Hindawi Publishing Corporation
ISRN Signal Processing
Volume 2013, Article ID 905685, 53 pages
http://dx.doi.org/10.1155/2013/905685

Review Article

Seven Challenges in Image Quality Assessment: Past, Present, and Future Research

Damon M. Chandler

School of Electrical and Computer Engineering, Oklahoma State University, Stillwater, OK 74078, USA

Correspondence should be addressed to Damon M. Chandler; damon.chandler@okstate.edu

Received 9 October 2012; Accepted 15 November 2012

Academic Editors: S. Li, C. S. Lin, and K. Wang

Image quality assessment (IQA) has been a topic of intense research over the last several decades. With each year comes an increasing number of new IQA algorithms, extensions of existing IQA algorithms, and applications of IQA to other disciplines. In this article, I first provide an up-to-date review of research in IQA, and then I highlight several open challenges in this field. The first half of this article provides discuss key properties of visual perception, image quality databases, existing full-reference, no-reference, and reduced-reference IQA algorithms. Yet, despite the remarkable progress that has been made in IQA, many fundamental challenges remain largely unsolved. The second half of this article highlights some of these challenges. I specifically discuss challenges related to lack of complete perceptual models for: natural images, compound and suprathreshold distortions, and multiple distortions, and the interactive effects of these distortions on the images. I also discuss challenges related to IQA of images containing nontraditional, and I discuss challenges related to the computational efficiency. The goal of this article is not only to help practitioners and researchers keep abreast of the recent advances in IQA, but to also raise awareness of the key limitations of current IQA knowledge.

1. Introduction

Digital imaging and image-processing technologies have revolutionized the way in which we capture, store, receive, view, utilize, and share images. Today, we have come to expect the ability to instantly share photos online, to send and receive multimedia MMS messages at a moment's notice, and to stream live video across the globe instantaneously. Today, these conveniences are possible because the digital cameras and photo-editing systems used by photographers and artists, the compression and transmission systems used by distributors and network engineers, and the various multimedia and display technologies enjoyed by consumers all have the ability to process images in ways that were unthinkable just 20 years ago.

But despite the innovation and rapid advances in technology and despite the prevalence of higher-definition and more immersive content, one thing has remained constant throughout the digital imaging revolution: the biological hardware used by consumers—the human visual system. Although personal preferences can and do change over time and can and do vary from person to person, the underlying neural circuitry and biological processing strategies have changed very little over measurable human history. As a result, digital processing can alter an image's appearance in ways that humans can reliably and consistently judge to be either detrimental or beneficial to the image's visual quality.

Because of the prevalence of these alterations, a crucial requirement for any system that processes images is a means of assessing the impacts of such alterations on the resulting visual quality. To meet this need, numerous algorithms for image quality assessment (IQA) have been researched and developed over the last several decades. Today, IQA research has emerged as an active subdiscipline of image processing, and many of the resulting techniques and algorithms have begun to benefit a wide variety of applications. Variations of IQA algorithms have proved useful for applications such as image and video coding (e.g., [1–3]), digital watermarking (e.g., [4–8]), unequal error protection (e.g., [9]), denoising (e.g., [10]), image synthesis (e.g., [11, 12]), and various other areas (e.g., for predicting intelligibility in sign language video [13]).

Many of the techniques employed by modern IQA algorithms are founded in the early research on quality evaluation

of optical systems and analog television broadcast and display systems (e.g., [14–25]). For example, in their 1940 paper titled "Quality in Television Pictures," Goldmark and Dyer [16] stated that

> *"The factors which chiefly determine the quality of a television picture are (1) definition, (2) contrast range, (3) gradation, (4) brilliance, (5) flicker, (6) geometric distortion, (7) size, (8) color, and (9) noise." [16].*

Although no objective quality assessment formulae were presented in [16], many of today's IQA algorithms do indeed employ measures of one or more of these factors. Later work by Winch [17], on the topic of color TV quality, further pushed toward objective quality assessment by providing guidelines for how photometric and colorimetric properties could be used to derive "characteristic data for correlation with the subjective preferences"; such properties are now commonly employed in modern IQA algorithms. On the optics front, in their 1955 paper titled "On the Assessment of Optical Images," Fellgett and Linfoot proposed two key strategies and associated numerical measures of image quality: "*assessment by similarity*" and "*assessment by information content*" [19]. Indeed, variations of these ideas have been used by many of today's IQA algorithms.

