Zhou Wang and Alan C. Bovik

Reduced- and No-Reference Image Quality Assessment

The natural scene statistic model approach

Recent years have witnessed dramatically increased interest and demand for accurate, easy-to-use, and practical image quality assessment (IQA) and video quality assessment (VQA) tools that can be used to evaluate, control, and improve the perceptual quality of multimedia content in a wide variety of practical multimedia signal acquisition, communication, and display systems. There is a vast and increasing proliferation of such content over both wireline and wireless networks. Think of the Internet: Youtube, Facebook, Google Video, Flickr and so on; networked high-definition television (HDTV), Internet Protocol TV (IPTV) and unicast home video-on-demand (Netflix and Hulu, for example); and an explosion of wireless video traffic that is expected to more than double every year over the next five years [1]. In such an environment of extreme growth, limited bandwidths, and diverse content, resolutions, and quality, there is considerable concern regarding how the quality of service (QoS) of videos being delivered can be managed. In short, there is currently no practical method for accurately monitoring the perceptual quality of this vast proliferation of video data.

A number of successful algorithms have been created that can predict subjective visual quality of a distorted image or video signal—in agreement with human opinions of visual quality—when a (presumed) "pristine" signal is fully available [2].

Digital Object Identifier 10.1109/MSP.2011.942471
Date of publication: 1 November 2011

Yet, in most present and emerging practical real-world visual communication environments, such full-reference (FR) methods are not useful since the reference signals are not accessible at the receiver side (or perhaps at all). What are really needed are I/VQA algorithms that can operate with little or no reference signal information at all; in other words, by operating on the visual signal of interest directly, rather than by extensive comparison. Indeed, the assumption of a supposedly "pristine" reference image or video is highly suspect; even under the most ideal and controlled circumstances, a captured optical signal will inevitably suffer from some kind of distortion [3].

CREATING AUTONOMOUS ALGORITHMS THAT DEPEND ON MUCH LESS SPECIFIC INFORMATION FROM ANY REFERENCE SIGNAL IS NOW AN INTENSE FOCUS OF RESEARCH.

Thus, creating autonomous algorithms that depend on much less specific information from any reference signal is now an intense focus of research. Such algorithms fall into two categories: reduced-reference (RR) and no-reference (NR) (or blind) IQA and VQA algorithms. In the former category, a reference signal is assumed only partially accessible (in the form of selected features); typically, the amount of data from the reference signal is significantly less than in the reference signal itself. In the latter category, reference signal information is deemed completely inaccessible [3]. Although RR and NR algorithms (that accord with perception) have been desired for a very long time, progress has been slow. Yet the need is large since today's video consumers have become increasingly savvy about the capabilities of digital video, and expectations regarding the QoS of delivered visual multimedia have risen significantly.

If such RR and/or NR visual QA algorithms could be created, they could be deployed as agents over wide-area data communications and visual surveillance networks by embedding them in smart routers, set-top boxes, smart phones, cameras, tablets and laptops. They could be used as primary QoS tools that could feed back time-varying visual signal quality information, enabling source adaptation and distributed network control mechanisms to adapt resource allocation, source and channel coding, and other network parameters. In today's increasing video-centric consumer data communications environment, we think that such algorithms could represent a sea change in visual multimedia data delivery.

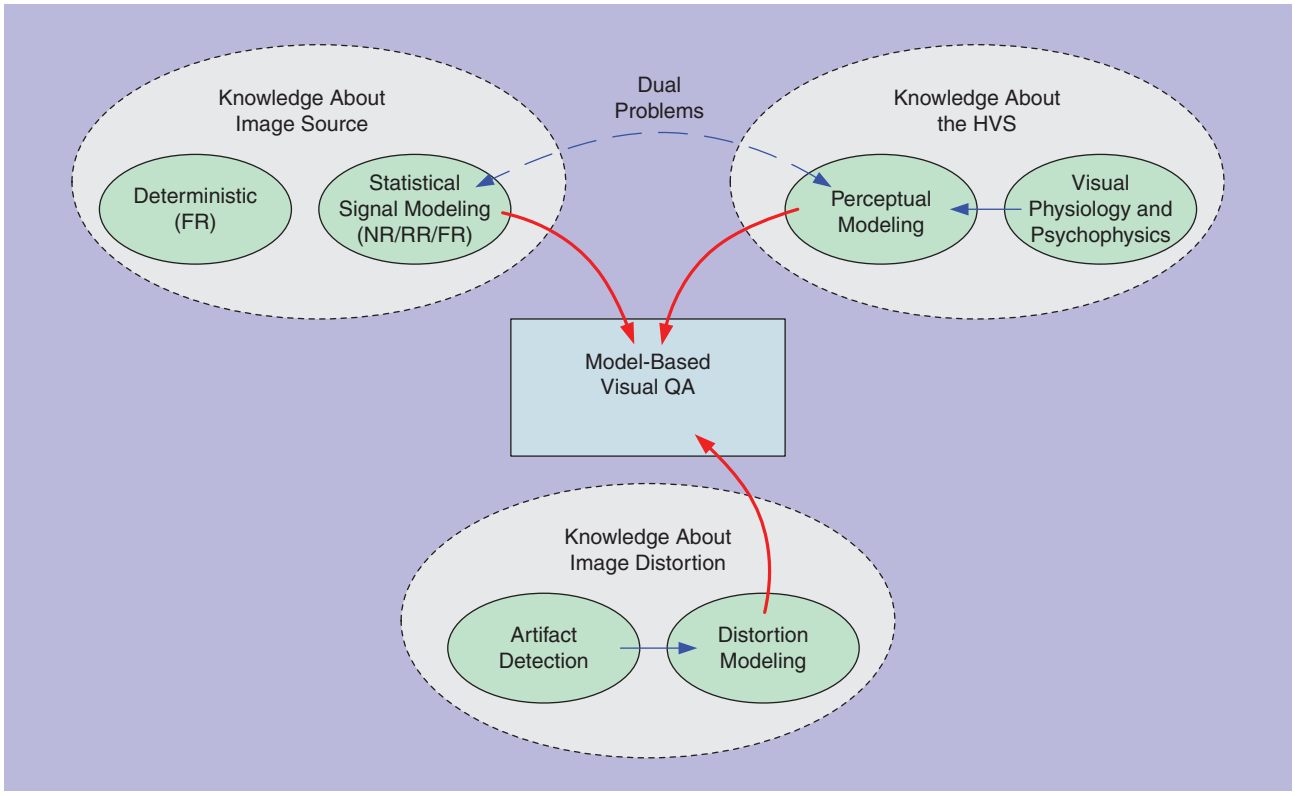
To date, a wide variety of inventive methods have been deployed toward solving the RR and NR I/VQA problems, and the universe of ideas are quite large. In attempting to describe and help the readers to understand the field, we are confronted also with the fact that the various approaches attempt to solve diverse problems that operate under different assumptions. It is our goal in this tutorial to clarify the issues to be solved and how they might be pragmatically approached, without clouding things by attempting a broad survey of the field. In doing so, we take a certain viewpoint regarding how things ought to be done.

