INVITED PAPER

Automatic Prediction of Perceptual Image and Video Quality

This paper discusses the principles and methods of modern algorithms for automatically assessing the perceived quality of visual signals.

By Alan Conrad Bovik, *Fellow IEEE*

ABSTRACT | Finding ways to monitor and control the perceptual quality of digital visual media has become a pressing concern as the volume being transported and viewed continues to increase exponentially. This paper discusses the principles and methods of modern algorithms for automatically predicting the quality of visual signals. By casting the problem as analogous to assessing the efficacy of a visual communication system, it is possible to divide the quality assessment problem into understandable modeling subproblems. Along the way, we will visit models of natural images and videos, of visual perception, and a broad spectrum of applications.

KEYWORDS | Digital photography; image quality; objective quality; video quality; visual perception; wireless spectrum crunch

I. INTRODUCTION

The human appetite for electronic visual content is apparently insatiable. The capability of digital cameras, smartphones, and tablet computers to acquire and display high-resolution images and videos continues to advance rapidly, and consumer demand is increasing just as fast. Indeed, it is estimated that Americans took about 80 billion digital photographs in 2011 [1] and that this number will increase by more than 30% by 2015, with half of digital photographs being taken by mobile devices. The proliferation of captured digital image data presents significant challenges to the consumer regarding how to store, share, assess, and cull digital photos.

Beyond photographs or "still pictures," commercial digital cameras and smartphones now routinely capture standards-compliant digital high-definition (HD) videos. Consumers are finding that choosing and organizing digital videos is an even more onerous task, since they occupy much larger data volumes and require considerable time to review. Dedicated social sites such as Facebook, Youtube, Google+, and Flickr enable an increasingly video-savvy public to acquire, upload, and view copious numbers of pictures and videos of diverse sizes, durations, and levels of quality. A quick visit to any of these websites reveals that the visual content typically suffers from a wide variety of annoying distortions.

Streaming video continues to proliferate as well: stored video-on-demand sites such as Netflix and Hulu already deliver a very large and growing percentage of Internet traffic [2], and live online video, such as video telephony (e.g., Skype) is expanding rapidly as well. As the volume of video traffic continues to grow exponentially, finding ways to deliver good quality content is a focal concern of service providers, carriers, and equipment vendors.

This is particularly true in the wireless realm: fourth-generation/long-term evolution (4G/LTE) wireless networks now enable high-speed mobile web videos, IP telephony, video gaming, mobile HDTV, video conferencing and even mobile 3-D TV. Because of the convenience and freedom afforded by high-performance video-optimized mobile devices, wireless video traffic is exploding.

Indeed, the wireless telecommunication industry is facing a watershed "moment" where foreseeable capacity may soon

Manuscript received May 14, 2012; revised January 27, 2013; accepted March 25, 2013. Date of publication July 26, 2013; date of current version August 16, 2013. This work was supported in part by the U.S. National Science Foundation under Grants IIS 1116656, IIS 0917175, CNS 0854905, CCF 0728748, and CCF 0310969, and also by Intel Corporation and Cisco Corporation under the Video Aware Wireless Networks (VAWN) program.

The author is with the Wireless Networking and Communications Group and the Institute for Neuroscience, The University of Texas at Austin, Austin, TX 78712 USA (e-mail: bovik@ece.utexas.edu).

Digital Object Identifier: 10.1109/JPROC.2013.2257632

fail to meet demand. In 2010, global mobile data traffic increased by nearly 200%—for the third year in a row [3]. The current trend of annual doubling of wireless data traffic is expected to continue, and it is largely being driven by video. Indeed, the Cisco Visual Networking Index reports that video traffic already accounts for more than half of all mobile traffic, and this fraction may exceed 75% by 2015 [3]. Given accelerating sales of tablet computers, which consume more spectrum than smartphones, these trends are likely to continue.

Unfortunately, the glut of video-driven data is already straining capacity as evidenced by sporadic poor mobile data performance in high population centers, sharper limits on data usage imposed by wireless carriers, and extreme measures such as "throttling," whereby a mobile user's data rate is dramatically reduced once a threshold is reached (or at the carrier's whim!). Unless something is done (and soon), the wireless "spectrum crunch" could reduce the attractiveness of the wireless video medium. Shortfalls in capacity could lead to sharper limits on use or lower quality viewing experiences as service providers and carriers cope with an excess of both video supply and video demand.

Increasing capacity and improving bandwidth efficiency are difficult goals. Future massively broadband technologies such as 60-GHz "WirelessHD" may deliver high-volume video streams, but only over very short distances [4]. Femtocells could greatly expand local capacity by maximizing spatial reuse of the wireless spectrum [5]; yet large-scale deployments may be five to ten years away.

While it may be argued that consumer appetite for visual content is already peaking, it is likely that we are seeing the tip of an emerging iceberg. As visual creatures we are drawn to visual realism, naturalness, and, increasingly, immersion. Our behavior and conscious awareness are strongly correlated with vision: more than 30% of cortical neurons are devoted to vision [6]. Multimodal, immersive sensorial experiences are a focus of industry R&D efforts and include augmented reality, haptics [7], and bandwidth-dense 3-D [8], [9]. The rollout of smaller, cheaper, and better mobile digital cameras continues, and larger, thinner, flexible, rollable, and foldable displays are on the way. These developments will magnify the video data "crunch" as users find them attractive and convenient to use.

Given that increasingly knowledgeable users demand better quality image and video acquisition and display, it is highly desirable to be able to automatically and accurately predict visual signal quality as would be perceived and reported by these users. Such predictive capability can be used to monitor image and video traffic, and to improve the perceptual quality of visual signals via "quality-aware" processing, computing, and networking. "Quality assessment" algorithms can be used to improve picture quality, e.g., by perceptually optimizing the process of image or video acquisition, by modifying video transmission rates, by reallocating resources to geographically balance video quality across a network, by postprocessing, or by combining these kinds of "quality aware" ideas.

Being able to accurately and objectively predict visual quality in agreement with humans requires detailed mathematical models of picture signals and their distortions and of how both are perceived. These topics involve modeling image and video statistics, understanding how distortions change these statistics, and predicting how distortions of images and videos are perceived using low- and intermediate-level perceptual models. The goal of the following sections is to supply the reader with an understanding of the essential elements of visual quality assessment models, and to inspire practitioners to apply image quality assessment (IQA) and video quality assessment (VQA) algorithms to solve a broader array of practical problems.

II. A VISUAL COMMUNICATION ANALOGY

A convenient and intuitive approach to conceptualizing the multifaceted visual quality assessment problem is to make an analogy with the classical communication problem. We will make extensive use of Fig. 1, which depicts a number of important concepts, beginning with the lower part of the figure, which depicts the elements and flow of a transmitter–channel–receiver communication system. As the caption explains, in this analogy, the "transmitter" is the world of visible radiant energy and of object surfaces that interact in a physically and statistically lawful manner, projecting through lenses to images that are incident on a sensor. Because of the coherent nature of matter and the physical properties of light, "natural" images that are formed by an optical process obey laws that can be statistically expressed. Section III elaborates on this, but for now the term "natural image" or "natural video" may be construed to mean an optical signal sensed from visible light that has been captured by an ordinary good quality camera equipped with a low-distortion lens under reasonably good conditions. This means that the image is not distorted by noticeable aliasing from poor resolution, over or under exposure from poor lighting, or other impairments arising from poor technique. Further, "naturalness" implies that the signal was not synthetically created by computer graphics techniques, nor had its appearance altered by them.

In Fig. 1, the "channel" is more than just the communication medium. Instead, it encapsulates all phases of image or video capture, processing, and display. The sensing step might use a camera of known characteristics or perhaps might be unknown, e.g., an image from a web search or a video from Youtube. Front–end processing might include source compression and encoding, artifact reduction, or format conversion. Digital communication might be as simple as storage into memory on a camera or transmission over a cable or wireless channel. This might include sophisticated error protection or error concealment protocols. Back–end processing might include decompression, correction of compression or transmission artifacts, or preprocessing for display. The monitor could

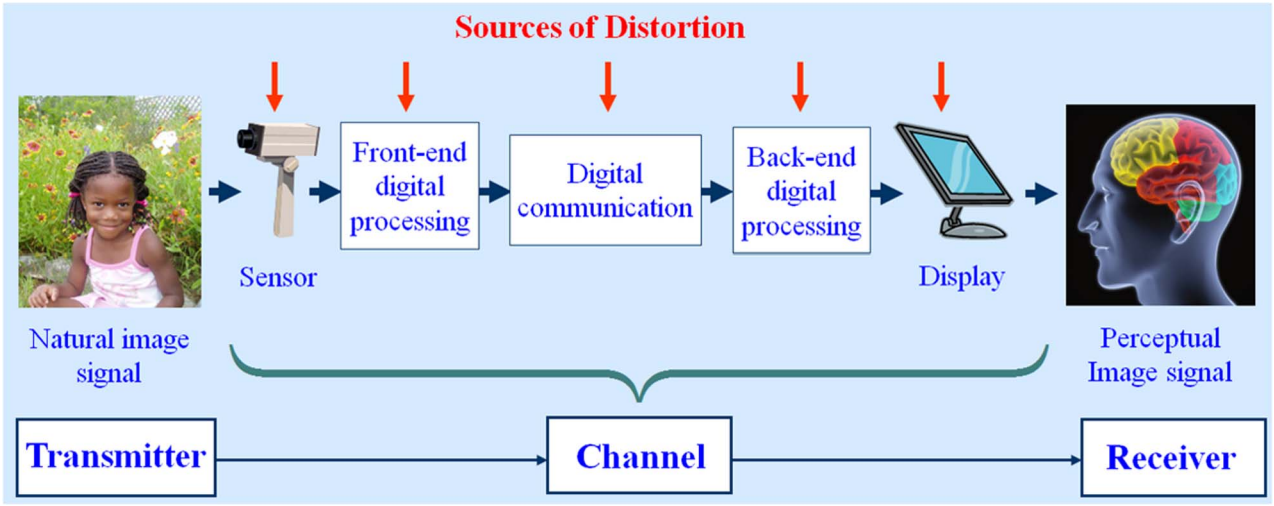


Fig. 1. *Predicting the perceived, subjective quality of a natural image or video that has been artificially acquired, processed, communicated, and displayed is analogous to the classic problem of analyzing the end-to-end efficacy of a visual communication system. In this analogy, the "transmitter" is the physical world that reflects and emits radiation, while the "receiver" is the human visual system. The "channel" is all manipulation of the visual signal beginning with sensing and culminating with display.*

be a small form-factor (but nevertheless high-resolution) smartphone display or a large-format HDTV.

All of the above stages of capture, processing, and display are potential sources of signal distortion, as indicated by the red arrows in Fig. 1. It may be argued that distortion also occurs in the "transmitter," e.g., from light scattering, and in the "receiver," e.g., from imperfect visual optics or neural noise. However, the "channel," as depicted in Fig. 1, defines those points in the flow where the visual signal is ordinarily *digital* and *accessible* by objective visual quality assessment algorithms. This is relevant since, if a "reference signal" is to be used for comparison, then the earliest (and most usual) point at which it can be obtained is immediately post-digitization, i.e., at the sensor. Conversely, the last point at which a "test" image or video may be digitally quality assessed is immediately prior to display. The signal quality can be measured at any or multiple points along the "channel," revealing where quality is most affected.

Of course, any of the above channel substages (aside from capture and display) may be omitted, or others added, depending on what actually occurs between acquisition and viewing of the visual signal. Next, we will start by discussing relevant and commonly used models for the essential transmitter (natural scenes) and receiver [human visual system (HVS)], followed by an overview of distortion models.