It is interesting to note that nearly all of these early research efforts up through the 1960s mentioned the need to take into account the characteristics of human vision during the quality assessment process. Five of the earliest efforts to explicitly model properties of the human visual system (HVS) for IQA were published in the early 1970s by Sakrison and Algazi [22], by Budrikis [23], by Stockham [24], by Mannos and Sakrison [26], and by Schade [25]. Although no extensive IQA algorithms were presented in these early papers, many of the properties which are used in modern HVS-based IQA algorithms—such as luminance and contrast sensitivity and visual masking—were also suggested in these papers. At that time, Budrikis forecasted that

> *"Full evaluations are as yet impossible but seem very likely for the foreseeable future, although probably entailing considerable computational tasks." [23].*

Today, 40 years later, we still have yet to achieve *full* evaluations of quality, though remarkable progress has been made—as I will point out in this paper.

At a glance, the IQA problem for digital images may not seem as a difficult task as reported in the literature. After all, digital processing alters an image's pixel values, and the task of estimating quality requires merely mapping these numerical changes to corresponding visual preferences. Of course, anything that involves the human visual system is rarely straightforward. Humans do not see images as collections of pixels, and consequently, the appropriate mapping varies depending on the image, on the type of processing, on the numerical and psychological interaction between these two, and on numerous additional factors. As an example, Figure 1 shows an original image and 11 altered versions of that image, each with the same peak signal-to-noise ratio in comparison to the reference. Clearly, a mapping based only on the energy of the differences in pixel values cannot capture the wide range of visual qualities exhibited by these images.

The task of judging quality in Figure 1 is facilitated by the presence of an original, undistorted reference image. In his seminal 1975 collection titled "Image Quality: A Comparison of Photographic and Television Systems," Schade [25] stated that

> *"Image quality is a subjective judgment made by a mental comparison of an external image with image impressions stored and remembered more or less distinctly by the observer. . .. Moreover, the rating of a given image may be greatly influenced by the availability of a much better image for comparison purposes." [25].*

Most IQA algorithms operate in this relative-to-a-reference fashion; these are so-called *full-reference* algorithms, which take as input a reference image and a processed (usually distorted) image and yield as output either a scalar value denoting the overall visual quality or a spatial map denoting the local quality of each image region (see Section 3). More recently, researchers have begun to develop *no-reference* and *reduced-reference* algorithms, which attempt to yield the same quality estimates either by using only the processed/distorted image (no-reference IQA; see Section 4.1) or by using the processed/distorted image and only partial information about the reference image (reduced-reference IQA; see Section 4.2).

All three types of IQA algorithms can perform quite well at predicting quality. Some of today's best-performing full-reference algorithms have been shown to generate estimates of quality that correlate highly with human ratings of quality, typically yielding Spearman's and Pearson's correlation coefficients in excess of 0.9. Research in no-reference and reduced-reference IQA is much less mature; however, recent methods have been shown to yield quality estimates which also correlate highly with human ratings of quality, sometimes yielding correlation coefficients which rival the most competitive full-reference methods.

The field of image quality assessment is rapidly advancing. With each year comes an increasing number of papers on new IQA algorithms, extensions of existing IQA algorithms, and applications of these IQA techniques to other disciplines. Here, the objective of this paper is not only to provide an overview of the strategies used in IQA algorithms, but also—and more so—to highlight the current challenges in this field. This paper is meant to complement previous reviews and chapters on IQA [28–36] (see also [37, 38] for related reviews on video quality assessment). Here, I first provide a more recent survey of research in IQA to help practitioners and researchers keep abreast of the recent advances in IQA. Next, I discuss several open research challenges which are needed to further push IQA algorithms toward achieving the "full evaluations" envisioned by Budrikis.

In the first three sections of this paper, I provide an up-to-date survey of research in IQA. As in previous reviews, Section 2 summarizes several important properties of human visual perception which are used, at least to some extent,

Figure 1: Original and distorted versions of the image *child swimming* from the CSIQ database [27]. Note the large variation in perceived quality, despite the fact that all distorted images have the same peak signal-to-noise ratio of 22.5 dB.

directly or indirectly by the vast majority of IQA algorithms. However, I also discuss with each of these properties some of the early experiments in vision science that were performed to uncover the properties; the goal of this discussion is to provide a context which can help define bounds on the applicability of each of the properties. This section also provides an up-to-date survey of the publicly available ground-truth datasets (image quality databases) that can be used to quantify the performances of IQA algorithms in predicting quality. (Note that research on color perception and the specific effects of color on image quality are not covered in this paper. Color-perception research has its own long history, most of which predates research in IQA. The reader is referred to [39–41] for discussions on the influences of color on image quality.).