TOWARD MODEL-BASED VISUAL QA

To better cast the foregoing discussion against the "big picture," Figure 1 depicts a "knowledge map" that contains what we regard as the essential building blocks in the design of successful visual QA models. We think that accurate modeling is the key to successful I/VQA algorithm design. Essentially, three types of knowledge may be exploited to build such models. The first is knowledge about the image source that captures the essence of what a signal ought to look like when not distorted. This can be either deterministic when the reference signal is fully accessible (FR case) or statistical when certain statistical models are available to regulate the undistorted images. The second is knowledge about image distortion, which may help detect particular artifacts created by specific distortion processes (e.g., blocking artifacts generated in JPEG compression). Furthermore, mathematical models of how distortions change image content (by degradation process, artifacts, or loss of statistical naturalness) can be used to predict distortion severity in an observed visual signal. Finally, since most applications direct visual signals toward human viewers, the third type of knowledge is about the HVS, which is based on perceptual models originated from visual physiology and psychophysical studies. We believe that models derived from these three types of knowledge should not be disjoint. In particular, statistical signal modeling and perceptual modeling may be understood as dual problems, as discussed later. By analogy, communication systems embody knowledge of transmitter (analogously visual signal model), channel (distortion model), and receiver (perceptual model). The more information that is available regarding transmitter, channel, and receiver, the better job of communication that can be accomplished. The most successful design will use and combine models of all three. Most visual QA algorithms can be understood using such a unified modeling framework, even if not expressed in such terms.

NATURAL SCENE STATISTIC MODELS

We believe that there is a category of statistical models that comes close to embodying the three-fold modeling objectives just described, and that provides the most promising basis for successful RR and NR QA algorithm design. As we shall see, these so-called natural scene statistic (NSS) models are highly attractive in a number of ways: they reliably capture low-level statistical properties of images (hence are very general and flexible models); they can be used to measure the destruction of "naturalness" introduced by distortions (enabling effective distortion models); and they accurately describe the statistics to which the visual apparatus has adapted and evolved over the millennia (and so, are regarded as direct duals of low-level perceptual models). A "natural scene" is one captured by an optical camera, and can include both naturalistic (e.g., trees and grass) content as well as man-made indoor and outdoor scenery. The term is meant to distinguish "natural" from artificial image



[FIG1] **Knowledge map expressing the elements required to construct successful visual QA models.**

creation processes such as computer graphics. We will describe specific NSS models that appear to be well suited for RR and NR QA algorithm design. We will also point out ways in which NSS models can be improved for QA applications, e.g., by incorporating perceptual information into them.

NSS models seek to capture the natural statistical behavior of images, rather than assuming deterministic knowledge of the image source (as in FR QA). Such prior models of image statistics enables the use of a rich groundwork of Bayesian statistical methods, and are rooted in the widely accepted view of biological perceptual systems in computational neuroscience and psychophysics, that the visual apparatus is highly adapted to the natural environment, and has evolved to most efficiently extract visual information from it [4], [5].

What constitutes a useful model of natural image or video statistics? The most important criteria is that the model be regular, in the sense that any natural image that has not been distorted by unnatural distortions can be expected to follow the model with a high degree of confidence. In recent years, a number of NSS models have been derived for still images that are highly regular. Common NSS models used for both modeling of perception and for image processing applications have been developed around theories of sparse coding by the visual brain [4], [18], [19] and by the observed (self-similar) scaling properties of natural images [28], [29]. Importantly, the visual brain appears to have both to have evolved to "match" the statistics of natural images [19] and to seek a efficient, decorrelated representations of image information, as evidenced by the fact that the principal components (or independent components) of natural images closely resemble the spatial responses of cortical neurons [61].

While it is beyond the scope of this article to describe the function and coding processes in visual cortex, or the broader spectrum of NSS models that have been proposed (a good introduction can be found in [62]), the connections between NSS and perceptual processes are important, since the ultimate goal of objective visual QA algorithms is to predict human behavioral responses when evaluating visual quality. Of particular importance in this context are NSS models that are sensitive to image "unnaturalness" introduced by distortion. The NSS models that we describe later on are quite useful in this regard, as they can be used to successfully predict the type and degree of perceptual quality loss introduced by common distortions.

Of course, the QA of moving pictures, or videos, can be accomplished in a limited manner by applying still-image NSS models on a frame-by-frame basis. More desirable, however, would be NSS models that capture the statistics of naturalistic videos in a natural manner. There is evidence that the spatio-temporal vision system has adapted to the natural statistics of moving images to achieve efficient encoding of the large volume of data [63]. While some progress has been made on characterizing the statistics of optical flow fields under rather rigid

assumptions [64], there does not yet exist any established statistically regular model of the natural spatiotemporal statistics of video data. The problem is greatly complicated by the complexity of the natural motions of objects, and of the sensor. As such, NSS-based VQA remains very much an open problem. Because of this, the design of (FR) VQA algorithms that are not distortion-specific has been largely driven by structural or perceptual models [35], [55], [65], [66].

THE IDEA OF RR QA WAS FIRST CONCEIVED IN THE 1990s AS A PRAGMATIC APPROACH TO REAL-TIME VIDEO QUALITY MONITORING OVER MULTIMEDIA COMMUNICATION NETWORKS.

STATISTICAL IMAGE QA

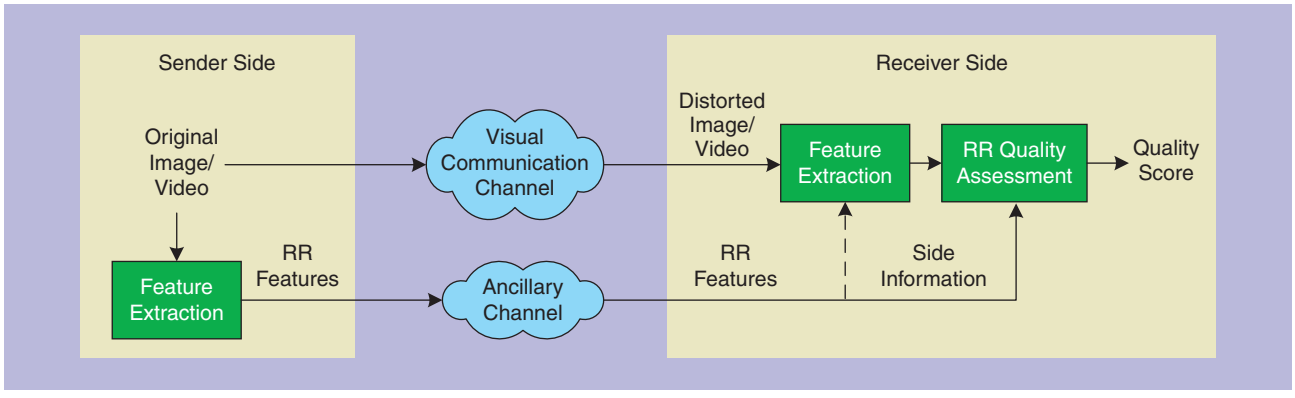
In essence, NSS-based IQA algorithms seek to capture statistical regularities of natural images and to quantify how these regularities are modified or lost when distortions occur. Since these methods do not necessarily rely on directly detecting or quantifying specific image artifacts, such algorithms have the potential to be more widely applicable than distortion-specific approaches. Such "holistic" NR QA algorithms could operate by measuring a "distance" from naturalness, thereby gauging how severely distorted a visual signal is by how "far" it lies from the space of natural images. However, it is also possible to distinguish distortions by type according to the "direction" in which the distorted signal lies away from "NSS space."

The first successful NSS model-based IQA algorithm was the FR Visual Information Fidelity (VIF) index [6], which uses a wavelet-domain Gaussian scale mixture (GSM) model (as described later in the context of RR algorithms) [7] that captures both the marginal distributions of wavelet coefficients and the magnitude dependencies of neighboring coefficients across space, scale, and orientation. This algorithm has exhibited excellent performance in two large human studies [8], [9]. The GSM was also used to modify the FR multiscale structural similarity (SSIM) index [10] by weighting local information content. The resulting information-weighted SSIM (IW-SSIM) index delivers superior performance relative to the state of the art against human subjectivity, as shown on multiple public databases [11]. These results further motivate the use of NSS for the design of visual QA algorithms. How then, to use these models when the amount of information from the reference image is reduced or eliminated? As we will show, promising results have been achieved using NSS for still image RR and NR QA, for both distortion-specific and "holistic" problems.