III. THE TRANSMITTER MODEL: NATURAL SCENE STATISTICS

The statistics of natural images, commonly referred to as natural scene statistics (NSS), have been studied for more than 50 years by vision scientists and television engineers. The thesis behind NSS models is that photographic images of the world exhibit statistical regularities that reflect the physical world [10]. These regularities manifest in various ways. For examples, natural images exhibit statistical self-similarity or invariance with respect to scale, as exemplified by the "fractal" power law: the magnitude spectra of the spatial Fourier transforms of natural images follow a reciprocal power law [10]–[12]. Here is another example: The principal (and independent) components of natural images closely resemble edge sensitive filters used in computer vision algorithms and by vision scientists to model neuronal responses in visual cortex [13].

One particularly useful NSS model assumes that natural images that have had their lowest spatial frequencies removed obey a Gaussian scale mixture (GSM) probability distribution [14], [15]. If $I(\mathbf{x})$ is an image defined on spatial coordinates $\mathbf{x} = (x, y)$, and $H_\sigma(\mathbf{x})$ is a spatial filter that greatly attenuates low frequencies (such as slowly varying brightness variations from illumination), then the 2-D convolution response

$$J(\mathbf{x}) = H_\sigma(\mathbf{x}) * I(\mathbf{x}) \tag{1}$$

can be reliably modeled as

$$J(\mathbf{x}) = S(\mathbf{x})U(\mathbf{x}) \tag{2}$$

where U is a stationary white Gaussian stochastic process with mean 0 and variance 1. The process S is a scalar *variance field* embodying structured variation and correlation. U could be multivalued or defined over multiple bands or scales;

wavelet-domain GSM models have been successfully used in many image processing applications [16], [17]. In natural images, GSMs that fit the data well are symmetrically distributed with heavier tails than Gaussian [18], reflecting sparse occurrences of large responses to image singularities.

A very simple space-domain GSM model used in [14] and [15] can be used to explain a number of key concepts: let $J(\mathbf{x}) = I(\mathbf{x}) - G_\sigma(\mathbf{x}) * I(\mathbf{x})$, where $G_\sigma(\mathbf{x})$ is a 2-D unit-volume low-pass filter (e.g., a spatially truncated or windowed Gaussian). The image $J(\mathbf{x})$ is a weighted local mean-subtracted "predictive-coded" version of $I(\mathbf{x})$ that is approximately decorrelated. Forming a simple estimate $\hat{S}$ of S, e.g., a weighted sample variance (summed over the support of the filter window)

$$\hat{S}(\mathbf{x}) \sim \sqrt{\sum\sum G_\sigma(\mathbf{x})[J(\mathbf{x}-\mathbf{y})]^2} \tag{3}$$

then executing a contrast normalization step (C is a small constant factor that serves to stabilize the quotient.)

$$\hat{J}(\mathbf{x}) = \frac{J(\mathbf{x})}{[\hat{S}(\mathbf{x}) + C]} \tag{4}$$

yields a residual image signal that is approximately Gaussian. This behavior is broadly observed over natural images. The model (1)–(4) is not perfect and the residual generally retains small spatial dependencies. However, it is close enough and is remarkably regular across natural images [10], [19].

Fig. 2 depicts a natural image I and processed residual image $\hat{J}$ (top left and right, respectively). Also shown are the luminance histograms of both and scatter plots of horizontally adjacent pixels. The unprocessed image exhibits near-linear correlation before normalization (left column) while the residual is nearly Gaussian with a scatter plot that resembles that of white noise (right column).

If the image is subsampled iteratively, as in a wavelet tree, then the GSM model still holds at each scale; the statistics reflected by the GSM model nicely reflect the scale invariance of natural images.

Understanding the statistics of natural dynamic *videos* is a more elusive problem. Much effort has focused on trying to model the statistics of *optical flow* (motions of image luminance) in videos [19]–[22], but this has not been accomplished, except under extremely limiting conditions, e.g., that all flow is assumed due to egomotion (camera movement), viz., without any arising from the motions of independent objects [19], [20]. Thus far, the complexity of object and camera motions has rendered optical flow statistics difficult to model.

However, statistical regularities do exist in natural videos. For example, Dong and Atick [23] found that natural videos reliably obey a (global) space–time spectral model that does not require accounting for optical flow. Further, a simple and regular natural video statistic (NVS) model nicely describes filtered or transformed time-differential (or practically, frame-differenced) videos, without the need for computing optical flow. If $I(\mathbf{x}, t)$ is a natural video defined on 2-D space $\mathbf{x}$ and time t, then the wavelet coefficients or bandpass response to the difference video

$$D(\mathbf{x}, t) = I(\mathbf{x}, t) - I(\mathbf{x}, t-1) \tag{5}$$

will also reliably follow a GSM model [24].

These spatial and temporal scene models supply an incomplete picture of the statistics of the visual world, and much work could be done refining them. Our environment is populated by solid objects that carve out space and that follow motion trajectories, and current NSS/NVS models are far too simple to create realizations of images of these objects or the scenes they reside in. Yet, the perception of common distortions is largely an instantaneous, precognitive, and localized process in space and time. As such, low-order GSM natural scene models for images and videos supply a powerful basis for creating visual quality assessment models that accurately predict human visual responses to distortions.

IV. THE RECEIVER MODEL: HUMAN VISION

Substantial strides have been made toward understanding and modeling low-level visual processing in the human eye-brain system [25]. While high-level cognitive factors (such as semantic content and attention) can affect the perception of quality, distortion sensing (of still images at least) is largely pre-attentive and dominated by low-level processes [26].

Models of neuronal processing that affect quality perception mirror the discussion of natural scene statistics in Section III. Indeed, the architecture of neurons involved in early visual processing is generally regarded as having evolved to efficiently encode and analyze images that obey natural statistical laws [27], [28].

Early processing of the visual signal at the retina has the apparent function of predictive coding. Several types of sensory neurons near the surface of the retina, such as horizontal, amacrine, bipolar, and ganglion cells, collectively accept and sum the inputs from the photoreceptor neurons (cones or rods) to produce a "center-surround" excitatory–inhibitory response to local cone (or rod) cell signals and their surrounding neighbors, yielding a reduced-entropy residual signal. This behavior can be modeled as linear bandpass spatial filtering. Indeed, the collective local neural "impulse response" closely approximates a local difference-of-Gaussian (DoG) low-pass filter: $H_\sigma(\mathbf{x}) = G_{k\sigma}(\mathbf{x}) - G_\sigma(\mathbf{x})$ [29].

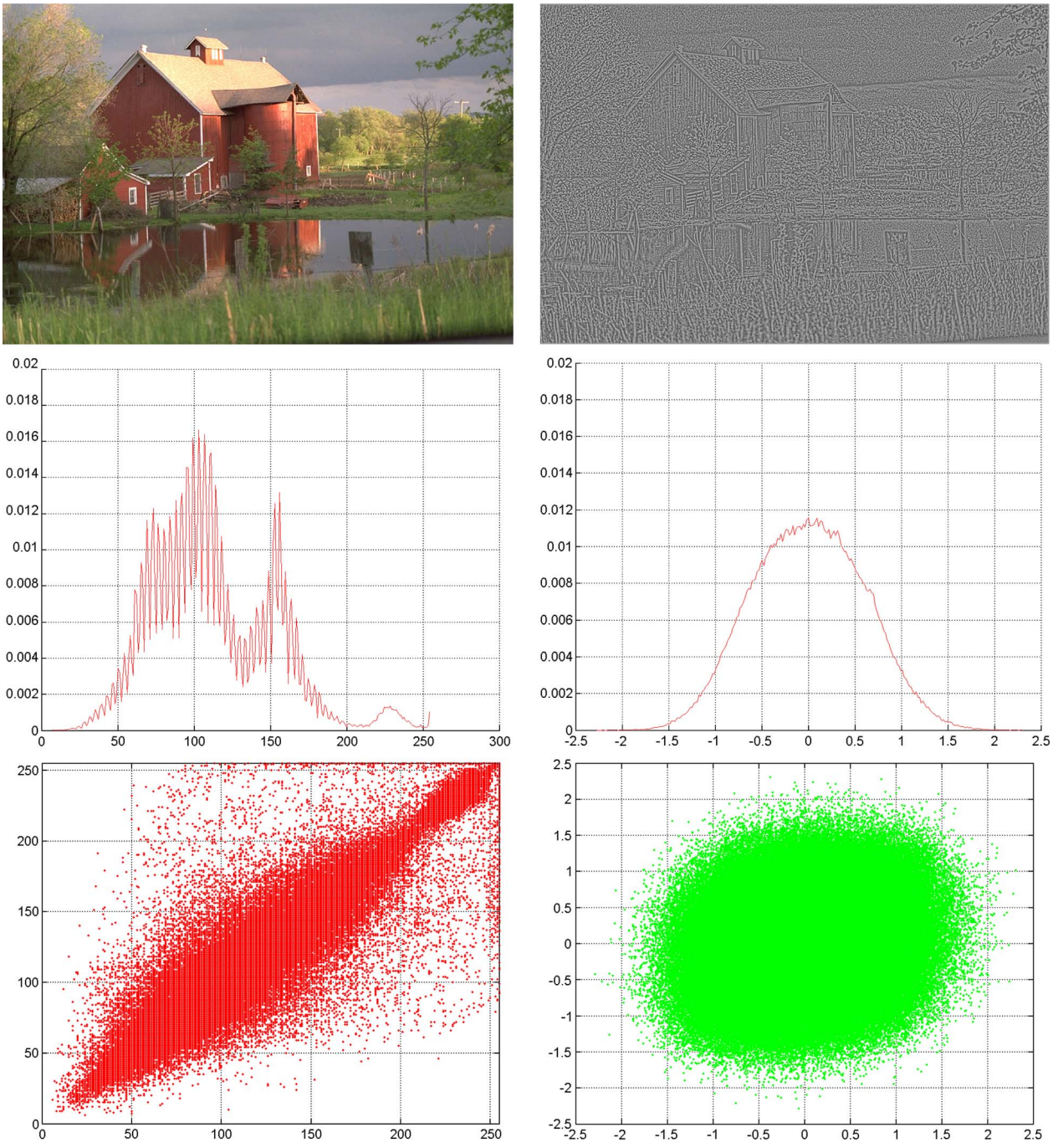


Fig. 2. *Local image debiasing and normalization produces a residual that is a nearly decorrelated and Gaussian distributed. Upper left: A natural image. Upper right: Debiased and divisively normalized residual. Middle left and right: Luminance histograms of original and residual images, respectively. Lower left and right: Scatter plot of horizontally adjacent luminances from original and residual images.*

Fig. 3 depicts this kind of decorrelating spatial summation that occurs at the retina. This occurs over different scales (sizes of the area of summation), orientations, and polarities. The local processing serves a number or purposes, but, in particular, the center-surround differencing accomplishes predictive coding of the retinal signal [30]. This also closely corresponds to the GSM model (1)–(4), and may be viewed as an evolutionary response to the statistics of the visual world.