Sections 3 and 4 provide concise surveys of previous and recent IQA algorithms. Again, the primary objective of these surveys is to help the reader keep abreast of the latest IQA techniques. Section 3 surveys full-reference IQA algorithms. Section 4 surveys no-reference and reduced-reference IQA algorithms. For a more specific and thorough discussion of the use of natural-scene statistics for image and video quality assessptment, I refer you to the recent review by Bovik [36].

One point should become evident after reading previous reviews and the reviews provided in Sections 2, 3, and 4 of this paper: remarkable progress has been made since the pioneering IQA work of Budrikis, Goldmark, Sakrison, Schade, Stockham, Winch, and others. Today's IQA algorithms can perform extremely well at predicting quality for a variety of images and distortion types.

Yet, beneath the surface of this seemingly orderly picture, behind the scenes of this wealth of IQA knowledge that we have gained lies a more cloudy portrait fueled by a growing number of counterexamples—images, distortions, and other alterations—which modern IQA algorithms are ill-equipped to handle. Under the covers of the numerous successes in IQA research lies a long list of unanswered questions and unsolved challenges.

In Section 5, I discuss seven of these challenges. Some of the challenges are fundamental; some are more application-specific; most of the challenges have been or are actively being researched. But all remain largely unsolved.

(1) Section 5.1 discusses the challenges IQA researchers face when designing a model of human visual processing which can cope with natural images. This section highlights the need for improved models of primary visual cortex, the need for more ground-truth data on natural images, and the need for models which incorporate processing by higher-level visual areas.

(2) Section 5.2 discusses the challenges researchers face when designing an algorithm that can cope with the variety of distortions that IQA algorithms can encounter. This section discusses the need for improved visual summation models which can handle the broadband nature of distortions, and the need for more research on the perception of suprathreshold distortions.

(3) Section 5.3 discusses the challenges researchers face when designing an IQA algorithm that can model the influence of the distortion on the image's appearance. This section discusses the differences between distortions which are perceived as additive and distortions which affect the image's objects. This section also highlights the need to consider the adaptive visual strategies and other higher-level effects that humans use when judging quality.

(4) Section 5.4 discusses the challenges researchers face when designing an IQA algorithm that can cope with images which are simultaneously distorted by multiple types of distortion. This section reviews previous work on the effects of multiple distortions on image quality, and it discusses the potential perceptual interactions between the distortions and their joint effects on images.

(5) Section 5.5 discusses the challenges researchers face when designing an IQA algorithm that can deal with geometric changes to images. This section reviews existing IQA algorithms which have been designed to handle basic geometric changes, and it discusses research efforts on IQA of textures, which can contain more radical geometric and photometric changes.

(6) Section 5.6 discusses the challenges researchers face when designing an IQA algorithm that can perform IQA of enhanced images. This section describes efforts to model the perceptual effects of enhancement on quality, and it discusses the need for more thorough image quality databases which contain enhanced images.

(7) Section 5.7 discusses the challenges surrounding runtime performance and memory requirements of IQA algorithms. This section reviews previous efforts to accelerate existing IQA algorithms, and it discusses the need for further related performance analyses and accelerations.

It is important to note that these seven challenges are by no means an exhaustive list of research topics in IQA that require further investigation. Rather, I have selected these particular challenges to highlight some key limitations of current IQA knowledge and to point out areas which can begin to answer broader questions on IQA. Additional important open challenges can be garnered from the *Proceedings of SPIE* "*Image Quality and System Performance*" and "*Human Vision and Electronic Imaging*," among others.

2. Image Quality Assessment by Humans

A common approach toward designing an IQA algorithm is to first consider the physical attributes of images that humans find pleasing or displeasing. By understanding how these physical changes give rise to perceptual changes, one can begin to develop an estimate of image quality based on measures of the physical changes. Numerous studies in the fields of visual psychophysics and visual neuroscience have quantified relationships between the physical attributes of visual stimuli and the corresponding psychological and neurophysiological responses. The results of these studies have provided important insights into the goals and functions of the HVS, and many of these findings have been used in IQA algorithms. In Section 2.1, I provide a brief review of the basic properties of the HVS that are commonly taken into account—either explicitly or implicitly—in the vast majority of IQA algorithms.