TOWARD RR IMAGE QA

The idea of RR QA was first conceived in the 1990s [12] as a pragmatic approach to real-time video quality monitoring over multimedia communication networks. Figure 2 depicts the general idea of how an RR image or video QA system works [3]. At the sender side, a feature extractor is applied to the reference visual signal. The extracted features are transmitted to the receiver as side information through an ancillary channel. It is usually assumed that the ancillary channel is error free. When the distorted signal is transmitted to the receiver via an error-prone channel, a feature extractor is also applied at the receiver side. This could be the same process as at the sender, or it might be adapted according to the received side information. In the final QA stage, the RR features extracted from both reference and distorted visual signals are used to compute an overall score indicating the quality of the distorted signal.

A good RR approach must achieve a good balance between the accuracy of the quality predictions and the RR data rate. One would expect that accuracy would improve monotonically as more information about the reference image is made available [3]. RR algorithms lie within two extremes: if the data rate is high enough to deliver the reference signal as side information, then an FR method can be applied at the receiver side. Conversely, if the data rate is zero (i.e., no reference side information), then an NR method is required. In practice, a maximum RR data rate is specified, which is usually quite low, since bandwidth for reference side information is effectively "stolen," since it could be used to improve the quality of the transmitted



[FIG2] General framework of an RR image or image QA system.

visual signal. This limited data rate makes the design of RR algorithms a challenging task. It puts strong constraints on the selection of RR features, which constitute the most critical component of RR algorithms. A good set of RR features should

- efficiently summarize the content of the reference visual signal
- be sensitive to specific distortions or (if of the holistic variety) be sensitive to a broad spectrum of image distortion types
- embed aspects of the signal that are perceptually relevant.

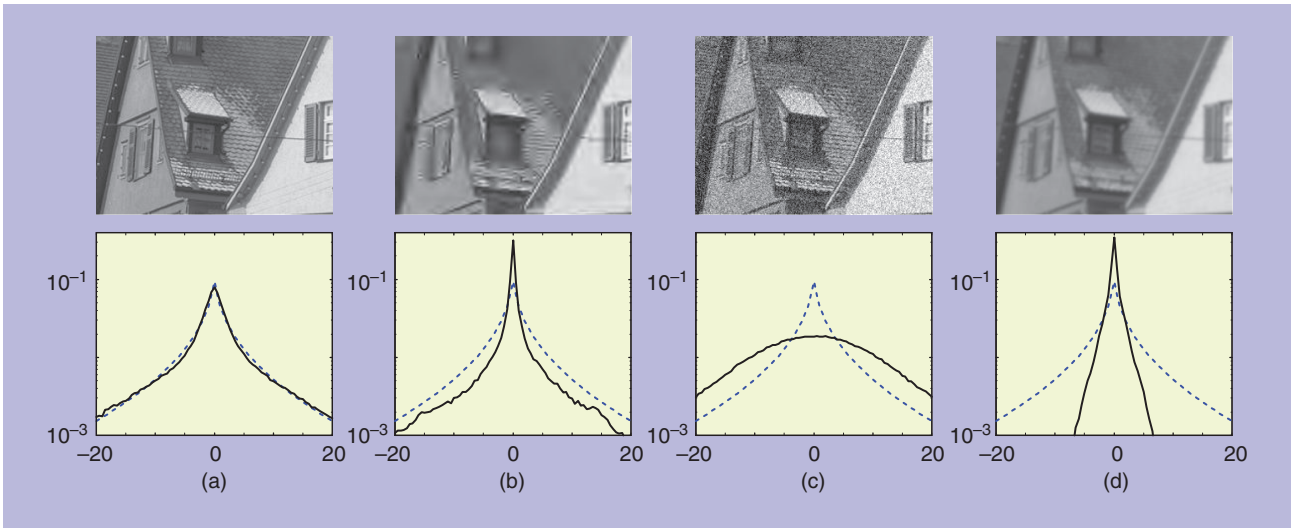
The significance of these properties can be demonstrated by a naïve example: At the sender side, randomly select image pixels (say, 1% of them) as RR features. When (side) transmitted to the receiver, they are compared on a pixel-wise basis with those in the received signal, so that the mean-squared error (MSE) or peak-signal-to-noise ratio (PSNR) between reference and received signals can be estimated. This approach is weak in several regards. First, it is difficult to keep the RR data rate low—even 1% of the pixels in a 512×512, 8 b/pixel image requires transmitting 20,976 b. An additional 47,196 b are required if the positions of the randomly selected pixels are also transmitted. This is a heavy burden—much greater than the NSS-based RR methods that we will discuss later. Second, the RR features sparsely sample the reference and do not adequately summarize the image. Third, some distortions may change only pixels not selected as RR features, and thus not be easily detected. Finally, the MSE and PSNR have poor correlations relative to the perception of visual quality [3], [8], [13], [14].

The problems exhibited in the above naïve example are instructive. Clearly, RR features should more efficiently summarize image information content, be more sensitive to image distortions, and have stronger perceptual relevance. NSS modeling provides a powerful means to approach these goals. To demonstrate this, we use an RR algorithm proposed in [15] and [60], where an NSS model of the marginal distribution of the image wavelet coefficients is employed. Owing to space limitations, we must assume that the reader is conversant with the discrete wavelet transform; otherwise an excellent easy tutorial and a deep treatment can be found in [22] and [23], respectively.

The choice of wavelet space for statistical modeling of images has a number of important underpinnings. Images are naturally multiscale and the early visual system decomposes image information in a multiscale manner, thereby representing visual signals simultaneously in localized space, frequency, and orientation [16], [17]. More importantly, natural images exhibit statistical regularities in wavelet space [5]. For example, it has been observed that the marginal distributions of natural image wavelet coefficients consistently have sharp peaks near zero and longer tails than Gaussian. This reflects specific intuitive properties of images of the real world: most of the world (and images of it) is smooth (hence many near-zero wavelet or bandpass responses). This smoothness is broken up by sparse, often large amplitude discontinuities (hence relatively many large bandpass responses). Such highly kurtotic distributions have important implications with respect to the sensory neural coding of natural scenes [18], and are a central focus of recent theories on the evolution and function of biological vision systems [19].

Figure 3 shows the histograms of the wavelet coefficients from a subband of a natural image and several distorted versions of it. A discovery in the literature of NSS is that the marginal distribution of the wavelet coefficients of natural images can be consistently well fitted by a two-parameter generalized Gaussian density (GGD) model with high accuracy [20]

$$p_m(x) = \frac{\beta}{2\,\alpha\,\Gamma(1/\beta)} \exp[-(|x|/\alpha)^{\beta}], \tag{1}$$



[FIG3] Wavelet coefficient histograms (solid curves) of (a) original "buildings" image; (b) compressed by JPEG2000; (c) with additive white Gaussian noise; and (d) blurred by a linear Gaussian kernel. The histogram in (a) is well fitted by a generalized GGD model (dashed curves). The shapes of the histograms (and the GGD fits) change in different ways for different types of distortions.

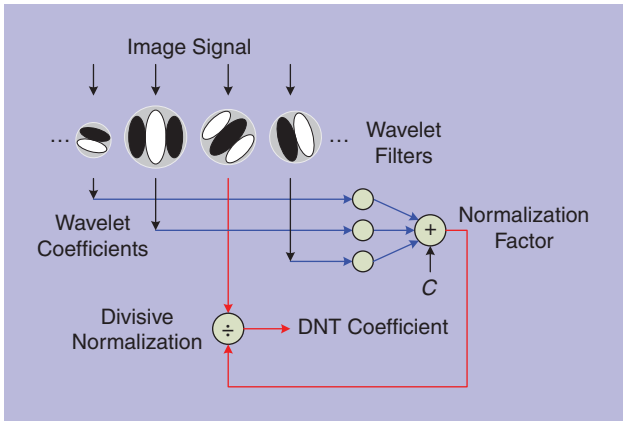
where $\Gamma(a) = \int_0^\infty t^{a-1}e^{-t}dt$ (for $a > 0$) is the Gamma function. Figure 3(a) also depicts GGD fit of the histogram of the natural images (dashed curves), which very closely approximates the true distribution (solid curve). This is important since only two parameters $\{\alpha, \beta\}$ are required to summarize the reference image coefficient histograms.

