott Award for Distinguished Engineering Research, University of Texas, the Distinguished Alumni Award from the University of Illinois at Urbana-Champaign, Urbana, in 2008, the IEEE Third Millennium Medal in 2000, and two Journal Paper Awards from the International Pattern Recognition Society in 1988 and 1993. He is a fellow of the Optical Society of America, the Society of Photo-Optical and Instrumentation Engineers, and the American Institute of Medical and Biomedical Engineering. He has been involved in numerous professional society activities, including Board of Governors, the IEEE Signal Processing Society from 1996 to1998. He is the co-founder and Editor-in-Chief of the IEEE TRANSACTIONS ON IMAGE PROCESSING from 1996 to 2002. He was on the Editorial Board of the PROCEEDINGS OF THE IEEE from 1998 to 2004, Series Editor for *Image, Video, and Multimedia Processing* (Morgan and Claypool Publishing Company, 2003), and Founding General Chairman of the first IEEE International Conference on Image Processing, Austin, in 1994.

Christophe Charrier (M'10) received the M.S. degree from the Nantes University of Science and Technology, Nantes, France, in 1993, and the Ph.D. degree from the University Jean Monnet, Saint-Étienne, France, in 1998.

He has been an Associate Professor with the Communications, Networks and Services Department, Cherbourg Institute of Technology, Cherbourg, France, since 2001. From 1998 to 2001, he was a Research Assistant with the Laboratory of Radio Communications and Signal Processing, Laval University, Quebec, QC, Canada. In 2008, he was a Visiting Scholar with the Laboratory for Image and Video Engineering, University of Texas, Austin. From 2009 to 2011, he was an Invited Professor with the Computer Department, University of Sherbrooke, Sherbrooke, QC, Canada. His current research interests include digital image and video coding, processing, quality assessment, and computational vision.