Another approach toward gaining insight into how humans judge quality is to directly collect quality ratings from a representative pool of human subjects on a database of altered images. Several such quality-rating studies have been conducted, and the results of these studies are commonly released in the form of the so-called *image quality databases*. These databases generally contain the set of reference and altered images used in the study, along with corresponding average quality ratings for each altered image. In Section 2.2, I provide a survey of the various publicly available image quality databases, including a brief discussion of how the data are used for quantifying the predictive performances of IQA algorithms.

2.1. Psychophysical Underpinnings of Image Quality. Research in visual psychophysics aims to provide a better understanding of the human visual system (HVS) by linking changes in the *physical* attributes of a visual stimulus the to corresponding changes in *psychological* responses (visual perception and cognition). These studies generally entail carefully designed experiments on human subjects using highly controlled visual stimuli and viewing conditions. Many of the most fundamental properties of visual perception which are used for IQA have been obtained from the results of such studies; the most commonly used of these properties are summarized in this section.

It must be stressed that the primary goal of the vast majority of research in visual psychophysics is to gain knowledge of how the HVS operates; any relations to image quality are usually secondary and are usually not extensively discussed in such studies. Consequently, it is often up to the designer of an IQA algorithm to decide how the psychophysical findings

relate to image quality. Nonetheless, due in part to the increasing popularity of IQA algorithms, an increasing number of psychophysical studies have been devoted specifically toward image quality (e.g., [42–68]).

2.1.1. Contrast Sensitivity Function. Psychophysical studies have shown that the minimum contrast needed to detect a visual target (e.g., distortions) depends on the spatial frequency of the target [69, 70]. This minimum contrast is called the *contrast detection threshold*, and the inverse of this threshold is called *contrast sensitivity*. When contrast sensitivity is plotted as a function of the spatial frequency of the target, the resulting profile is the *contrast sensitivity function* (CSF).

Contrast thresholds for sine waves were first measured by Schade [71] in an experiment that presented human observers with achromatic sine-wave gratings of various spatial frequencies. The key result of Schade's experiment was the discovery that contrast sensitivity varies with the spatial frequency of the grating; the resulting CSF is bandpass, indicating that we are least sensitive to very-low-frequency and very-high-frequency targets, with a peak in sensitivity near 4–6 cycles per degree of visual angle (c/deg).

The reduction in sensitivity at high frequencies has been attributed to the optics of the eye, to receptor spacing, and to quantum noise. Reduced sensitivity at low spatial frequencies is believed to occur, in part, by limited receptive field sizes and by masking effects imposed by the target's DC component. However, when contrast sensitivity is measured using Gabor functions, the CSF tends to be much more low pass [73], and such low-pass-type CSFs are most commonly utilized in IQA algorithms. The CSF has also been measured as a function of the orientation of the sine-wave grating, commonly resulting in reduced sensitivity to diagonal orientations as compared to horizontal and vertical orientations (the *oblique effect* [74, 75]). Alternative theories of the neural underpinnings of the CSF have also been proposed based on the statistical properties of natural scenes [76, 77].

In IQA algorithms, the CSF is commonly taken into account by prefiltering the images with a 2D spatial filter designed based on the psychophysical results. One popular CSF filter, which is shown in Figure 2, was proposed by Mannos and Sakrison [26] and further adjusted by Daly [72]; its frequency response, $H(f,\theta)$, is given by

$$H(f,\theta) = \begin{cases} 2.6\left(0.019 + \lambda f_\theta\right) e^{-(\lambda f_\theta)^{1.1}}, & \text{if } f \geq f_{\text{peak}} \text{ c/deg}, \\ 0.981, & \text{otherwise}, \end{cases} \tag{1}$$

where f denotes the radial spatial frequency in c/deg and $\theta \in [-\pi, \pi]$ denotes the orientation and where

$$f_\theta = \frac{f}{[0.15\cos(4\theta) + 0.85]} \tag{2}$$

accounts for the oblique effect (see [72]). In Figure 2, the parameter λ was set to $\lambda = 0.228$, resulting in the CSF taking on its maximum value of 0.981 at $f_{\text{peak}} = 4$ c/deg (and forced to be this value for frequencies below f_{peak}) when $\theta = 0$ or $\pi/2$. A very thorough treatment of the use of the CSF in IQA has been published by Barten [78].