Evident from Figure 3(b)–(d) is the fact that the marginal distributions of the wavelet coefficients can change in diverse ways as the image undergoes different distortions. This is ideal for RR QA, since departures from the reference distributions (characterized by $\{\alpha, \beta\}$) can be used as a common measure to quantify the degree of distortions. One such measure is the Kullback-Leibler divergence (KLD) [21] between the model (1) and the marginal distribution of the distorted signal $q(x)$, viz., between the solid and dashed curves in Figure 3(b)–(d)

$$d(p_m \| q) = \int p_m(x) \log \frac{p_m(x)}{q(x)} dx. \tag{2}$$

This approach was used in [15] to create a holistic algorithm that achieves competitive performance relative to the full reference PSNR, using a side channel RR data rate of only 162 b/image. This is powerful evidence of the efficacy of this NSS model for IQA.

A number of earlier statistical approaches operated without any models, by extracting local sample statistics from the image. For example, in [22], simple local statistical descriptors of spatiotemporal edge and orientation activity were extracted from video sequences to form an RR video QA index for monitoring perceptual degradations in visual communication systems. In [23], local harmonic magnitudes were extracted from local image patches containing edges to create a harmonic activity map. The authors showed that different types of distortions (blocking and blurring) alter these maps in different ways, thereby providing a way to evaluate image quality with reduced reference. In [24], perceptually motivated "structural information features" (orientation, length, width, and contrast) were extracted near predicted visual fixation points. These RR features were transformed to be invariant to zoom, translation, and rotation, and stored in a database of "visual image memory." It shows good performance when predicting the quality of images compressed by JPEG and JPEG2000. In [25], the authors attempt to model the visual pathway from "front end" to cortex (starting from display, through the eyes, and ending in visual cortex), producing a set of local statistics that provide a "reduced description" of an image.



[FIG4] A simple divisive normalization scheme. A wavelet response (red output) is normalized by the summed responses of a cluster of wavelets neighboring in space, scale, and frequency. The constant *C* accounts for the saturation effect and stabilizes the DNT when the neighbor responses are low in energy. This model accords closely with neurological and psychophysical evidence, nicely explains the perceptual masking effect, and improves visual QA algorithms.

We believe it is useful to design perceptually motivated approaches using NSS-based frameworks, making it thereby possible to use Bayesian methods to achieve statistically and perceptually optimized QA. Such a stratagem is taken in [26], where NSS and perceptual models are combined in just such a manner, as described next.

Our goal here is not to give a tutorial on current models of neural processing, but there are certain aspects that require explanation in order that their relevance to QA be understood. As mentioned earlier, neurons in visual cortex effectively perform frequency- and orientation-selective waveletlike decompositions of visual data arriving from the two eyes. A nonlinear neural mechanism that has been observed is adaptive gain control (AGC), whereby each neuron's response (or in our case, wavelet coefficient response) is divided by the energy of a cluster of neighboring neuronal responses (neighboring wavelet coefficients in space, scale, and orientation) [27], as depicted in Figure 4. Such a divisive normalization transform (DNT) of the neuronal (wavelet) responses has been shown to significantly reduce statistical dependencies between the responses (coefficients) which can lead to efficient representation. Further, the statistics of the normalized coefficients have been shown to closely follow Gaussian marginal distributions [28]. More importantly, this DNT deeply affects the degree of visibility of image distortions.

The DNT is a critical perceptual model that is both useful for IQA algorithm design and that melds seamlessly with the Gaussian scale mixture (GSM) model mentioned earlier in the context of the VIF index. In fact, VIF uses a form of divisive normalization [6].

We describe the GSM model next. Suppose that $\mathbf{y}$ is a vector of wavelet responses that are locally clustered over neighboring space, scales and/or orientations. Then the GSM model is given by $\mathbf{y} = z\mathbf{u}$, where $\mathbf{u}$ is a multidimensional zero-mean Gaussian random vector, and z is a scalar random variable called a mixing multiplier. If we assume that z takes a fixed value for each selected cluster of wavelet coefficients, then putting all z values constitutes a variance field. If an accurate estimate $\hat{z}$ of z can be found for

each coefficient cluster, then dividing the observed vector of coefficients by $\hat{z}$, which accomplishes the DNT, produces a random vector that is Gaussian. This is a form of conditioning of $\mathbf{y}$ given knowledge of the variance field.

In [26], the authors used a maximum likelihood procedure [29] to estimate z and observed that the distribution of the DNT coefficients (or conditioned wavelet coefficients undergone DNT) of natural images are Gaussian, but changes in different ways in images altered with different types of distortions. Based on this observation, they form a DNT-based RR IQA algorithm, which computes the KLD (2) between the DNT coefficient histogram in the distorted image versus the best Gaussian fit to the DNT coefficients of the reference image. In their RR implementation, four features are computed from each wavelet subband: the above KLD, and the variance, kurtosis, and skewness of the DNT coefficients from that band. Their wavelet decomposition, which is a steerable pyramid [30], is taken over three scales and four orientations, yielding just 48 pieces of RR information. Yet the algorithm does quite well as measured against human subjectivity, matching the performance of the widely used FR index PSNR.

The DNT is relevant to a wide variety of neuroscience, perceptual, engineering, and in particular, QA issues. Since the DNT reduces the dependencies between the wavelet coefficients (or neural responses) over local space-scale-orientation regions, it supports the efficient coding hypothesis of early biological vision, wherein as much redundancy in representation is eliminated from low-level visual information before higher-level processes act upon it [4]. It also serves as a model of AGC, which serves to limit the dynamic range of retinal signals. AGC or DNT has an important perceptual byproduct that is easily observed: visual masking.

Visual masking is a process whereby one element of a visual signal reduces the visibility of another [31], typically of similar characteristics such as frequency or orientation. Figure 5 is an easy-to-see example, where the image of the woman is distorted everywhere by the same level of additive Gaussian noise. Although the noise statistics are unchanged across the image, it is only highly visible on the smooth regions (e.g., face), and much less visible on the more "textured" hair and scarf, and nearly imperceptible on the wicker chair backing. This effect occurs with other distortions that introduce artificial high frequencies, such as JPEG compression [32]. The significance of masking on the obscuration of spatial image distortions was observed by Girod [33] as well as Teo and Heeger [34], who proposed masking models not dissimilar to the DNT outlined above. The most successful FR IQA and VQA algorithms, such as SSIM [10] and its derivations [14], VIF [6], and Motion-Based Video Integrity Evaluator (MOVIE) [35], and the DNT-based RR algorithm outlined above, all embed masking mechanisms implemented by some kind of AGC in wavelet or scale-space.

THE DNT IS RELEVANT TO A WIDE VARIETY OF NEUROSCIENCE, PERCEPTUAL, ENGINEERING, AND IN PARTICULAR, QA ISSUES.

Naturally, NR image QA algorithms should also benefit by masking models, either by some type of DNT, or by conditioning on the signal content, or by adapting to the content in some other manner, perhaps through a training procedure. The difficulty arises since masking data is not available from a reference signal.