The visual signal is transmitted from the retinas to the rear of the brain [via the lateral geniculate nucleus (LGN); more on it later in this section] where it arrives at primary

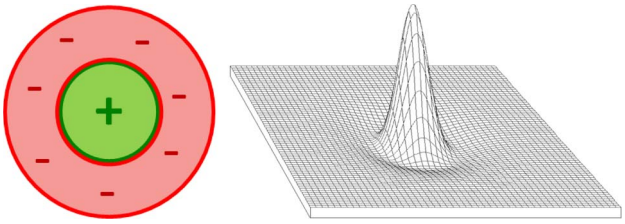


Fig. 3. *Left: Concept of center-surround processing of the visual signal at the retina. This is closely modeled by 2-D spatial linear filters that spatially decorrelate the signal, the basis of predictive coding. Right: A spatial filter model that provides good fits to measured retinal responses are 2-D DoG functions of diverse scales, bandwidths, and orientations. This data decomposition serves many purposes, one of which is to collectively debias and decorrelate the retinal signal.*

(or striate) visual cortex, termed "area V1." Returning to the right-hand side of Fig. 1, area V1 is roughly indicated by the blue area at the back of the brain. Much can be said about the function of the many neurons in V1, and much remains unknown [31]. However, it is clear that the spatial visual signal is decomposed over multiple orientations and scales/frequency bands, in a manner that closely resembles an overcomplete wavelet decomposition of the visual data into narrowband orientation and frequency channels [32]. A great variety of low-level image processing and computer vision algorithms are based on this model [33]–[35]. Cortical processing of visual signals may also be viewed as an evolutionary response to the naturally multiple-scale, multiple-orientation statistical properties of the visual world.

Important aspects of perception that are well modeled and that affect the perception of image quality are *masking principles*. Visual masking occurs when a signal reduces or eliminates the visibility of another signal, typically of similar frequency, orientation, motion, color, or other attribute.

A simple type of luminance masking that occurs is expressed by the Weber–Fechner law, which roughly states that the detectability of a deviation $I + \Delta I$ from a patch luminance I is proportional to the ratio $\Delta I/I$. In other words, a localized image distortion ΔI is more likely visible in a dark image region than a brighter one, largely as a consequence of the logarithmic response of the retinal photoreceptors [36]. This can be used, for example, for image compression [37] or for the design of image noise suppression models [38].

The second and more important type of masking that occurs is a byproduct of the adaptive gain control (AGC) mechanism in visual cortex. As discussed above, so-called "simple" V1 neurons conduct a wavelet-like orientation/scale "transform" of the visuals signals from the two eyes. Computation of the energies of these neural responses is facilitated by the fact that they are commonly found in colocated phase quadrature pairs that feed nonlinear "complex cells" that compute local energy responses to the visual stimulus [39].

AGC is a process of *divisive normalization*, whereby each complex cell's energy response is divided by a weighted sum of those of its neighbors [40], [41], as depicted in Fig. 4. This has the effect of normalizing the response to patterns in the presence of large contrasts, which are typically sparsely distributed, thereby reducing the tail weight of the image distribution, which becomes nearly Gaussian. Indeed, at a fixed scale, this is nicely modeled by (4), including the presence of a "saturation threshold" C. The spatial masking effect plays a central role in nearly every image quality model, and contrast masking models [42] have been used for a long time to perceptually improve such tasks as video compression [43], [44] and image watermarking [45].

Models also describe temporal decorrelation and wavelet-like decompositions along the visual pathway. Temporal decorrelation appears to occur midway between retina and cortex, in the LGN, which is the visual relay station of the thalamus [46]. Information projects from LGN to various brain centers but mostly to primary cortex. Temporal decorrelation, e.g., by linear temporal filtering [46] or by frame differencing (5), allows for greatly reduced data volume, something that was realized early on by video compression engineers, e.g., in the first patent on the topic [47].

Temporal cortical processing of the dynamic visual signal amounts to a space–time multiple-orientation (in

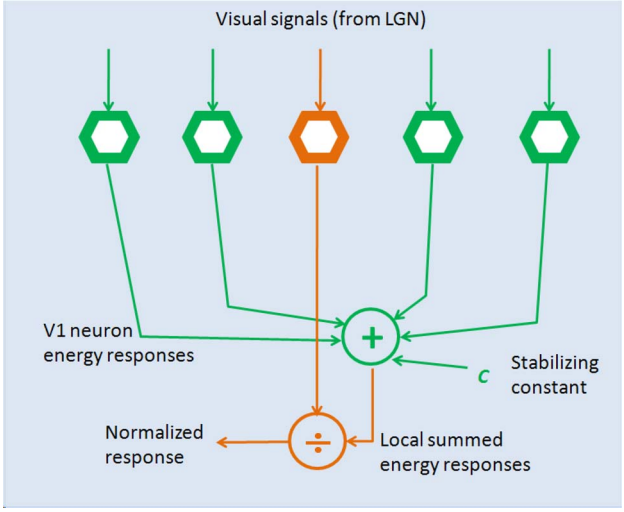


Fig. 4. *Illustration of divisive normalization of the response of a V1 "complex" cell response by the summed energies of neighboring cells that receive signals from similar retinal locations and that are tuned to similar frequencies and orientations. Complex cells compute envelope or energy responses from neighboring pairs of "simple cell" neurons arranged in phase quadrature. The responses of the other complex cells are similarly normalized.*

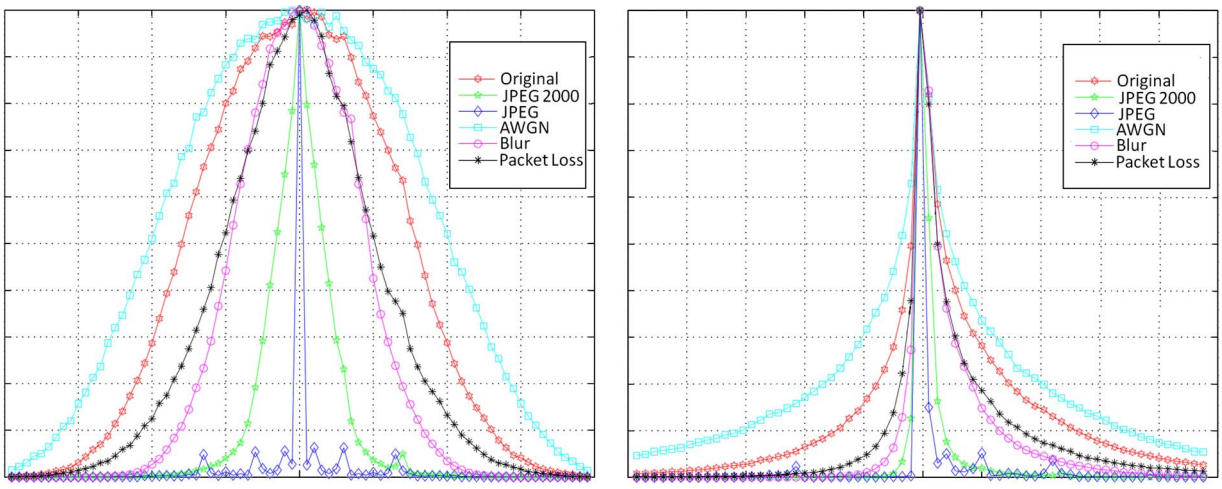


Fig. 5. ***The histograms of normalized distorted images are predictably modified. Left: Histograms of the natural image in Fig. 2 after being distorted in five different ways. Right: Histograms of products of adjacent pixels from the same distorted images.***

space time), multiple-frequency (in space) decomposition [48] that is approximately space–time separable [49]. These low-level spatio–temporal feature coefficients are transmitted to various brain centers, but notably, an entire retinotopic map is sent (i.e., *en masse*, representing the entire visual field) to middle temporal (MT) visual area V5, which is implicated in the formation of coherent motion representations over large spatial areas. Good models of visual processing in area MT exist [50], [51] and have been applied to the VQA problem [52].

While accounting for motion statistics in videos has proved difficult, motion percepts are relevant to understanding the visibility of temporal distortions. The development of temporal masking models is of high interest. Although temporal masking models have been proposed that simply mirror spatial models (e.g., using local energy normalization [53], [52]), the actual picture appears to be somewhat different.

Very recently, an observed visual "motion silencing" phenomenon has been described by Suchow and Alvarez and demonstrated through a series of remarkable visual illusions [54]. They have shown that the presence of large coherent object motions in a video renders local changes in luminance, hue, shape, or size of the objects invisible. Objects that are set in collective appear to stop changing [54]. This dramatic form of dynamic "change blindness" has exciting implications for understanding temporal distortion perception. While there is not yet any definitive explanation of this remarkable effect [55], our own studies suggest that motion coherency, foveation, and spatial and temporal "crowding" all play a role. We have developed a spatio–temporal filter model that accurately predicts when human subjects will judge that silencing has occurred, as a function of average object change rate and object velocity [56].

V. THE CHANNEL MODEL: ARTIFICIAL DISTORTIONS

Referring again to Fig. 1, next we will consider distortions that arise from sources between the "natural transmitter" and the "natural receiver."

Within the digital processing flow indicated by the red arrows in Fig. 1, digital images are subject to a wide variety of distortions, including, but hardly limited to, blur, noise, compression, blocking, false contours, ringing, overexposure/underexposure, quantization, and under sampling. Videos suffer from these spatial distortions and from additional temporal ones: ghosting, mosquitoing, texture flutter, jerkiness, motion estimation errors, and many others.

Models of specific distortion abound, e.g., for estimating the severity of JPEG blocking [57]–[61], ringing from JPEG2000 [62], blur [63]–[67] (e.g., via edge loss or the perceptually relevant "just-noticeable blur"), combinations of noise, blur and blocking [68], and MPEG-4 source/channel distortions [69]. Most of these focus on making direct measurements of the artifacts introduced, e.g., by analyzing spatial image structure or loss of structure, but without using a model of the underlying visual signal, or of the human receiver (with some exceptions, e.g., [65] and [66]).

However, it has been observed that the statistics of natural scenes are predictably modified by distortion [70], making it possible to determine the presence and severity of distortions without the need for specific distortion models. For example, Fig. 5 (left) plots the empirical histograms of 80 different natural images, each distorted by four common processes (JPEG, JPEG2000, white noise, and blur), then normalized via (5); each distortion characteristically affects the image distribution.

The two-parameter generalized Gaussian distribution (GGD)

$$f_1(a) = A_1 \exp\left(-\left|\frac{a}{\sigma}\right|^{\gamma}\right) \tag{6}$$

provides a good fit to the empirical histograms of distorted (and undistorted, when $\gamma \approx 2$ and $\sigma \approx 1$) images [41], [71], [72]. Estimates of the shape and scale parameters γ and σ can be used for identifying and/or assessing distortions, as we will see later (Section VIII).

Moreover, distortions typically introduce unnatural spatial dependencies, which can be measured by examining the distributions of local image correlations (products of adjacent pixels) following normalization (4). Letting $K(\mathbf{x}) = \hat{I}(\mathbf{x})\hat{I}(\mathbf{x} \pm \mathbf{1})$ denote the product between a normalized pixel and any of its eight neighbors (along cardinal or diagonal directions), then the asymmetric generalized Gaussian distribution (AGGD) model [73]

$$f_2(a) = \begin{cases} A_2 \exp\left(-\left|\dfrac{a}{\sigma_L}\right|^{\lambda}\right) \\ A_2 \exp\left(-\left|\dfrac{a}{\sigma_R}\right|^{\lambda}\right) \end{cases} \tag{7}$$

captures the shapes and distortion-driven asymmetries in the products $K(\mathbf{x})$ (in the absence of distortion, the distribution of $K(\mathbf{x})$ is symmetric [74]). Fig. 5 (right) depicts histograms of normalized distorted images showing distinct distortion signatures: spreads, degree of skew, and asymmetry. These are highly reliable indicators of distortions [75].

The model parameters (γ, σ) in (6) and (λ, σ_L, and σ_R as well as the mean μ) are distorted scene statistics (DSS) or conversely, quality-aware features that can be used to create IQA models. These kinds of features play an important role in image quality models that use little or no reference information, as described in Sections VII and VIII. These kinds of features are also important in VQA models, when applied on a frame-difference basis [24].