2.1.2. Visual Masking. Another finding from the visual perception research which is commonly taken into account in IQA algorithms is the fact that certain regions of an image can hide distortions better than other regions, a finding that can be attributed to *visual masking* [79]. Visual masking is a general term that refers to the perceptual phenomenon in which the presence of masking signal (the *mask*) reduces a subject's ability to detect a given target signal. The task of detection becomes a *masked detection*, and contrast thresholds denote *masked detection thresholds*. In IQA, it is commonly assumed that the image serves as the mask and the distortions serve as the target of detection.

Luminance masking and pattern masking are the two most common forms of masking employed in IQA algorithms. Detection thresholds tend to increase due to an increase in the luminance of the background (mask) upon which the target is placed (luminance masking, [70, 80]), a process which is believed to be mediated by retinal adaptation [81]. For masks consisting of spatial patterns, detection thresholds also tend to increase when the contrast of the mask is increased [69, 79, 82], a postretinal process believed to be attributable to cortical processing [81]. Current explanations of pattern masking can generally be divided into three paradigms:

(1) *noise masking*, which attributes the reduction in sensitivity to the corruptive effects of the mask on internal decision variables [83];

(2) *contrast masking*, which attributes reduction in sensitivity to contrast gain control [79] (discussed later);

(3) *entropy masking*, which attributes reduction in sensitivity to an observer's unfamiliarity with the mask [44].

Because a mask's contrast is readily computable, contrast masking has been exploited in a variety of IQA and image processing applications (e.g., [84–88]; see Section 3.1). The extent to which a mask constitutes visual noise and the extent to which an observer is unfamiliar with a mask are phenomena which are more difficult to quantify; accordingly, noise and entropy masking are less commonly used in IQA (though, see [89]).

Contrast masking results are commonly reported in the form of threshold-versus-contrast (TvC) curves, in which masked detection thresholds are plotted as a function of the contrast of the mask. Figure 3 depicts TvC curves for the detection of a sine-wave grating presented against noise and sine-wave-grating masks (after [79]). Masked detection thresholds generally increase as the contrast of the mask is increased and often demonstrate a region of facilitation (i.e., a decrease in threshold; "dipper effect") at lower mask contrasts, depending on the dimensional relationships between the target and the mask (e.g., differences in spatial frequency, orientation, and phase). Note that learning effects have been shown to lower the slopes of the TvC curves [87, 90].