TOWARD NR IMAGE QA

The NR (blind) QA problem is both tantalizingly important as well as technically difficult. Yet "NR" human judgments of image quality occur with little effort. Our visual systems easily distinguish high-quality against low-quality images, and "know" what is right and wrong about them, without seeing an "original." Moreover, humans tend to agree with each other to a rather high extent. What is the mystery behind perceptual judgments of quality? Do humans have an innate ability to judge the quality of pictures relative to an unseen high standard of quality?

The answer must be positive in the sense of adaptation. Just as people from nontechnological cultures without photography interpret pictures differently from those exposed to it all their lives [36], the customers targeted by purveyors of cable, satellite, and wireless video are quite "picture savvy" with high expectations regarding the picture quality they pay for. These customers, who have observed electronic images all their lives, have collectively adapted to high-quality visual signals and to the distortions that occur. Our neural plasticity extends not only over the eons of evolution (wherein the visual systems are exposed to a large variety of natural scenes), but also over shorter spans within our lifetimes. Short-term plasticity forms the basis for our abilities of visual recognition and

[FIG5] Easy-to-see example of visual masking of white Gaussian noise by high-frequency image content.

visual memory [37], and no doubt, affects our ability to perceive a loss of quality. In other words, there are models of high-quality "reference signals" in our brains, and a learned ability to use these models to assess picture quality. High-definition-equipped readers might try to recall viewing analog TV, and how their satisfaction with respect to this older visual experience has changed.

While we do not know the exact nature of these models, clues are available from prior work on FR and RR QA. We believe that the "prior model," on which the brain relies as it perceives levels of picture quality, must be statistical and reflect the statistics of natural scenes. In this regard, the kind of NSS models we have been discussing are likely well suited for adaptation into theories of visual quality perception, and hence NR QA algorithm design.

Prior work on NR IQA algorithm design has not emphasized statistical image modeling, and most approaches presume that the distortion affecting the visual signal is known. This methods typically estimating image blur (e.g., via edge loss) [38], [39] or JPEG and JPEG2000 compression artifacts by looking for artifact signatures in spatial or spectral domains [40]–[44]. Perceptual modeling remains underutilized, although the authors of [45] utilize a psychometric model derived from subjective tests to create a perceptual blur index, which interestingly attempts to estimate blur relative to image content. Another interesting approach to blur assessment is taken in [46], where a loss of local phase coherence in the complex wavelet domain quantifies departures from image "naturalness." One NSS model-based distortion-specific NR IQA algorithm uses a GSM model to characterize correlations between the wavelet coefficients of images over scales [47]. By measuring reductions in these correlations induced by distortion, good quality prediction performance was demonstrated on JPEG2000 compressed images.

All of the above-described prior work has been geared toward still images only. The field of NR video QA has seen less progress, and almost no work at all using statistical image models. As mentioned earlier, a primary reason for this is a dearth of regular statistical models of naturalistic videos. Many proposed algorithms measure blockiness in compressed videos, e.g., by MPEG-2 [48], [49] or by H.264 [50], [51]. A usual technique is to evaluate edge-strength at block boundaries then relate it to quality. One recent NR QA method for assessing H.264-compressed video quality does use an NSS image model (Laplace/Cauchy) of the transform coefficient distributions, along with a perceptual model of contrast sensitivity and eye movement. They report good performance using their own database of videos and subjective scores [52].

Only a small amount of work has been done on the extremely difficult problem of designing NR QA algorithms that are not fixed to a single type or source of distortion. One interesting approach taken in [53] observes that the statistics of natural images tend to be locally isotropic. The authors hypothesize that image distortions destroy this property, making it possible to detect distortions. They develop an algorithm to measure the degree of local anisotropy across the image using a form of (Renyi) entropy, which is then mapped to quality scores. We implemented and tested this algorithm on the Laboratory for Image and Video Engineering (LIVE) Image Quality database [67], where it achieved a poor correlation score relative to human subjectivity; however, the idea is sound and likely could be improved by an underlying NSS model and additional features. Indeed, inspired by this, we created a simple distortion-agnostic IQA algorithm called Blind Image Integrity Notator Using DCT Statistics (BLIINDS) that uses four simple DCT-domain sample statistics computed from local windows. Inspired by the work in [53], BLIINDS-I (as we will refer it, to distinguish it from a much more evolved version of the general idea) uses two local DCT-domain entropy features and two other simple DCT statistics (kurtosis and contrast) which were to fit to half of the (content-divided) LIVE IQA Database and tested on the other half, using a simple probabilistic prediction model. The method has the virtues of conceptual and computational simplicity and achieved prediction-performance parity with the FR PSNR metric [54]. BLIINDS-I does not rely on a statistical image model, however. Instead, it uses intuitive DCT-domains sample statistics.

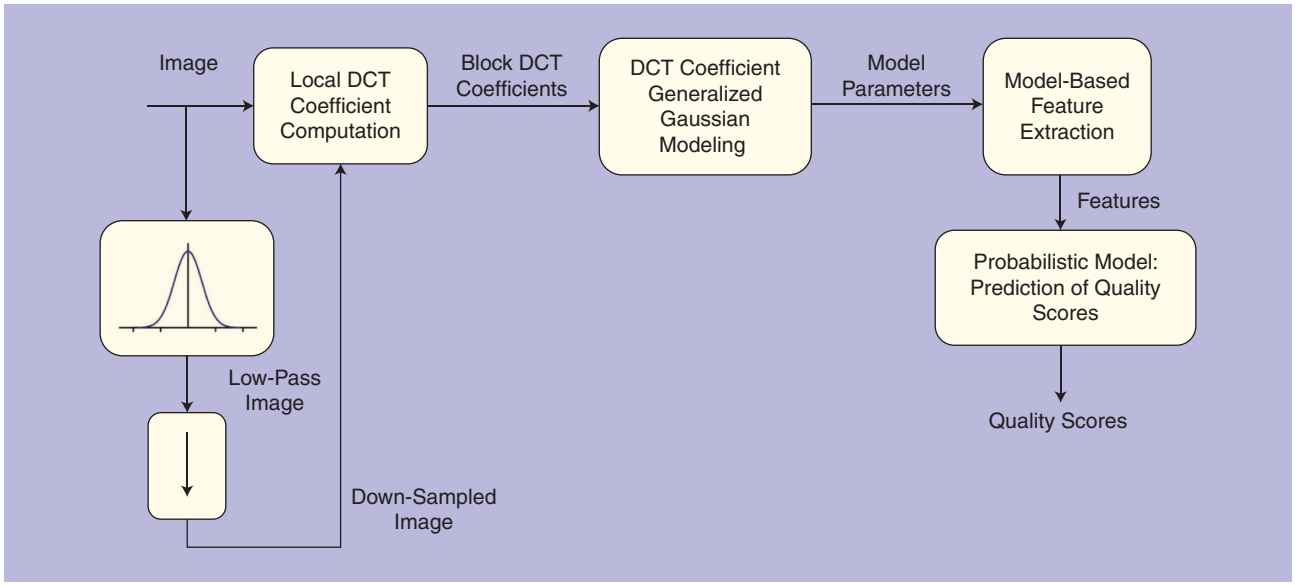
NSS MODEL-BASED APPROACHES TO NR IQA

Motivated by recent developments in NSS-based image modeling and NSS-based RR algorithm design, we have developed new NSS model-based approaches to the NR IQA problem. One exemplary method, named BLIINDS-II, retains the moniker since it still uses easily computed statistics of block DCT coefficients as features, trained on subjective scores [56]. However, BLIINDS-II is quite different from BLIINDS-I: it is model based, and uses very different features that are NSS model based.

Figure 6 diagrams the overall flow of the BLIINDS-II NR IQA algorithm. It is an instructive example, since the features used are simple and naturally defined on the NSS model.