VI. FULL REFERENCE VISUAL QUALITY ASSESSMENT AND APPLICATIONS

It is usual to divide image and video quality assessment models into three broad categories: full reference (FR), reduced reference (RR), and no reference (NR) or "blind." There exist a number of broad surveys [76]–[79], books [80], [81], and publicly available comparative studies [82]–[84] already. Source code for many IQA and VQA models is available at [85] and [86].

As such, we will not attempt a review, but rather discuss some representative high-performing models and how the underlying principles outlined in the preceding sections guides their function. Most existing IQA and VQA models can be regarded as utilizing some combination of elements of the transmitter, receiver, and channel models that we have been discussing.

We will also examine how I/VQA algorithms can be used in quality-driven applications of image and video processing, transmission, and analysis, and will follow this pattern also when discussing RR and NR algorithms in ensuing sections.

An FR IQA or VQA index assumes that a pristine signal is available to compare distorted versions of the signal against. While a reference is extremely useful, this also, of course, limits the applicability of these models in many applications. The most commonly used metric remains the mean squared error (MSE) or its equivalent, the peak-signal-to-noise ratio (PSNR). However, the MSE does not predict subjective judgments of visual quality well, despite its computational and analytic convenience [87]. Thus, perceptually relevant methods are rapidly gaining popularity.

Certainly the most successful *perceptual* FR IQA algorithm is the SSIM index [88], which can be made very fast [89], and which delivers highly competitive image quality predictions against human judgments, particularly in multiscale implementation (MS–SSIM) [90]. SSIM is defined as a product of three terms computed over small image patches: a "structural similarity" term $s(\mathbf{x})$, a luminance similarity term $l(\mathbf{x})$, and a contrast similarity term $c(\mathbf{x})$. The "structure" term measures the linear correlation similarity between corresponding reference and distorted image patches. The second term $l(\mathbf{x})$ has two virtues: it compares the luminance similarity between corresponding reference and distorted image patches, but also is a vanishing function of the ratio $\Delta\mu_I/\mu_I$, where μ_I is the mean luminance of the pristine patch and $\Delta\mu_I$ is the deviation from it in the distorted image. Thus, the SSIM index embodies a Weber–Fechner principle. Likewise, the contrast similarity term measures similarity between patch contrasts, but also, for a fixed contrast difference, $c(\mathbf{x})$ diminishes with the reference contrast. Thus, the SSIM index also has a contrast masking behavior. The MS–SSIM index thus contains all of the basic spatial perceptual principles discussed in Section IV.

However, the SSIM index can also be interpreted using natural scene models [91]. Indeed, under the GSM model, the SSIM index equates closely with another top-performing IQA index called visual information fidelity (VIF) [92], if the luminance similarity term is omitted. Some other IQA algorithms also deliver good performance: the visual-signal-to-noise ratio (VSNR), which uses multi-scale modeling of contrast masking to detect distortion visibility and subsequently assesses contrast degradation [93], the most apparent distortion (MAD) model [94], which uses a similar strategy to assess quality differently

depending on whether distortions are severe or moderate, and FSIM [95], which deploys comparisons of phase coherency, an idea first proposed in [96]. Variations of SSIM are also abundant, e.g., by weighting SSIM or MS–SSIM scores by saliency [26], by the type of image content [97], by local information measures [98], or by applying perceptual weights to the three SSIM terms [99].

SSIM models remain quite competitive with these variations on FR IQA, and are much simpler to implement than most. Because of this, SSIM has become a *de facto* choice for IQA applications. FR IQA models like SSIM can be used to benchmark image processing algorithms, since in simulation of algorithm performance, a reference image is usually available. The SSIM index is used to validate and compare the results of all kinds of image processing algorithms (far too many to list). A good practical example is the inclusion of SSIM in the H.264 video compression standard JM reference software [100], where it can be used to compare the before-and-after quality of video compression. Many examples of SSIM-driven algorithm benchmarking are given in [87] and [101].

A very exciting direction is the idea of *perceptual optimization* of image processing algorithms using SSIM or other IQA indices. SSIM was first used to perceptually optimize image restoration, by showing that the optimization can be cast as quasi-convex [102]. By expressing SSIM using a discrete cosine transform (DCT) formulation, it was shown possible to derive rate bounds on image DCT quantization under SSIM [103]. Since then, the SSIM index has been used to optimize multichannel image restoration [104], image histogram shaping [105], image denoising [106], [107], perceptually optimized image compression [108] including JPEG 2000 [109], and for estimating realistic compressed video distortions using variations of SSIM [110]. Recently, Brunet *et al.* [111] developed a number of interesting metric properties of SSIM useful for perceptually optimizing image processing problems in a natural and rigorous manner.

One method of "live" optimization that is of particular interest is perceptual rate control of compressed video. In the H.264 standard, the method for rate control is not dictated, and so there is considerable room for design control. A straightforward method of perceptual rate-distortion (R–D) optimization is to seek to minimize the cost functional

$$C = R + \lambda(1 - \mathrm{SSIM}) \tag{8}$$

where λ is a Lagrange multiplier that mediates a tradeoff between bit rate R and distortion 1-SSIM. Given a target bitrate, the perceptually relevant SSIM index is used to estimate the R–D curve in the interframe predictor. Huang *et al.* [112] and Ou *et al.* [113] solve this problem using a simple exponential model of distortion as a function of rate, achieving as much as a 25% improvement over the H.264 JM reference software recommendation.

The astute reader may have noticed that the SSIM index as used in these video applications is not a true VQA model, in the sense that it does not embody temporal models of either video statistics or perception. While extensions of SSIM to video have been shown to perform well [22], [114], the best performers on publicly available video quality databases are the ST–MAD [115] and MOVIE [52] indices, both of which are based on temporal perceptual designs. ST–MAD builds on the still image MAD model described earlier, while MOVIE uses a model of extra-cortical area MT motion perception. However, these algorithms require computing motion vectors, which is a considerable overhead for real-time applications such as rate control. The older ITU standard algorithm VQM [116] is fast, but only delivers performance similar to the spatial-only MS–SSIM index [84]. Other algorithms using perceptual criteria for FR VQA are those in [117] and [118] ("TetraVQM"); both model QA performance using perceptual factors based on motion computations, and deliver QA prediction performance comparable to VQM [116].

Since these algorithms require motion computations, real-time implementations are probably in the future. However, high-performing VQA algorithms such as those outlined above are ideally suited for benchmarking video processing algorithms and codecs.

VII. REDUCED REFERENCE VISUAL QUALITY ASSESSMENT AND APPLICATIONS

In real-world image quality monitoring applications, the requirement of a reference image or video signal is often problematic. This might be because no high-quality reference signal is available at all, in which case IQA/VQA must be done using an NR model, which is the topic of Section VIII.

However, if a reference signal is available, but is too costly to supply *in toto* to the location where quality assessment is to be accomplished, then reduced-reference (RR) approaches that transmit only a very small fraction of the reference information are of great interest.

For brevity, we will not consider distortion-specific RR algorithms here. The most prominent general-purpose RR IQA algorithms are based on NSS models. For example, the "quality-aware image" RR IQA method [119] deploys a GGD model (6) of the wavelet coefficients of images, and embeds NSS-based quality-aware features into the image via a watermarking method. Thus, no side channel is needed. Likewise, the "divisive normalization" RR IQA model in [120] operates under the GSM model (1)–(5) on the wavelet coefficients [121] of the reference and distorted images. Following a divisive normalization stage similar to (5), the histograms of wavelet coefficients of the original and distorted images are fit to parametric

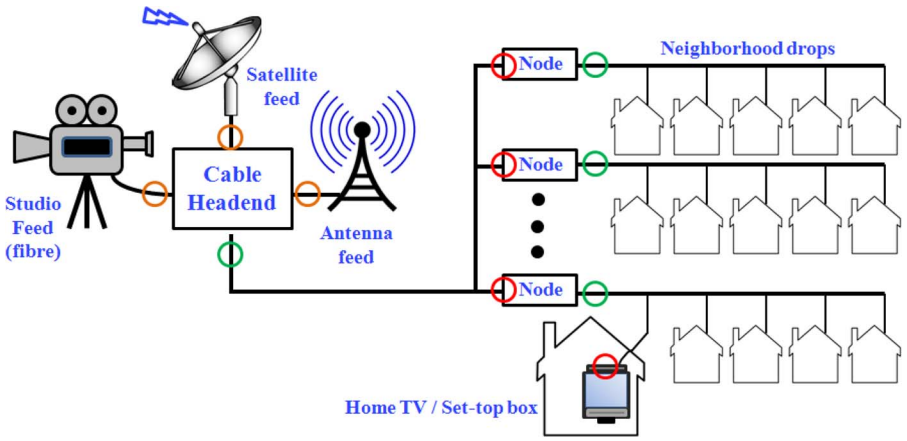


Fig. 6. *Diagram of cable television broadcast system and points at which reference and "test" videos may be defined. Quality may be compared between any two points along the path from signal source to set-top box. However, the green circles ◯ indicate points where reference videos may be defined for comparison with the signal further along the path. Possible points of testing the quality relative to reference are indicated by red circles ◯. Another possibility is that the cable provider may test the videos arriving from content providers using an NR video quality assessment algorithm, indicated by orange circles ◯.*

distributions and compared using a mutual information measure. More recently, the RRED indices [122] are a family of RR IQA models that also use the GSM model. Rather than employing divisive normalization, these models compute wavelet-domain entropies conditioned on the variance field. IQA is then based on comparing the conditional entropies of reference and distorted images.

It is difficult to closely contrast the performance of RR IQA models, since the amount of "side" information that may be sent varies. Generally, the quality prediction power they offer is somewhat better than that of PSNR when very little information is sent (e.g., the "single number" version of RRED [122]), but close to the performance of the best FR models when a lot is sent (still a very small percentage of that sent by an FR algorithm).

Models for general-purpose RR *video* QA are few in number. This owes in part to the slow development of video statistics models and motion masking models. Nevertheless, the need for algorithms of this type is quite significant. A good example is the cable television quality monitoring problem depicted in Fig. 6. Cable viewers have constantly increasing expectations of larger, higher quality viewing experiences in the "home theater," so cable providers expend considerable effort in testing and maintaining video quality. However, highly visible video impairments are common. Several companies offer products for accomplishing point-to-point video quality testing, some using SSIM and others using proprietary FR, RR, or NR algorithms. The latter generally test for technical error conditions rather than using perceptual or video source models, so there is considerable room for improvement.

In this direction, the RRED concept has been extended to the RR VQA problem, by applying the GSM model (1)–(3) to video frame differences, computing wavelet-domain conditional entropies on these differences, then comparing the entropic differences between reference and distorted videos [24]. In direct comparisons with top-performing FR VQA algorithms such as VQM, MOVIE, and ST-MAD, the so-called spatio–temporal RRED (ST–RRED) indices perform quite competitively, at much lower cost, since no motion computations are required. Source code is freely available at [86].

VIII. NO-REFERENCE VISUAL QUALITY ASSESSMENT AND APPLICATIONS

If no image or video reference information is available, then IQA/VQA must be conducted without it. This so-called NR I/VQA or "blind" problem is the most interesting and potentially the most important practical QA problem. The numerous applications that are appearing in the video-intensive handheld and mobile landscape do not allow the possibility of reference data. Further, the concept of "reference signal" is very difficult to define; any signal suffers from distortion. A broad review of the many challenges and possible sources of information embedded in the NR QA problem is given in [123].