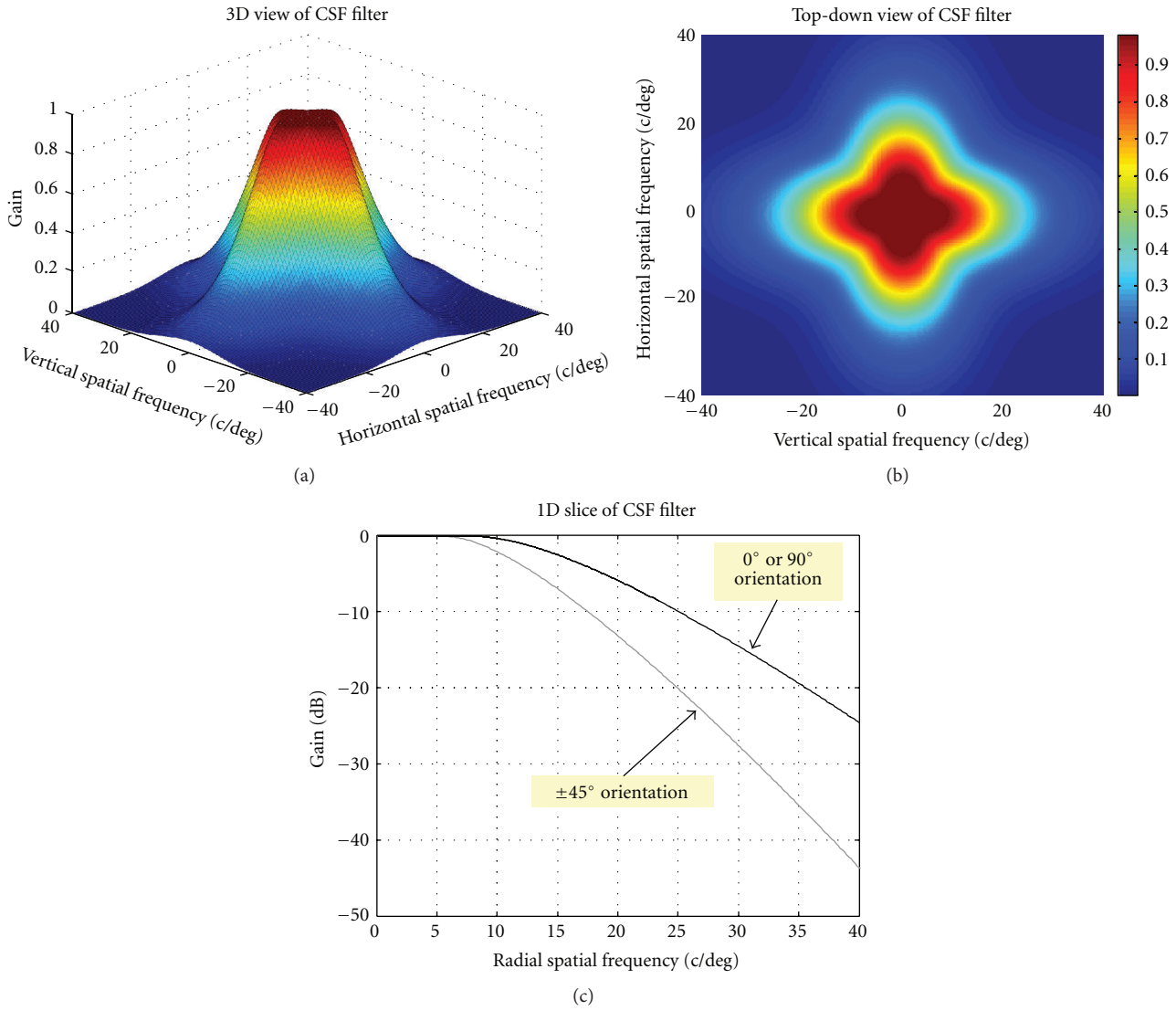


FIGURE 2: Frequency response of a 2D filter used to model the CSF. This particular model is from Mannos and Sakrison [26] with further adjusting by Daly [72]; see (1).

In IQA, a variety of methods have been used to account for masking, particularly in full-reference IQA (discussed later in Section 3). A common approach to explicitly account for masking is to measure the local luminance and contrast in the reference image and then attenuate the estimate of the visibility of the distortions in the distorted image based on these measures (e.g., using a power-function relationship between attenuation and contrast). Other IQA algorithms implicitly incorporate masking either by using local statistical measures which take into account the local contrast or by adjusting the simulated neural responses in the context of a computational neural model of the HVS (see Section 3).

2.1.3. Multichannel Model of the HVS. Schade used sine-wave gratings in his CSF study based on the notion that any stimulus can be described as a superposition of sine-waves. Campbell and Robson [91] extended this idea by measuring detection thresholds for both sine-wave and square-wave gratings. Because a square wave is composed of numerous sine waves, the peak-to-peak contrast of a square wave will always be lower than the peak-to-peak contrast of its fundamental sine wave (by a factor of approximately 1.3 in [91]). The results from Campbell and Robson's experiment revealed that the thresholds for the square-wave gratings were indeed approximately 1.3 times lower than those found for the sine-wave gratings. They concluded from this finding that the HVS performs a local spatial-frequency decomposition of a stimulus in which the frequency components are detected independently via multiple spatial-frequency *channels*. This paradigm is known as the *multichannel model* of human vision [82].

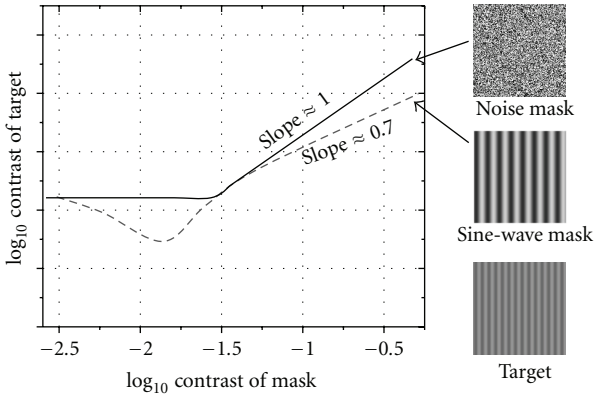


Figure 3: Demonstrative threshold-versus-contrast (TvC) curves for detection of a target consisting of a sine-wave grating in the presence of noise or sine-wave-grating masks. The horizontal axis denotes the contrast of the mask; the vertical axis denotes the contrast of the target (relative to the contrast threshold for detecting the target in the unmasked condition). Note that learning effects have been shown to lower the slopes of the TvC curves [90].