It operates over three scales (performance has been found to remain constant if more scales are added). At the finest scale, nonoverlapping 5×5 image blocks are DCT-transformed and the resulting (non-DC) coefficients used to define statistical model-based features. The coarser scales are obtained by downsampling with a 3×3 Gaussian anti-aliasing kernel. The multiscale DCT basis is used to balance the need to approximate the natural waveletlike multiscale representation of images in the brain with the need for computational efficiency and compatibility with existing DCT-based image processing algorithms. Fortunately, a simple NSS model applies with excellent regularity to the local DCT data.

The BLIINDS-II features are also simply defined. The essential NSS model that is used is the GGD model given in (1) that has been successfully used in RR algorithm design. Specifically, the non-DCT coefficients are modeled as GGD by fitting each



[FIG6] **Flow diagram of BLIINDS-II NR IQA algorithm.**

block DCT histogram with the best-fitting GGD function. Each block is also divided into subblocks (Figure 7(a) and (b), respectively) designed to capture the radial frequency and orientation behavior, and the histogram fit is done on each of these subblocks as well. In this way, the estimated NSS model parameters are used to create all features used in BLIINDS-II.

Only very simple parametric features are extracted from the fit to the GGD NSS model: the GGD shape parameter b (sensitive to distortion signatures); coefficient of variation (CoV) z 5 $\sigma_{|X|}/\mu_{|X|}$ of the magnitudes of the GGD variates X (a normalized energy measure useful for assessing the amount of local image energy, which also accounts for masking); the ratios of energy between the radial frequency bands shown in Figure 7(a)

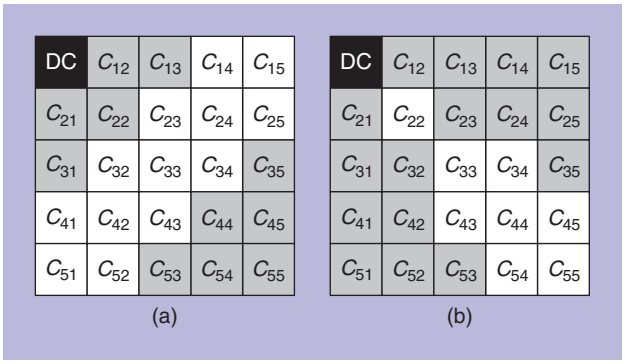
$$R_n = \frac{|E_n - \frac{1}{n-1}\sum_{j<n} E_j|}{E_n + \frac{1}{n-1}\sum_{j<n} E_j}, \tag{3}$$

where the subband energy is the band variance (the ratios indicate the relative proportions of high-, mid-, and low-frequency content in each block). Indexing the subbands outward from DC, two energy ratios are used as features: R_2 and R_3. Finally, two orientation features computed from the DCT block subdivisions shown in Figure 7(b) are used: the standard deviations of the CoV and of the shape parameters: σ_ζ and σ_β, which effectively capture changes in the orientation statistics over the three bands shown (for each scale).

Thus, just a few (six) NSS-based features are extracted using the model fit to the local DCT data. These features are computed over three scales. The features are then pooled in two ways: 1) the block feature values are averaged over the image (standard mean pooling); and (2) only the upper or lower 10% of the block feature values are averaged (depending on the trend of feature against quality). This percentile pooling strategy exploits a previously behavioral observation that the worst distortions in an image affect subjective judgments most strongly [55], [57].

BLIINDS-II is a holistic NR IQA algorithm in the sense that no specific distortion model is used to guide the design of the algorithm, and moreover, it is intended to be useful. In the absence of any distortion model, it is necessary to train NSS model based IQA algorithms using a sufficiently large and diverse database of distorted images with associated subjective data, in the form of mean opinion scores (MOS) or difference of MOS (DMOS) [67], [68].

The method of training and testing is important since any "large-scale" image database still represents a very sparse sampling of the space of all possible images. Training with such a small number of samples (relative to the image space) can lead to over fitting. Thus NR QA algorithms that are designed using train-test procedures should



[FIG7] **Block DCT coefficients divided into bands. (a) Radial frequency bands; (b) orientation bands.**

follow a few simple "rules of conduct." First, there exists a standardized protocol for obtaining the human scores that compose the subjective portion of the database [69]. Second, training and testing should be cross-validated on multiple randomized divisions of the database. We suggest a minimum of 1,000 train-test sequences over which average/median performance and standard error are taken. Lastly, each train-test sequence should randomly divide the database by content, so that content is not learned and used by the algorithm.

THE METHOD OF TRAINING AND TESTING IS IMPORTANT SINCE ANY "LARGE-SCALE" IMAGE DATABASE STILL REPRESENTS A VERY SPARSE SAMPLING OF THE SPACE OF ALL POSSIBLE IMAGES.

BLIINDS-II was trained using the above cross-validation procedure on the LIVE IQA database: 1,000 randomized train-test divisions, using 80% of the content (and distorted versions) for training, and the other 20% for testing. Training was accomplished in a simple manner: the mean and covariance of the algorithm scores and the subjective scores were fit to a multivariate Gaussian distribution to form a probability model. In the test phase, prediction is accomplished by maximizing the conditional likelihood of the subjective scores, given the observed features. Despite the simplicity of the model, the features, and the training method, the performance of BLIINDS-II is remarkably good. Over the 1,000 sequences, the Spearman rank order correlation coefficient (SROCC) and linear correlation coefficient were computed, yielding nearly equal median values of 0.91 against subjectivity, soundly beating the FR PSNR and matching the established performance of the FR SSIM index. Table 1 shows SROCC scores of BLIINDS-II against several leading FR IQA algorithms. Although the performance of BLIINDS-II does not quite match that of the best-performing FR algorithms such as multiscale SSIM (MS-SSIM) or VIF, the level of performance attained by BLIINDS-II is remarkably close to that achieved by the best algorithms that have available the reference image for comparison.

A completely different and specialized approach that can be taken is to attempt distortion identification followed by QA. Such a *two-stage* approach is taken in [58] and [59], where GSM and GGD NSS models in the wavelet-domain are used to create a holistic NR IQA algorithm with very consistent performance comparable to BLIINDS-II. The method is complementary to BLIINDS, since it seeks to determine what distortion(s) afflict an image by computing likelihoods that each distortion is present; these are used to weight multiple distortion-specific QA algorithm scores derived from the same NSS models. When trained using 1,000 iterations of cross-validation on the LIVE IQA database, very good performance is also attained, equivalent to both BLIINDS-II and the SSIM index (Table 1). The reader is referred to [59], since the Distortion Identification-Based Image Verity and Integrity Evaluator (DIIVINE) method is much more involved than the BLIINDS-II index, although it delivers more information regarding the quality of the distorted image. The division of QA tasks in DIIVINE makes it more useful for such important tasks as post-QA distortion reduction, but much less useful for real-time QA applications, as in a video network.

ENVISIONING THE FUTURE

Despite significant recent progress on the very old visual QA problems, there remains significant room for improvement. There is a gap in prediction performance between current performance and what we believe is possible (RR and NR IQA models that predict subject image quality as well as FR algorithms). There is a rather rich literature of NSS models, among which only a small proportion have been successfully exploited in the context of RR and NR QA [5], [18], [28].

A wide spectrum of novel functionalities could be added to RR/NR systems, making them more flexible and versatile in user-centric multimedia communication environments. One desirable feature is rate-scalability, wherein RR features are aligned (and possibly coded) to a continuous bit stream, and ordered according to importance. Such a bit stream could be truncated at any location, and the quality of the distorted images evaluated based on the truncated RR features. Quality prediction could be improved with increased length of the received bit stream. Ideally, such a rate-scalable method could

[TABLE 1] IQA ALGORITHM SCORES AGAINST HUMAN DMOS SCORES FROM LIVE IQA DATABASE [67]. SROCC OVER THE ENTIRE DATABASE FOR FR IQA INDICES PSNR, SS-SSIM, MS-SSIM, AND VIF. MEDIAN SROCC OVER 1,000 RANDOMIZED TRAIN-TEST SEQUENCES FOR NR IQA MODELS BLIINDS-II AND DIIVINE.