Many algorithms have appeared that seek to blindly assess a single distortion or combination of distortions. These are too many and diverse to comprehensively survey, but they include methods to blindly assess blur [63]–[67], blocking (e.g., from JPEG) [57]–[61], [126], ringing (e.g., from JPEG2000) [62], [127], combinations of these [68], [124], [125], and MPEG/H.264 distortions [69], [128], [129]. The idea of identifying and distinguishing

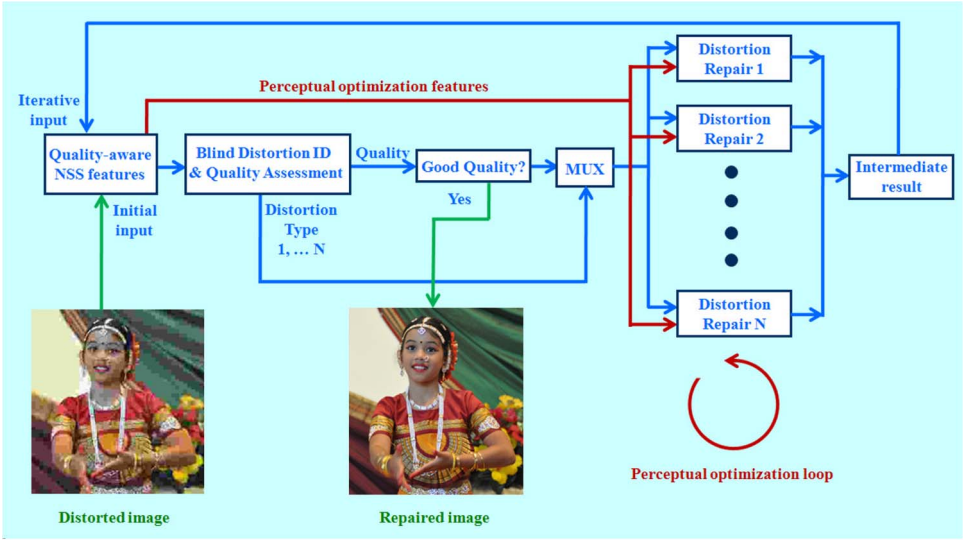


Fig. 7. ***General-purpose image repair using quality-aware NSS features to identify the dominant distortion type afflicting an image and assess the perceptual quality level resulting from it. Once the distortion is identified, an off-the-shelf image repair (denoising, deblurring, deblocking, etc.) algorithm is applied to reverse the distortion. Importantly, an inner perceptual optimization loop selects the repair algorithm parameters to deliver the highest quality image. Once repaired, the image being processed returns to the outer loop, and the process iterates until a good enough quality is reached, or another stopping criteria is met.***

multiple distortions was introduced in [70] and of quality assessing them in [130].

Of broader interest are blind IQA models that, like SSIM, are agnostic to distortion type. While the idea is not new [131], general-purpose NR IQA models that provide competitive performance have only recently begun to appear within the last two years [72], [73], [75], [130], [132]–[137]. These methods usually deploy some form of training, clustering, or other kind of machine learning principle, since the mapping from specific distortions to perception is poorly modeled. Through the use of suitable "quality-aware" NSS features in the wavelet domain [72], [130], [135], the DCT domain [132], [133] (generalizing the Laplacian model of DCT statistics [134]), or the spatial domain [73], [75], or using features that reflect NSS, such as image edges [136], or that map perceptual features [137], algorithms can be created that learn human responses to distortion by training them on large databases of human opinion scores [82], [138]. These algorithms achieve performance that is comparable to the best FR and RR IQA algorithms, with the caveat that their range of application is limited to the distortion types they have been trained on.

The NR IQA Blind/Referenceless Image Spatial QUality Evaluator (BRISQUE) index [73], [75] is simple to understand and apply, particularly since we have done most of the work defining it already. Utilizing the model of point NSS statistics (6), which has two feature parameters γ and σ, and the point-product NSS model (7), which has four features (λ, σ_L, σ_R, and μ) measured along four directions (the two cardinal and the two diagonal) results in 18 features; these are measured over two scales in BRISQUE yielding a total of only 36 features extracted from each image, all of which can be extracted at relatively little cost. Like other general-purpose NR models, BRISQUE is trained on large databases of human opinion scores, expressed as either mean opinion scores (MOS), or a variation, difference MOS (DMOS), which is a method of debiasing opinions from content (e.g., an image of something attractive should not rate higher than one of something unappealing, if they are distorted at the same perceptual level). In [73], a machine learning engine known as support vector regressor (SVR) is used to train the BRISQUE index. In application, BRISQUE computes the same 36 features from the image to be quality assessed (an efficient process), these are fed to the SVR, and a quality score is produced. Currently, BRISQUE delivers the highest level of predictive IQA performance among general-purpose NR IQA models on the LIVE database of distorted images [82], [138], while also offering computational efficiency.

However, it is possible to achieve nearly the same level of quality predictive performance using the same perceptual features, without training on human opinions of distorted images [139] or without exposure to any kind of distortion at all [140]. This is done by comparing the empirical distribution of each image to be quality assessed against that of a representative (and sufficiently large) corpus of high-quality images. The resulting "completely blind" natural image quality evaluator (NIQE) model is thus really a measure of "image naturalness" [140].

General-purpose NR VQA algorithms are only very recently being developed. One promising model, termed ST–BLIINDS [141], extends a DCT-domain NR IQA index

to the temporal domain by training an SVR on an NVS model of frame differences [24], and by weighting the features using a measure of spatial coherency inspired by the Suchow–Alvarez phenomenon. ST–BLIINDS is competitive with FR VQA algorithms, although ostensibly limited to the video distortions in the database it is trained on.

A particularly exciting application of NR IQA (and eventually NR VQA) models is the possibility of applying quality-aware features to conduct automatic picture repair and capture [142]. This requires that the concept of NR IQA be extended to a two-stage process, whereby the dominant distortion type in the image is identified by the QA algorithm. This two-stage distortion identification followed by quality-assessment model was first proposed in [130] and fully developed in the two-stage DIIVINE index [72]. The BRISQUE model we have been describing also can be implemented in this manner with high efficiency. Fig. 7 depicts the overall concept [138]. Two perceptual optimization loops occur: in the outer loop, the image is processed to determine the dominant distortion (from among those the system is trained on) in the image, and the perceptual quality assessed. If the quality is inadequate, the distorted image is passed through a multiplexer that guides it to an appropriate state-of-the-art image repair engine, to conduct denoising, deblurring, deblocking, etc. The parameters of the repair algorithm are found using the quality-aware features in an inner *perceptual optimization loop*. Once repaired, processing returns to the outer distortion identification/quality assessment loop. Fig. 8 shows an example of this iterative process.

This broad framework for image repair is quite different from any existing method. It simultaneously offers the opportunity for achieving new levels of performance in perceptual image repair, but also presents new challenges regarding how to differentiate distortion type and how to determine the convergence of a particular implementation of the model. Of course, the simplest way to ensure convergence is to limit the number of iterations, and output the maximum quality image created across iterations.

One immediate application of this general paradigm, modified for acquisition rather than postprocessing, is digital camera control, wherein control of the camera parameters, such as ISO, aperture, exposure, etc., are perceptually optimized before the image is "snapped." In this application, quality-aware features would be used as above to perceptually optimize each setting via an inner perceptual optimization loop while an outer loop determines which (if any) setting needs to optimized next.

IX. THE FUTURE OF VISUAL QUALITY ASSESSMENT

At this point, we have begun to reach the "outer limits" of image quality research. Yet, there are a few areas of pressing interest not touched on here, since work on these topics remains in an early stage.

Color quality is an important consideration, yet there are not currently any well-accepted models of perceptual quality prediction of color images. However, pretty good results and some improvement relative to "luminance-only" processing can be obtained by applying standard single-channel QA models to one or more chromatic channels, then combining these in various ways [143]–[146]. However, progress remains to be made on this problem, given the complexities of color perception, including opponency [147], [148], and the lack of current models of color distortion perception.

Assessing the quality of stereoscopic (3-D) images is also a topic of pressing interest. The main problem is geometry and visual comfort, which is difficult to ensure without a large Hollywood budget (and even then!).

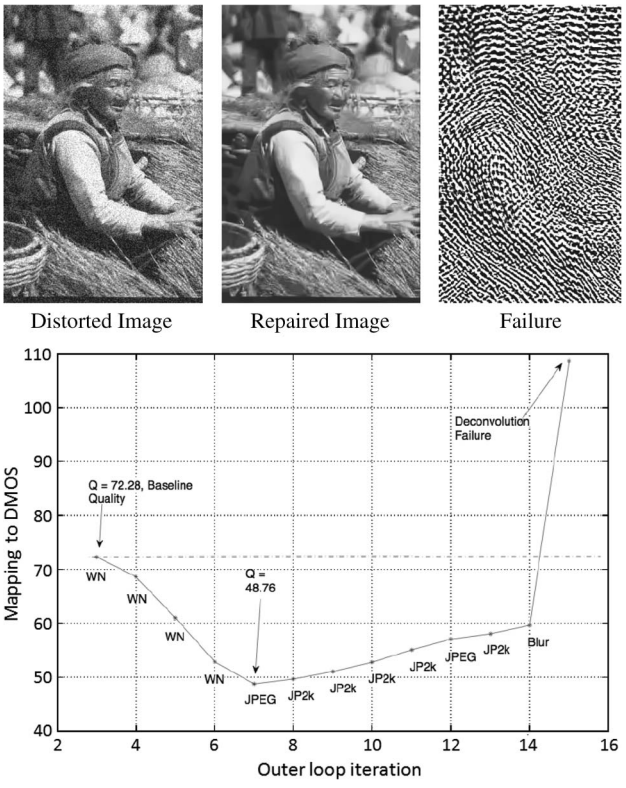


Fig. 8. *Example of perceptually optimized general-purpose image repair. The distorted image is processed by a two-stage NR IQA index to determine the dominant distortion type and quality level at each iteration of the outer image repair loop. The lower plot shows the outer loop iterations and the dominant distortion identified at each iteration from among four possibilities: additive white noise (WN), JPEG compression, JPEG2K compression, or Gaussian blur. At each iteration, the parameters of the repair algorithm are perceptually optimized using quality-aware features. The vertical axis is quality expressed in terms of human opinion (DMOS) learned from the LIVE image quality database, where lower scores indicate higher quality.*

However, our topic here is the perception of distortions, and the role of depth on distortion perception remains quite murky, and like color, no entirely successful method of stereo image quality assessment has been found. While several approaches have been proposed for 3-D stereoscopic QA [149]–[152], 2-D quality models applied to stereopairs commonly perform as well as, or better than, 3-D QA models that utilize depth or computed disparity maps [153]. Generally, the development of effective 3-D QA models has been hindered by poor definitions of stereoscopic distortions and 3-D distortion perception. Moorthy *et al.* [153] observe this and obtain promising results by carefully modeling several 3-D perceptual processes [154].

The statistics of natural chromatic images and of 3-D depth or disparity images are worth deeply exploring for the same reasons as monochromatic 2-D natural image and video statistics. The statistics of color and depth appear to be quite regular [155]–[157], are likely to be implicated in the function of the color and depth senses, and are likely relevant to chromatic and 3-D image quality.

Last, an exciting direction of inquiry is the interaction between visual quality and visual *task*. Certainly, quality plays a role that should be defined within the context of the visual task that is being conducted, and specifically, with regards to how measured quality affects execution of the task. Of course, the main visual task is viewing images or videos for information or entertainment [158], which is the context we have been discussing.

Hemami *et al.* take a broad view of quality versus visual task [159]–[161]. Recognizing the importance of perceptual principles in both visual tasks and in quality assessment, the authors study human and machine recognition of objects as a function of objective image quality as measured by the MS–SSIM and VIF IQA indices. They find that perception-driven FR IQA indices can successfully predict image recognizability. Likewise, Bedagkar-Gala and Shah [162] find that SSIM can be used to predict the performance of tracking algorithms with a high degree of confidence.