Further evidence in support of the multichannel model has been provided by visual adaptation and summation experiments [69, 82]. The CSF measured for a subject adapted to a sine-wave grating of a particular spatial frequency or orientation shows attenuation only within a narrow band of frequencies/orientations around the frequency/orientation of the grating (approximately 1-2 octaves, 15–30 degrees) [70, 82]. Visual summation experiments have revealed that a compound target (e.g., a plaid composed of two sine waves) is detectable only when one of its components reaches its own detection threshold, a finding which is consistent with a multichannel model with independent channels [82, 92–97] (the components of the compound target must be separated in spatial frequency by at least one octave or in orientation by at least 30°–45°; see [82]). Similar experiments have shown channels tuned to other dimensions such as color and direction of motion [69, 82].

The multichannel model has also been used to explain the shape of the CSF. Brady and Field [76] and Graham et al. [77] predicted the shape of the CSF via a model with equally sensitive spatial-frequency channels; reduction in detection performance for high spatial frequencies was attributed to extrinsic noise that dominates the response of channels tuned to high frequencies, thus resulting in decreased signal-to-noise ratios for these higher-frequency channels.

2.1.4. Computational Neural Models of V1. The multichannel model has inspired several related computational neural models of primary visual cortex (V1). These computational models have been used both to predict masking results and for IQA [87, 88, 98–101]. Models of this type first compute modeled neural responses to the reference image (mask), then compute modeled neural responses to the distorted image (mask + target), and then deem the distortions (target) detectable if the two sets of neural responses sufficiently differ. Quality can be estimated based on the predicted masked thresholds and/or the difference in simulated neural responses.

Figures 4 and 5 show block diagrams of the stages used in a typical computational neural model of V1 used to predict masked detection thresholds (Figure 4) or used to estimate quality (Figure 5). The pixel values of the reference and distorted images are first converted to either luminance or lightness values, and then both images are filtered with a 2D spatial filter designed to mimic the CSF. Alternatively, the CSF can be accounted for by scaling the coefficients of the frequency-based decomposition used to mimic the neural array. Next, two sets of simulated neural-array responses (one set for the reference image, one set for the distorted image) are computed via a filterbank. Further adjustments are made to account for neural nonlinearities and interactions (gain control) [99, 102–104]. The adjusted neural responses are then compared and collapsed across space, frequency, and orientation. The resulting threshold prediction or quality estimate is determined based on the comparison, that is, based on the extent to which the simulated neural responses to the reference image (mask) differ from the simulated neural responses to the distorted image (mask + target).

Frequency-Based Decomposition. To simulate an array of visual neurons in primary visual cortex (V1), and to account for the multichannel analysis performed by the HVS, the computational models employ some form of local frequency-based decomposition. Standard approaches to this decomposition include a steerable pyramid (e.g., [88]), a Gaussian pyramid (e.g., [105]), an overcomplete wavelet decomposition (e.g., [106]), radial filters (e.g., [107]), and cortex filters (e.g., [87, 99, 108]).

As shown in Figure 6, at a particular scale/orientation of this local frequency-based decomposition, the resulting matrix of transform coefficients represents the initially linear responses of a simulated array of neurons located at each spatial position in the image. Although real neurons cannot yield negative responses, negative coefficients are permitted and assumed to model the responses of co-located neurons that are tuned 180° out-of-phase (i.e., with an inhibitory central region and excitatory flanking regions).

The parameters of the decomposition are often tuned based on psychophysical and neurophysiological data (e.g., five or more radial frequency bands with 1-2 octave bandwidths, 4–12 orientations with 15°–30° bandwidths). In an IQA setting, the spatial-frequency decomposition is applied to both the reference image and the distorted image, yielding two sets of coefficients. The resulting coefficients are meant to simulate the initial linear responses of the neurons; they must be further adjusted to account for the neurons' nonlinear response properties.

Gain Control. Numerous studies have shown that the responses of neurons in V1 are nonlinearly related to the contrast of the stimulus to which the neurons are exposed (see [69, 109]). In the low-contrast regime, the neurons exhibit a threshold-type behavior in which a minimum contrast is required in order to yield any response. In the high-contrast regime, the neurons exhibit a saturation-type behavior in