IQA ALGORITHM (RN MODELS IN BOLD)	JPEG 2000	JPEG	WHITE NOISE	GAUSSIAN BLUR	FAST FADING NOISE	ALL DATA
PSNR	0.90	0.83	0.99	0.78	0.89	0.87
SS-SSIM [10]	0.94	0.95	0.96	0.91	0.94	0.91
MS-SSIM [70]	0.97	0.96	0.98	0.95	0.94	0.95
VIF [6]	0.97	0.96	0.98	0.97	0.97	0.96
BLIINDS-II [56]	0.95	0.94	0.98	0.94	0.93	0.91
DIIVINE [59]	0.91	0.91	0.98	0.92	0.86	0.92

cover the full range of QA methods (NR, RR, and FR) within a unified framework. It would also be interesting to make the RR approach reverse-directional, where RR features are returned to the sender and compared with the reference features. This would be useful in a networking scenarios (e.g., broadcasting) where central quality control is at the sender side and the RR features from the receiver could help the sender make adaptive adjustments. Even further, one could design bidirectional RR systems, where the RR features could be sent either from the sender to the receiver or vice-versa.

AS PROGRESS CONTINUES, THE NEED FOR REFERENCE-FREE METHODS IS BECOMING MORE PRONOUNCED.

As progress continues, the need for reference-free methods is becoming more pronounced. This will be a very fast-growing area in the next five to ten years, driven by the needs of practical applications and by the many open problems that need to be solved. One of the most important problems to be solved, as we hinted at, is the RR/NR VQA problem, which will require the discovery of comprehensive video NSS models that are statistically regular, and that are sensitive to losses of naturalness induced by distortion. Another important problem is three-dimensional (3-D) stereoscopic image and video QA. Here also, there is a deficit of accurate and regular QA models, and good-performing algorithms (relative to two-dimensional algorithms applied to 3-D data) do not yet exist. We hope that this tutorial article can help attract and inspire more academic researchers and industrial practitioners to this fast-evolving field.

ACKNOWLEDGMENTS

The research of Alan C. Bovik was supported in part by Intel and Cisco Corporation under the VAWN Program and by the U.S. National Science Foundation under the IIS program. Zhou Wang was supported in part by the National Science and Engineering Research Council of Canada and by the Ontario Early Researcher Award program.

AUTHORS

Zhou Wang (zhouwang@ieee.org) is an associate professor in the Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research interests include image processing and multimedia communications. He has more than 90 publications in these fields with over 7,000 citations. He was an associate editor of *IEEE Signal Processing Letters* (2006–2010). He is currently an associate editor of *Pattern Recognition* (2006–present) and *IEEE Transactions on Image Processing* (2009–present). He received the 2009 IEEE Signal Processing Best Paper Award, ICIP 2008 IBM Student Paper Award (as senior author), and 2009 Ontario Early Researcher Award.

Alan C. Bovik (bovik@ece.utexas.edu) is the Curry/Cullen Trust Chair Professor at The University of Texas at Austin. He is the director of LIVE in the Department of Electrical and Computer Engineering and the Institute for Neurosciences. His many awards include the 2011 IS&T Imaging Scientist of the Year Award and the 2009 IEEE Signal Processing Society Best Paper Award. He created the IEEE International Conference on Image Processing and cofounded *IEEE Transactions on Image Processing*. His books, articles on education, and award-winning online courseware and SIVA software attest to his dedication to engineering education.

REFERENCES

[1] Cisco Corporation. (Feb. 2011). Cisco visual networking index: Global mobile data traffic forecast update, 2010–2015. [Online]. Available: http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.pdf

[2] A. C. Bovik, "Meditations on video quality," *IEEE Multimedia Commun. E-Lett.*, vol. 4, no. 4, pp. 4–10, May 2009.

[3] Z. Wang and A. C. Bovik, *Modern Image Quality Assessment.* San Rafael, CA: Morgan & Claypool, 2006.

[4] H. B. Barlow, "Possible principles underlying the transformation of sensory messages," in *Sensory Communications,* W. A. Rosenblith, Ed. Cambridge, MA: MIT Press, 1961, pp. 217–234.

[5] E. P. Simoncelli and B. Olshausen, "Natural image statistics and neural representation," *Annu. Rev. Neurosci.*, vol. 24, pp. 1193–1216, May 2001.

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

[7] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain," *IEEE Trans. Image Processing*, vol. 12, pp. 1338–1351, Nov. 2003.

[8] H. R. Sheikh and A. C. Bovik, "An evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Processing*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.

[9] N. Ponomarenko, M. Carli, V. Lukin, K. Egiazarian, J. Astola, and F. Battisti, "Color image database for evaluation of image quality metrics," in *Proc. Int. Workshop Multimedia Signal Processing*, Australia, Oct. 2008, pp. 403–408.

[10] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Processing*, vol. 13, pp. 600–612, Apr. 2004.

[11] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Trans. Image Processing*, vol. 20, no. 5, pp. 1185–1198, May 2011.

[12] A. A. Webster, C. T. Jones, M. H. Pinson, S. D. Voran, and S. Wolf, "An objective video quality assessment systems based on human perception," *Proc. SPIE*, vol. 1913, pp. 15–26, 1993.

[13] B. Girod, "What's wrong with mean-squared error?" in *Visual Factors of Electronic Image Communications*. Cambridge, MA: MIT Press, 1993.

[14] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? A new look at signal fidelity measures," *IEEE Signal Processing Mag.*, vol. 26, no. 1, pp. 98–117, Jan. 2009.

[15] Z. Wang and E. P. Simoncelli, "Reduced-reference image quality assessment using a wavelet domain natural image statistic model," in *Proc. SPIE Conf. Human Vision Electronic Imaging*, Jan. 2005, vol. 5666, pp. 149–159.

[16] S. Mallat, *A Wavelet Tour of Signal Processing,* 2nd ed. San Diego, CA: Academic, 1999.

[17] 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.

[18] D. J. Field, "What is the goal of sensory coding?" *Neural Comput.*, vol. 6, no. 4, pp. 559–601, 1994.

[19] W. S. Geisler and R. L. Diehl, "Bayesian natural selection and the evolution of perceptual systems," *Phil. Trans. R. Soc. Lond. B*, vol. 357, no. 1420, pp. 419–448, Apr. 2002.

[20] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, July 1989.

[21] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[22] S. Wolf and M. H. Pinson, "Spatial-temporal distortion metric for in-service quality monitoring of any digital video system," *Proc. SPIE*, vol. 3845, pp. 266–277, Sept. 1999.

[23] I. P. Gunawan and M. Ghanbari, "Reduced-reference picture quality estimation by using local harmonic amplitude information," in *Proc. London Communications Symp.*, 2003, pp. 137–140.

[24] M. Carnec, P. Le Callet, and D. Barba, "An image quality assessment method based on perception of structural information," in *Proc. IEEE Int. Conf. Image Processing*, Barcelona, Spain, Sept. 2003, pp. 185–188.

[25] M. Carnec, P. Le Callet, and D. Barba, "Objective quality assessment of color images based on a generic perceptual reduced reference," *Signal Process. Image Commun.*, vol. 23, pp. 239–256, Apr. 2008.

[26] Q. Li and Z. Wang, "Reduced-reference image quality assessment using divisive normalization-based image representation," *IEEE J. Select. Topics Signal Process. (Special Issue on Visual Media Quality Assessment)*, vol. 3, no. 2, pp. 202–211, Apr. 2009.