A small body of work exists on how quality affects biometric tasks (human recognition from iris, face, or fingerprint) [163]–[169]. The area is rather new and key ideas are ill-defined; e.g., *recognizability* is often used interchangably with *quality*. For example, ISO/IEC 19794-5 [170] specifies a list of factors (spectacles, pose, expression, head shape, and so on) affecting "face quality." While these do affect detection and recognition, they are not aspects of visual *quality* as normally defined (e.g., a high-quality image may be taken of a smudged fingerprint or an averted face; conversely, a pristine fingerprint image may be impaired by blur, compression, and/or transmission distortions, thus impairing recognizability).

In any case, as these diverse fields converge with mutual recognition of the importance of understanding, measuring, monitoring, and acting upon the quality of visual signals, principled approaches are certain to emerge whereby the effects of blindly measured quality degradations on visual tasks can be established.

It is quite possible that within a few years, image and video quality "agents" will be pervasive and a normal element of switches, routers, wireless access points, cameras and other mobile devices, as well as displays. Agents such as these could interact over large-scale networks, enabling distributed control and optimization of visual quality as the traffic becomes increasingly congested. ■

Acknowledgment

The author would like to thank A. Mittal and A. K. Moorthy for their discussions and for their help in generating some of the figures used in this paper.

REFERENCES

[1] "Image obsessed," *Nat. Geographic*, vol. 221, p. 35, Apr. 2012.

[2] D. Hardawar, "Netflix now accounts for 25% of North American Internet traffic," *VentureBeat*, May 17, 2011. [Online]. Available: http://venturebeat.com/2011/05/17/netflix-north-america-traffi/

[3] Cisco Corporation, Cisco Visual Networking Index: Global mobile data traffic forecast update, 2010–2015, 2011. [Online]. Available: http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white 520862.html

[4] T. S. Rappaport, J. N. Murdock, and F. Gutierrez, "State of the art in 60-GHz integrated circuits and systems for wireless communications," *Proc. IEEE*, vol. 99, no. 8, pp. 1390–1436, Aug. 2011.

[5] V. Chandrasekhar, J. G. Andrews, and A. Gatherer, "Femtocell networks: A survey," *IEEE Commun. Mag.*, vol. 46, no. 9, pp. 59–67, Sep. 2008.

[6] F. Crick and C. Koch, "Towards a neurobiological theory of consciousness," *Seminars Neurosci.*, vol. 2, pp. 263–275, 1990.

[7] J. G. Apostolopoulos, P. A. Chou, B. Culbertson, T. Kalker, M. D. Trott, and S. Wee, "The road to immersive communication," *Proc. IEEE*, vol. 100, no. 4, pp. 974–990, Apr. 2012.

[8] X. Cao, A. C. Bovik, and Y. Wang, "Converting 2D video to 3D: An efficient path to a 3D experience," *IEEE Multimedia*, vol. 18, no. 4, pp. 12–17, Apr. 2011.

[9] M. Tanimoto, M. P. Tehrani, T. Fujii, and T. Yendo, "FTV for 3-D spatial communication," *Proc. IEEE*, vol. 100, no. 4, pp. 905–917, Apr. 2012.

[10] E. P. Simoncelli, "Capturing visual image properties with probabilistic models," in *The Essential Guide to Image Processing*, A. C. Bovik, Ed. New York, NY, USA: Academic, 2009.

[11] B. Mandelbrot, *The Fractal Geometry of Nature*. San Francisco, CA, USA: Freeman, 1982.

[12] D. J. Tolhurst, Y. Tadmoor, and T. Chao, "Amplitude spectra of natural images," *Opthalmic. Physiol. Opt.*, vol. 12, pp. 229–232, 1992.

[13] A. J. Bell and T. J. Sejnowski, "The 'independent components' of natural images are edge filters," *Vis. Res.*, vol. 37, pp. 3327–3338, Dec. 1997.

[14] D. L. Ruderman, "The statistics of natural images," *Netw. Comput. Neural Syst.*, vol. 5, no. 4, pp. 517–548, 1994.

[15] D. L. Ruderman and W. Bialek, "Statistics of natural images: Scaling in the woods," *Phys. Rev. Lett.*, vol. 73, pp. 814–817, 1994.

[16] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using a scale mixture of Gaussians in the wavelet domain," *IEEE Trans. Image Process.*, vol. 12, no. 11, pp. 1338–1351, Nov. 2003.

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

[18] 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, Jul. 1989.

[19] E. P. Simoncelli, E. H. Adelson, and D. J. Heeger, "Probability distributions of optical flow," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Maui, HI, USA, Jun. 1991, pp. 310–315.

[20] D. Calow, N. Kruger, F. Worgotter, and M. Lappe, "Statistics of optic flow for self-motion through natural scenes," in *Dynamic Perception*, U. Ilg, H. Bulthoff, and H. Mallot, Eds. St. Augustin, Germany: Infix Verlag, 2004, pp. 133–138.

[21] S. Roth and M. Black, "On the spatial statistics of optical flow," *Int. J. Comput. Vis.*, vol. 74, no. 1, pp. 33–50, Jan. 2007.

[22] K. Seshadrinathan and A. C. Bovik, "A structural similarity metric for video based on motion models," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Honolulu, HI, USA, Apr. 2007, vol. 1, pp. 869–872.

[23] D. W. Dong and J. J. Atick, "Statistics of natural time-varying images," *Netw., Comput. Neural Syst.*, vol. 6, pp. 345–358, 1995.

[24] R. Soundararajan and A. C. Bovik, "Video quality assessment by reduced reference spatio-temporal entropic differencing," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 4, pp. 684–694, Apr. 2012.

[25] M. Carandini, J. B. Demb, V. Mante, D. J. Tolhurst, Y. Dan, B. A. Olshausen, J. L. Gallant, and N. C. Rust, "Do we know what the early visual system does?" *J. Neurosci.*, vol. 25, no. 46, pp. 10577–10597, Nov. 2005.

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

[27] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Amer. A*, vol. 4, pp. 2379–2394, 1987.

[28] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by V1?" *Vis. Res.*, vol. 37, no. 23, pp. 331–3325, 1997.

[29] H. K. Hartline and F. Ratliff, "Inhibitory interactions in the retina of Limulus," in *Handbook of Sensory Physiology*, vol. 7, M. G. Fuortes, Ed. Berlin, Germany: Springer-Verlag, 1972, pp. 381–447.

[30] M. V. Srinivasan, S. B. Laughlin, and A. Dubs, "Predictive coding: A fresh view of inhibition in the retina," *Proc. Roy. Soc. Lond. B*, vol. 216, no. 1205, pp. 427–459, Nov. 1982.

[31] B. A. Olshausen and D. J. Field, "How close are we to understanding V1?" *Neural Comput.*, vol. 17, pp. 1665–1699, 2005.

[32] J. G. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *J. Opt. Soc. Amer.*, vol. 2, no. 7, pp. 1160–1169, Jul. 1985.

[33] 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.

[34] L. Wiskott, J. M. Fellous, N. Kruger, and C. V. Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 775–779, Jul. 1997.

[35] B. S. Manjunath and W. Y. Ma, "Texture features for browsing and retrieval of image data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 8, pp. 837–842, Aug. 1996.

[36] S. E. Palmer, *Vision Science*. Cambridge, MA, USA: MIT Press, 1999.

[37] D. Chen and A. C. Bovik, "Visual pattern image coding," *IEEE Trans. Commun.*, vol. 38, no. 12, pp. 2137–2146, Dec. 1990.

[38] J. Shen, "On the foundations of vision modeling I. Weber's law and Weberized TV (total variation) restoration," *Physica D, Nonlinear Phenomena*, vol. 175, no. 3–4, pp. 241–251, 2003.

[39] D. A. Pollen and S. F. Ronner, "Phase relationships between adjacent simple cells in the visual cortex," *Science*, vol. 212, no. 4501, pp. 1409–1411, Jun. 1981.

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

[41] Z. Wang and A. C. Bovik, "Reduced and no-reference image quality assessment," *IEEE Signal Process. Mag.*, vol. 28, no. 6, pp. 29–40, Nov. 2011.

[42] G. E. Legge and J. M. Foley, "Contrast masking in human vision," *J. Opt. Soc. Amer.*, vol. 70, no. 12, pp. 1458–1470, Dec. 1980.

[43] D. J. Sakrison and J. L. Mannos, "The effects of a visual fidelity criterion on the encoding of images," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 4, pp. 525–536, Jul. 1974.

[44] A. N. Netravali and B. Prasada, "Adaptive quantization of picture signals using spatial masking," *Proc. IEEE*, vol. 65, no. 4, pp. 536–548, Apr. 1977.

[45] M. D. Swanson, B. Zhu, and A. H. Tewfik, "Multiresolution scene-based watermarking using perceptual models," *J. Sel. Areas Commun.*, vol. 16, no. 4, pp. 540–550, May 1998.

[46] D. W. Dong and J. J. Atick, "Temporal decorrelation: A theory of lagged and nonlagged responses in the lateral geniculate nucleus," *Netw., Comput. Neural Syst.*, vol. 6, no. 2, pp. 159–178, Feb. 1995.

[47] R. D. Kell, "Improvements related to electric picture transmission systems," U.S. Patent 341 811, 1931.

[48] A. B. Watson and A. J. Ahumada, Jr., "Model of human visual-motion sensing," *J. Opt. Soc. Amer. A.*, vol. 2, no. 2, pp. 322–342, Feb. 1985.

[49] E. H. Adelson and J. R. Bergen, "Spatiotemporal energy models for the perception of motion," *J. Opt. Soc. Amer. A.*, vol. 2, no. 2, pp. 284–299, Feb. 1985.

[50] D. J. Heeger, E. P. Simoncelli, and J. A. Movshon, "Computational models of cortical visual perception," *Proc. Nat. Acad. Sci.*, vol. 93, pp. 623–627, Jan. 1996.

[51] E. P. Simoncell and D. J. Heeger, "A model of neuronal responses in visual area MT," *Vis. Res.*, vol. 38, no. 5, pp. 743–761, 1998.

[52] 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.

[53] B. Girod, "Information theoretical significance of spatial and temporal masking in video signals," *Proc. SPIE—Int. Soc. Opt. Eng.*, vol. 1077, pp. 178–187, 1989.

[54] J. W. Suchow and G. A. Alvarez, "Motion silences awareness of visual change," *Current Biol.*, vol. 21, pp. 140–143, Jan. 2011.

[55] D. Burr and P. Thompson, "Motion psychophysics: 1985–2010," *Vis. Res.*, vol. 51, pp. 1431–1456, 2011.

[56] L. K. Choi, A. C. Bovik, and L. K. Cormack, "A flicker detector model of the motion silencing illusion," presented at the Annu. Meeting Vis. Sci. Soc., Naples, FL, May 2012.

[57] Z. Wang, A. C. Bovik, and B. Evans, "Blind measurement of blocking artifacts in images," in *Proc. IEEE Int. Conf. Image Process.*, Vancouver, BC, Canada, Sep. 2000, vol. 3, pp. 981–984.

[58] A. C. Bovik and S. Liu, "DCT-domain blind measurement of blocking artifacts in DCT-coded images," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Salt Lake City, UT, USA, May 2001, vol. 3, pp. 1725–1728.

[59] L. Meesters and J. Martens, "A single-ended blockiness measure for JPEG-coded images," *Signal Process.*, vol. 82, pp. 369–387, 2002.

[60] 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, Dec. 2002.

[61] F. Pan, X. Lin, S. Rahardja, W. Lin, E. Ong, S. Yao, Z. Lu, and X. Yang, "A locally adaptive algorithm for measuring blocking artifacts in images and videos," *Signal Process., Image Commun.*, vol. 19, no. 6, pp. 499–506, Jun. 2004.

[62] X. Feng and J. P. Allebach, "Measurement of ringing artifacts in JPEG images," *Proc. SPIE—Int. Soc. Opt. Eng.*, vol. 6076, Feb. 2006, DOI: 10.1117/12.645089.

[63] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "A no-reference perceptual blur metric," in *Proc. IEEE Int. Conf. Image Process.*, Rochester, NY, USA, Sep. 2002, vol. 3, pp. 57–60.

[64] E. Ong, W. Lin, Z. Lu, X. Yang, S. Yao, F. Pan, L. Jiang, and F. Moschetti, "A no-reference quality metric for measuring image blur," in *Proc. Int. Symp. Signal Process. Appl.*, Paris, France, Jul. 2003, vol. 1, pp. 469–472.

[65] R. Ferzli and L. Karam, "Human visual system based no-reference objective image sharpness metric," in *Proc. IEEE Int. Conf. Image Process.*, Atlanta, GA, USA, Oct. 2006, pp. 2949–2952.

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

[67] X. Zhu and P. Milanfar, "A no-reference sharpness metric sensitive to blur and noise," in *Proc. 1st Int. Workshop Quality Multimedia Experience*, San Diego, CA, USA, Jul. 2009, pp. 64–69.

[68] M. Farias, M. Moore, J. Foley, and S. Mitra, "Perceptual contributions of blocky, blurry, and fuzzy impairments to overall annoyance," *Proc. SPIE—Int. Soc. Opt. Eng.*, vol. 5292, pp. 109–120, Jun. 2004.

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

[70] A. K. Moorthy and A. C. Bovik, "Statistics of natural image distortions," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Dallas, TX, USA, Mar. 2010, pp. 962–965.

[71] Z. Wang and E. P. Simoncelli, "Reduced-reference image quality assessment using a wavelet domain natural image statistic model," *Proc. SPIE—Int. Soc. Opt. Eng.*, vol. 5666, pp. 149–159, Jan. 2005.

[72] 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.

[73] A. Mittal, A. K. Moorthy, and A. C. Bovik, "Blind/referenceless spatial image quality evaluator," in *Proc. Annu. Asilomar Conf. Signals Syst. Comput.*, Monterey, CA, USA, Nov. 2011, pp. 723–727.

[74] A. H. Nuttall, "Accurate efficient evaluation of cumulative or exceedance probability distributions directly from characteristic functions," NUSC, Tech. Rep. 7023., Oct. 1983.

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

[76] W. Lin and C.-C. Kuo, "Perceptual visual quality metrics: A survey," *J. Vis. Commun. Image Representation.*, vol. 22, no. 4, pp. 297–312, Apr. 2011.

[77] K. Seshadrinathan and A. C. Bovik, "Automatic prediction of perceptual quality of multimedia signals—A survey," *Int. J. Multimedia Tools Appl.*, vol. 51, no. 1, pp. 163–186, Jan. 2011.

[78] K. Seshadrinathan, T. N. Pappas, R. J. Safranek, J. Chen, Z. Wang, H. R. Sheikh, and A. C. Bovik, "Image quality assessment," in *The Essential Guide to Image Processing*, A. C. Bovik, Ed. New York, NY, USA: Elsevier, 2009.

[79] K. Seshadrinathan and A. C. Bovik, "Video quality assessment," in *The Essential Guide to Video Processing*, A. C. Bovik, Ed. New York, NY, USA: Elsevier, 2009.

[80] S. Winkler, *Digital Video Quality: Vision Models and Metrics*. Hoboken, NJ, USA: Wiley, Mar. 2005.

[81] Z. Wang and A. C. Bovik, *Modern Image Quality Assessment*. New York, NY, USA: Morgan & Claypool, 2006.

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

[83] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti, "TID2008—A database for evaluation of full-reference visual quality assessment metrics," *Adv. Modern Radioelectron.*, vol. 10, pp. 30–45, 2009.

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

[85] MeTriX MuX Visual Quality Assessment Package. [Online]. Available: http://foulard.ece.cornell.edu/gaubatz/metrix_mux/

[86] H. R. Sheikh, Z. Wang, A. C. Bovik, and L. K. Cormack, "Image and video quality assessment research at LIVE." [Online]. Available: http://live.ece.utexas.edu/research/quality/

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

[88] 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.

[89] M.-J. Chen and A. C. Bovik, "Fast structural similarity index algorithm," *J. Real-Time Image Process.*, vol. 6, no. 4, pp. 281–287, Dec. 2011.

[90] Z. Wang, E. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," in *Proc. Annu. Asilomar Conf. Signals Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2003, vol. 2, pp. 1398–1402.

[91] K. Seshadrinathan and A. C. Bovik, "Unifying analysis of full reference image quality assessment," in *Proc. IEEE Int. Conf. Image Process.*, San Diego, CA, USA, Oct. 2008, pp. 1200–1203.

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

[93] 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.

[94] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *J. Electr. Imaging*, vol. 19, no. 1, pp. 011006-1–011006-21, Jan.–Mar. 2010.

[95] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.

[96] Z. Wang and E. P. Simoncelli, "Local phase coherence and the perception of blur," in *Advances in Neural Information Processing Systems*, vol. 16. Cambridge, MA, USA: MIT Press, May 2004, pp. 786–792.

[97] C. Li and A. C. Bovik, "Content-partitioned structural similarity index for image quality assessment," *Signal Process., Image Commun.*, vol. 25, no. 7, pp. 517–526, Jul. 2010.

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

[99] C. Charrier, K. Knoblauch, L. T. Maloney, A. C. Bovik, and A. K. Moorthy, "Optimizing multi-scale SSIM for compression via MLDS," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4682–4694, Dec. 2012.

[100] A. M. Tourapis, A. Leontaris, K. Suhring, and G. Sullivan, "H.264/14496-10 AVC reference software manual," Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, Jul. 2009.

[101] Z. Wang, "Applications of objective image quality assessment methods," *IEEE Signal Process. Mag.*, vol. 28, no. 6, pp. 137–142, Nov. 2011.

[102] S. S. Channappayya, A. C. Bovik, C. Caramanis, and R. W. Heath, "Design of linear equalizers optimized for the structural similarity index," *IEEE Trans. Image Process.*, vol. 17, no. 6, pp. 857–872, Jun. 2008.

[103] S. S. Channappayya, A. C. Bovik, and R. W. Heath, "Rate bounds on SSIM index of quantized images," *IEEE Trans. Image Process.*, vol. 17, no. 9, pp. 1624–1639, Sep. 2008.

[104] M. Temerinac-Ott and M. Burkhardt, "Multichannel image restoration based on optimization of the structural similarity index," in *Proc. Asilomar Conf. Signals Syst. Comput.*, Pacific Grove, CA, USA, 2009, pp. 812–816.

[105] A. N. Avanaki, "Exact global histogram specification optimized for structural similarity," *Opt. Rev.*, vol. 16, no. 6, pp. 613–621, Nov. 2009.

[106] S. S. Channappayya, A. C. Bovik, and R. W. Heath, "Perceptual soft thresholding using the structural similarity index," in *Proc. IEEE Int. Conf. Image Process.*, San Diego, CA, USA, Oct. 2008, pp. 569–572.

[107] A. Rehman, Z. Wang, D. Brunet, and E. R. Vrscay, "SSIM-inspired image denoising using sparse representations," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Prague, Czech Republic, 2011, pp. 1121–1124.

[108] Z. Wang, Q. Li, and X. Shang, "Perceptual image coding based on a maximum of minimal structural similarity criterion," in *Proc. IEEE Int. Conf. Image Process.*, San Antonio, TX, USA, 2007, vol. 2, pp. 121–124.

[109] T. Richter and K. J. Kim, "A MS-SSIM optimal JPEG 2000 encoder," in *Proc. Data Compression Conf.*, Snowbird, UT, USA, 2009, pp. 401–410.

[110] A. C. Brooks, X. Zhao, and T. N. Pappas, "Structural similarity quality metrics in a coding context: Exploring the space of realistic distortions," *IEEE Trans. Image Process.*, vol. 17, no. 8, pp. 1261–1273, Aug. 2008.

[111] D. Brunet, E. R. Vrscay, and Z. Wang, "On the mathematical properties of the structural similarity index," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1488–1499, Apr. 2012.

[112] Y. H. Huang, T. S. Ou, P. Y. Su, and H. H. Chen, "Perceptual rate-distortion optimization using structural similarity index as quality metric," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 11, pp. 1614–1624, Nov. 2010.

[113] T.-S. Ou, Y.-H. Huang, and H. H. Chen, "SSIM-based perceptual rate control for video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 5, pp. 682–691, May 2011.

[114] Z. Wang, L. Lu, and A. C. Bovik, "Video quality assessment based on structural distortion measurement," *Signal Process., Image Commun.*, vol. 19, no. 2, pp. 121–132, Feb. 2004.

[115] P. V. Vu, C. T. Vu, and D. M. Chandler, "A spatiotemporal most-apparent-distortion model for video quality assessment," in *Proc. IEEE Int. Conf. Image Process.*, Brussels, Belgium, Sep. 2011, pp. 2505–2508.

[116] M. H. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," *IEEE Trans. Broadcast.*, vol. 50, no. 3, pp. 312–322, Sep. 2004.

[117] A. Ninassi, A. Le Meur, P. Le Callet, and D. Barba, "Considering temporal variations of spatial visual distortions in video quality assessment," *IEEE J. Sel. Top. Signal Process.*, vol. 3, no. 2, pp. 253–265, Apr. 2009.

[118] M. Barkowsky, J. Bialkowski, B. Eskofier, R. Bitto, and A. Kaup, "Temporal trajectory aware visual quality measure," *IEEE J. Sel. Top. Signal Process.*, vol. 3, no. 2, pp. 253–265, Apr. 2009.

[119] Z. Wang, G. Wu, H. R. Sheikh, E. Simoncelli, E. Yang, and A. C. Bovik, "Quality-aware images," *IEEE Trans. Image Process.*, vol. 15, no. 5, pp. 1680–1689, May 2006.

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

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

[122] R. Soundararajan and A. C. Bovik, "RRED indices: Reduced reference entropic differencing for image quality assessment," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 517–526, Feb. 2012.

[123] S. S. Hemami and A. R. Reibman, "No-reference image and video quality estimation: Applications and human-motivated design," *Signal Process., Image Commun.*, vol. 25, no. 7, pp. 469–481, Aug. 2010.

[124] M. C. Q. Farias and S. K. Mitra, "No-reference video quality metric based on artifact measurements," in *Proc. IEEE Int. Conf. Image Process.*, Genoa, Italy, Nov. 2005, vol. 3, pp. 141–144.

[125] M. C. Q. Farias and S. K. Mitra, "A methodology for designing no-reference video quality metrics," in *Proc. Int. Workshop Video Process. Quality Metrics Consumer Electron.*, Scottsdale, AZ, USA, Jan. 2009, pp. 140–145.

[126] Z. Wang, H. R. Sheikh, and A. C. Bovik, "No-reference perceptual quality assessment of JPEG compressed images," in *Proc. IEEE Int. Conf. Image Process.*, Rochester, NY, USA, Sep. 2002, vol. 1, pp. 477–480.

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

[128] T. Oelbaum, K. Diepold, and C. Keimel, "Rule-based no-reference video quality evolution using additionally coded videos," *J. Sel. Topics Signal Process.*, vol. 3, no. 2, pp. 294–303, 2009.