[27] D. J. Heeger, "Normalization of cell responses in cat striate cortex," *Vis. Neurosci.*, vol. 9, no. 2, pp. 181–198, 1992.

[28] D. L. Ruderman, "The statistics of natural images," *Network: Comput. Neural Syst.*, vol. 5, no. 4, pp. 517–548, 1996.

[29] M. J. Wainwright and E. P. Simoncelli, "Scale mixtures of Gaussians and the statistics of natural images," *Adv. Neural Inform. Process. Syst.*, vol. 12, pp. 855–861, 2000.

[30] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable multi-scale transforms," *IEEE Trans. Inform. Theory*, vol. 38, no. 2, pp. 587–607, Mar. 1992.

[31] J. Foley, "Human luminance pattern mechanisms: Masking experiments require a new model," *J. Opt. Soc. Amer.*, vol. 11, no. 6, pp. 1710–1719, 1994.

[32] A. C. Bovik, "What you see is what you learn," *IEEE Signal Processing Mag.*, vol. 27, no. 5, pp. 117–123, Sept. 2010.

[33] B. Girod, "The information theoretical significance of spatial and temporal masking in video signals," in *Proc. SPIE Conf. Human Vision, Visual Processing, Digital Display*, 1989, vol. 1077, pp. 178–187.

[34] P. C. Teo and D. J. Heeger, "Perceptual image distortion," in *Proc. IEEE Int. Conf. Image Processing*, Austin, TX, Nov. 1994, pp. 982–986.

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

[36] J. M. Kennedy, *A Psychology of Picture Perception*. San Francisco, CA: Jossey-Bass, 1974.

[37] I. van der Linde, U. Rajashekar, A. C. Bovik, and L. K. Cormack, "Visual memory for fixated regions of natural scenes dissociates attraction and recognition," *Perception*, vol. 38, no. 8, pp. 1152–1171, Aug. 2009.

[38] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "A no-reference perceptual blur metric," in *Proc. IEEE Int. Conf. Image Processing,* Rochester, NY, Sept. 2002, pp. 57–60.

[39] X. Zhu and P. Milanfar, "A no-reference sharpness metric sensitive to blur and noise," in *Proc. 1st Int. Workshop Quality of Multimedia Experience*, San Diego, CA, July 2009.

[40] Z. Wang, A. C. Bovik, and B. Evans, "Blind measurement of blocking artifacts in images," in *Proc. IEEE Int. Conf. Image Processing,* Vancouver, BC, Canada, Sept. 2000, pp. 981–984.

[41] S. Liu and A. C. Bovik, "Efficient DCT-domain blind measurement and reduction of blocking artifacts," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 12, pp. 1139–1149, 2002.

[42] Z. Wang, H. R. Sheikh and A. C. Bovik, "No-reference perceptual quality assessment of JPEG compressed images," in *Proc. IEEE Int. Conf. Image Processing.*, Rochester, NY, 2002, pp. 477–480.

[43] L. Meesters and J. Martens, "A single-ended blockiness measure for JPEG-coded images," *Signal Processing*, vol. 82, no. 3, pp. 369–387, 2002.

[44] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "Perceptual blur and ringing metrics: Applications to JPEG2000," *Signal Process. Image Commun.*, vol. 19, pp. 163–172, Feb. 2004.

[45] R. Ferzli and L. J. Karam, "A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB)," *IEEE Trans. Image Processing*, vol. 18, no. 4, pp. 717–728, Apr. 2009.

[46] R. Hassen, Z. Wang, and M. Salama, "No-reference image sharpness assessment based on local phase coherence measurement," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing,* Dallas, TX, Mar. 2010, pp. 2434–2437.

[47] H. R. Sheikh, A. C. Bovik, and L. K. Cormack, "No-reference quality assessment using natural scene statistics: JPEG2000," *IEEE Trans. Image Processing*, vol. 14, no. 11, pp. 1918–1927, Nov. 2005.

[48] K. Tan and M. Ghanbari, "Blockiness detection for MPEG2-coded video," *IEEE Signal Processing Lett.*, vol. 7, no. 8, pp. 213–215, 2000.

[49] T. Vlachos, "Detection of blocking artifacts in compressed video," *Electron. Lett.*, vol. 36, no. 13, pp. 1106–1108, 2000.

[50] M. Ries, O. Nemethova, and M. Rupp, "Motion based reference-free quality estimation for H.264/AVC video streaming," in *Proc. Int. Symp. Wireless Pervasive Computing*, 2007, pp. 355–359.

[51] M. F. Sabir, R. W. Heath, and A. C. Bovik, "Joint source-channel distortion modeling for MPEG-4 video," *IEEE Trans. Image Processing*, vol. 18, no. 1, pp. 90–105, Jan. 2009.

[52] T. Brandao and M. P. Queluz, "No-reference quality assessment of H.264/AVC encoded video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 11, pp. 1437–1447, Nov. 2010.

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

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

[55] M. H. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," *IEEE Trans. Broadcast.*, vol. 50, pp. 312–322, Sept. 2004.

[56] M. A. Saad, A. C. Bovik, and C. Charrier, "DCT statistics model-based blind image quality assessment," submitted for publication.

[57] A. K. Moorthy and A. C. Bovik, "Visual importance pooling for image quality assessment," *IEEE J. Special Topics Signal Processing*, vol. 3, pp. 193–201, Apr. 2009.

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

[59] A. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," submitted for publication.

[60] Z. Wang, G. Wu, H. R. Sheikh, E. P. Simoncelli, E.-H. Yang, and A. C. Bovik, "Quality-aware images," *IEEE Trans. Image Processing*, vol. 15, no. 6, pp. 1680–1689, June 2006.

[61] J. H. Van Hateren and A. Van Der Schaaf, "Independent component filters of natural images compared with simple cells in primary visual cortex," *Proc. R. Soc. Lond. B Biol. Sci.*, vol. 265, no. 1394, pp. 359–366, 1998.

[62] E. P. Simoncelli, "Capturing visual image properties with probabilistic models," in *The Essential Guide to Image Processing,* A. C. Bovik, Ed. San Diego, CA: Academic, 2009.

[63] J. H. van Hateren and D. L. Ruderman, "Independent component analysis of natural image sequences yields spatio-temporal filters similar to simple cells in primary visual cortex," *Proc. R. Soc. Lond. B,* vol. 265, no. 1412, pp. 2315–2320, 1998.

[64] S. Roth and M. J. Black, "On the spatial statistics of optical flow," *Int. J. Comput. Vis.*, vol. 74, pp. 33–50, Aug. 2007.

[65] K. Seshadrinathan and A. C. Bovik, "A structural similarity metric for video based on motion models," in *Proc. IEE Int. Conf Acoustics, Speech, Signal Processing,* Honolulu, HI, Apr. 2007.

[66] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "Study of subjective and objective quality assessment of video," *IEEE Trans. Image Processing*, vol. 19, no. 6, pp. 1427–1441, June 2010.

[67] H. R Sheikh, Z. Wang, L. K. Cormack, and A. C. Bovik. (2005, Sept.). *LIVE Image Quality Database* [Online]. Available: http://live.ece.utexas.edu/research/quality/subjective.htm

[68] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "An evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Processing*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.

[69] International Telecommunication Union. (2003). Methodology for the Subjective Assessment of the Quality for Television Pictures. ITU-R Rec. BT 500-11. [Online]. Available: http://www.dii.unisi.it/~menegaz/DoctoralSchool2004/papers/ITU-R_BT.500-11.pdf

[70] Z. Wang, E. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," in *Proc. Annu. Asilomar Conf. Signals, Systems, Computing, Pacific Grove*, CA, Nov. 2003, pp. 1398–1402.

[SP]