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

[130] 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.

[131] H. Tong, M. Li, H. Zhang, C. Zhang, J. He, and W. Ma, "Learning no-reference quality metric by examples," in *Proc. Int. Multimedia Model. Conf.*, Jan. 2005, pp. 247–254.

[132] 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.

[133] M. A. Saad and A. C. Bovik, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3339–3352, Aug. 2012.

[134] E. Y. Lam and J. W. Goodman, "A mathematical analysis of the DCT coefficient distributions for images," *IEEE Trans. Image Process.*, vol. 9, no. 10, pp. 1661–1666, Oct. 2000.

[135] J. Shen, Q. Li, and G. Erlebacher, "Hybrid no reference natural image quality assessment of noisy, blurry, JPEG 2000 and JPEG images," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2089–2098, Aug. 2011.

[136] H. Tang, N. Joshi, and A. Kapoor, "Learning a blind measure of perceptual image quality," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Toronto, ON, Canada, Jun. 2011, pp. 305–312.

[137] P. Ye and D. Doermann, "No reference image quality assessment based on visual codebook," in *Proc. IEEE Int. Conf. Image Process.*, Brussels, Belgium, Nov. 2011, pp. 3089–3092.

[138] H. R Sheikh, Z. Wang, L. K. Cormack, and A. C. Bovik, "LIVE Image Quality Assessment Database Release 2, 2005. [Online]. Available: http://live.ece.utexas.edu/research/quality

[139] A. Mittal, G. S. Muralidhar, J. Ghosh, and A. C. Bovik, "Blind image quality assessment without human training using latent quality factors," *IEEE Signal Process. Lett.*, vol. 19, no. 2, pp. 75–78, Feb. 2012.

[140] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 21, no. 3, pp. 209–212, Mar. 2013.

[141] M. Saad and A. C. Bovik, "Blind quality assessment of videos using a model of natural scene statistics and motion coherency," in *Proc. Annu. Asilomar Conf. Signals Syst. Comput.*, Monterey, CA, USA, Nov. 2012, pp. 332–336.

[142] A. Moorthy, "Natural scene statistic based blind image quality assessment and repair," Ph.D. dissertation, Dept. Elec. Comput. Engr., Univ. Texas at Austin, Austin, TX, USA, 2012.

[143] S. Chen, A. Beghdadi, and A. Chetouani, "A new color image quality index," in *Proc. Int. Workshop Video Process. Quality Metrics Consumer Electron.*, Scottsdale, AZ, USA, Jan. 2010, pp. 267–271.

[144] M. Hassan and C. Bhagvati, "Structural similarity measure for color images," *Int. J. Comput. Appl.*, vol. 43, no. 14, pp. 7–12, Apr. 2012.

[145] A. Kolaman and O. Yadid-Pecht, "Quaternion structural similarity: A new quality index for color images," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1546–1536, Apr. 2012.

[146] K. Okarma, "Colour image quality assessment using structural similarity index and singular value decomposition," in *Lecture Notes in Computer Science*, vol. 5101. Berlin, Germany: Springer-Verlag, 2008, pp. 55–65.

[147] X. Zhang and B. Wandell, "Color image fidelity metrics evaluated using image distortion maps," *Signal Process.*, vol. 70, pp. 201–214, 1998.

[148] C. Oleari, "Color opponencies in the system of the uniform color scales of the Optical Society of America," *J. Opt. Soc. Amer. A*, vol. 21, no. 5, pp. 677–682, May 2004.

[149] A. Benoit, P. Le Callet, P. Campisi, and R. Cousseau, "Quality assessment of stereoscopic images," *EURASIP J. Image Video Process.*, 2008, DOI: 10.1155/2008/659024.

[150] J. Yang, C. Hou, Y. Zhou, Z. Zhang, and J. Guo, "Objective quality assessment method of stereo images," in *Proc. IEEE 3DTV Conf.*, Potsdam, Germany, 2009, DOI: 10.1109/3DTV.2009.5069615.

[151] R. Olsson and M. A. Sjostrom, "A depth dependent quality metric for evaluation of coded integral imaging based 3D-images," in *Proc. IEEE 3DTV Conf.*, Kos, Greece, 2007, DOI: 10.1109/3DTV.2007.4379397.

[152] P. Gorley and N. Holliman, "Stereoscopic image quality metrics and compression," *Proc. SPIE—Int. Soc. Opt. Eng.*, vol. 6803, 2008, DOI: 10.1117/12.763530.

[153] A. K. Moorthy, C.-C. Su, A. Mittal, and A. C. Bovik, "Subjective evaluation of stereoscopic image quality," *Signal Process., Image Commun.*, Sep. 2012. [Online]. Available: http://dx.doi.org/10.1016/j.image.2012.08.004

[154] R. Bensalma and M. C. Larabi, "A perceptual metric for stereoscopic image quality assessment based on the binocular energy," *Multidimensional Syst. Signal Process.*, vol. 24, no. 2, pp. 281–316, Jun. 2013, DOI: 10.1007/s11045-012-0178-3.

[155] C.-C. Su, A. C. Bovik, and L. K. Cormack, "Natural scene statistics of color and range," in *Proc. IEEE Int. Conf. Image Process.*, Brussels, Belgium, Sep. 2011, pp. 257–260.

[156] Y. Liu, L. K. Cormack, and A. C. Bovik, "Disparity statistics in natural scenes," *J. Vis.*, vol. 8, no. 11, pp. 1–14, Aug. 2008.

[157] Y. Liu, A. C. Bovik, and L. K. Cormack, "Statistical modeling of 3D natural scenes with application to Bayesian stereopsis," *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2515–2530, Sep. 2011.

[158] G. W Cermak, "Consumer opinions about frequency of artifacts in digital video," *IEEE J. Sel. Top. Signal Process.*, vol. 3, no. 2, pp. 336–343, Apr. 2009.

[159] D. Rouse, R. Pepion, S. S. Hemami, and P. Le Callet, "Image utility assessment and a relationship with image quality assessment," *Proc. SPIE—Int. Soc. Opt. Eng.*, vol. 7240, Jan. 2009, doi: 10.1117/12.811664.

[160] D. Rouse and S. S. Hemami, "Quantifying the use of structure in cognitive tasks," *Proc. SPIE—Int. Soc. Opt. Eng.*, vol. 6492, Jan. 2007, DOI: 10.1117/12.707539.

[161] D. Rouse and S. S. Hemami, "Analyzing the role of visual structure in the recognition of natural image content with multi-scale SSIM," *Proc. SPIE—Int. Soc. Opt. Eng.*, vol. 6806, Jan. 2008, DOI: 10.1117/12.768060.

[162] A. Bedagkar-Gala and S. K. Shah, "Joint modeling of algorithm behavior and image quality for algorithm performance prediction," in *Proc. British Mach. Vis. Conf.*, Aberystwyth, U.K., Sep. 2010, DOI: 10.5244/C.24.31.

[163] M. Abdel-Mottaleb and M. H. Mahoor, "Algorithms for assessing the quality of facial images," *IEEE Comput. Intell. Mag.*, vol. 2, no. 2, pp. 10–17, May 2007.

[164] R.-L. V. Hsu, J. Shah, and B. Martin, "Quality assessment of facial images," in *Proc. IEEE Biometrics Symp.*, Baltimore, MD, USA, Sep. 2006, DOI: 10.1109/BCC.2006.4341617.

[165] Y. Chen, S.-C. Dass, and A. K. Jain, "Localized iris image quality using 2-D wavelets *Lecture Notes in Computer Science*, vol. 3832. Berlin, Germany: Springer-Verlag, 2005, pp. 373–381.

[166] N. D. Kalka, J. Zuo, N. A. Schmid, and B. Cukic, "Image quality assessment for iris biometric," *Proc. SPIE—Int. Soc. Opt. Eng.*, vol. 6202, 2006, DOI: 10.1117/12.666448.

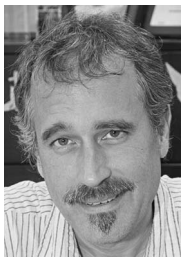
[167] K. W. Bowyer, K. Hollingsworth, and P. J. Flynn, "Image understanding for iris biometrics: A survey," *Comput. Vis. Image Understand.*, vol. 110, no. 2, pp. 281–307, Feb. 2008.

[168] M. Subasic, S. Loncaric, T. Petkovic, and H. Bogunovic, "Face image validation system," in *Proc. Int. Symp. Image Signal Process. Anal.*, 2008, pp. 30–33.

[169] L. Breitenbach and P. Chawdhry, "Image quality assessment and performance evaluation for multimodal biometric recognition using face and iris," in *Proc. Int. Symp. Image Signal Process. Anal.*, 2009, pp. 550–555.

[170] Information Technology—Biometric Data Interchange Formats—Part 5: Face Image Data, ISO/IEC 19794-5, Jun. 2005.

ABOUT THE AUTHOR



Alan Conrad Bovik (Fellow, IEEE) received the B.S., M.S., and Ph.D. degrees from the University of Illinois at Urbana-Champaign, Urbana, IL, USA, in 1980, 1982, and 1984, respectively.

He holds the Keys and Joan Curry/Cullen Trust Endowed Chair at The University of Texas at Austin, Austin, TX, USA, where he is a Professor in the Department of Electrical and Computer Engineering and in the Institute for Neuroscience, and Director of the Laboratory for Image and Video Engineering (LIVE). During spring 1992, he held a visiting position in the Division of Applied Sciences, Harvard University, Cambridge, MA, USA. He has made contributions to, among others, the fields of image and video processing, computational vision, digital microscopy, and modeling of biological visual perception. He has published over 650 technical articles in these areas and holds several U.S. patents. He is the author or coauthor of *The Handbook of Image and Video Processing* (New York, NY, USA: Academic, 2nd ed., 2005), *Modern Image Quality Assessment* (New York, NY, USA: Morgan & Claypool, 2006), *The Essential Guide to Image Processing* (New York, NY, USA: Academic, 2008), and *The Essential Guide to Video Processing* (New York, NY, USA: Academic, 2008).

Prof. Bovik is a Fellow of the Optical Society of America (OSA) and a Fellow of The International Society for Optical Engineers (SPIE). He has received a number of major awards from the IEEE Signal Processing Society, including: the Best Paper Award (2009), the Education Award (2008), the Technical Achievement Award (2005), the Distinguished Lecturer Award (2000), and the Meritorious Service Award (1998). Recently, he was named Honorary Member of IS&T (2012) and received the SPIE Technology Achievement Award (2012). He was also the IS&T/SPIE Imaging Scientist of the Year for 2011. He was the recipient of the Hocott Award for Distinguished Engineering Research from the Cockrell School of Engineering at The University of Texas at Austin (2008), the Distinguished Alumni Award from the University of Illinois at Champaign-Urbana (2008), the IEEE Third Millennium Medal (2000), and several paper awards. He has served in many and various capacities, including Board of Governors of the IEEE Signal Processing Society (1996–1998), Editor-in-Chief of the IEEE TRANSACTIONS ON IMAGE PROCESSING (1996–2002), Overview Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING (2009–present), Editorial Board of THE PROCEEDINGS OF THE IEEE (1998–2004), Senior Editorial Board of the IEEE JOURNAL ON SELECTED TOPICS IN SIGNAL PROCESSING (2005–2009), Associate Editor of the IEEE SIGNAL PROCESSING LETTERS (1993–1995), and Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING (1989–1993). He founded and served as the first General Chairman of the IEEE International Conference on Image Processing, held in Austin, TX, USA, in November 1994. He is a registered Professional Engineer in the State of Texas and is a frequent consultant to legal, industrial, and academic institutions.