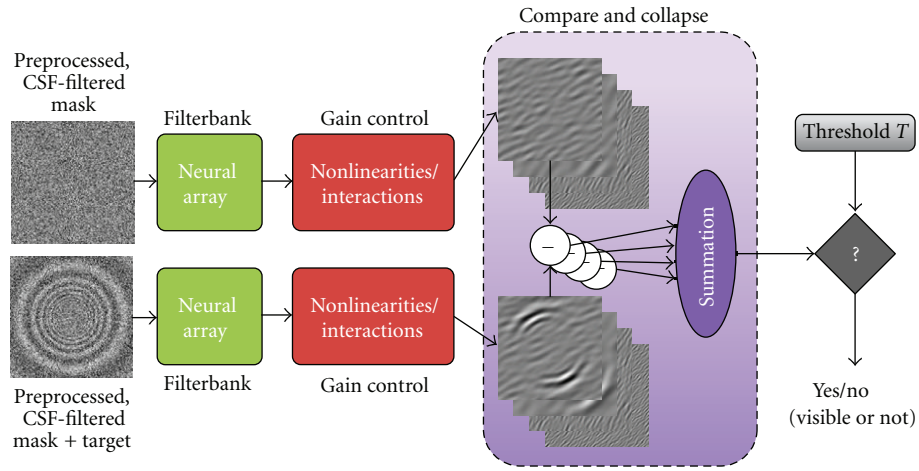


Figure 4: Block diagram of a typical computational neural model of V1 used to predict masked detection thresholds (here with a noise mask and a circular sine-wave target). Three key stages are employed in most neural models: (1) a frequency-based decomposition which models the initially linear responses of an array of visual neurons, (2) application of a pointwise nonlinearity to the decomposition coefficients and inhibition based on the values of other coefficients [99, 102–104], and (3) pointwise differences between the adjusted coefficients and summation of these adjusted coefficient differences across space, spatial frequency, and orientation so as to arrive at a single scalar differential response value or a map of differential response values.

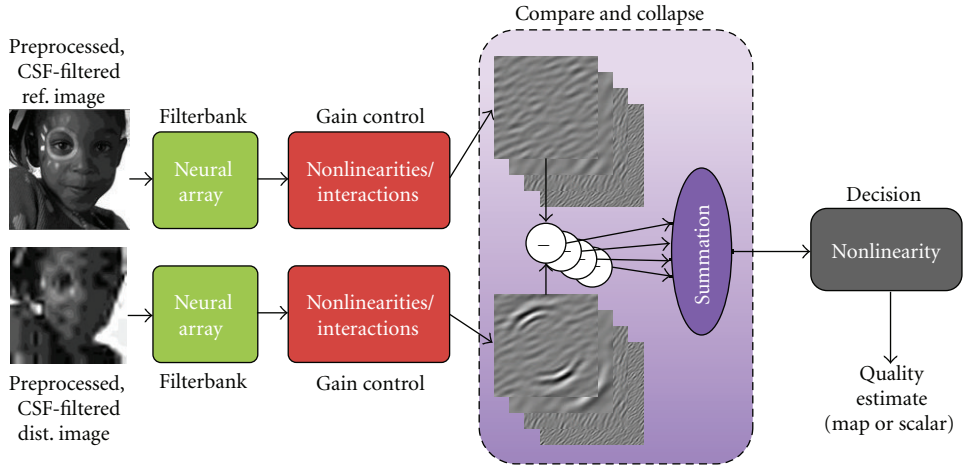


Figure 5: Block diagram of a typical computational neural model of V1 used to estimate quality. Three key stages are employed in most neural models: (1) a frequency-based decomposition which models the initially linear responses of an array of visual neurons, (2) application of a pointwise nonlinearity to the decomposition coefficients and inhibition based on the values of other coefficients [99, 102–104], and (3) pointwise differences between the adjusted coefficients and summation of these adjusted coefficient differences across space, spatial frequency, and orientation so as to arrive at a single scalar differential response value or a map of differential response values.

which further increases in contrast yield no corresponding increases in response. Studies have also shown that responses of V1 neurons can be inhibited by neighboring neurons in space, frequency, and orientation (the inhibitory pool). This inhibition from the neighboring neurons is commonly attributed to a *gain control* mechanism which is designed to keep the neuron operating in its linear regime and thus prevent saturation.

To account for these response properties, neural models apply a divisive normalization to the coefficients of the local frequency-based decomposition. Let $x(u_0, f_0, \theta_0)$ correspond to the coefficient at location u_0, center frequency f_0, and orientation θ_0. The (nonlinear) response of a neuron tuned to these parameters, $r(u_0, f_0, \theta_0)$, is most often simulated via

$$r(u_0, f_0, \theta_0) = g \cdot \frac{(w(f_0, \theta_0)\, x(u_0, f_0, \theta_0))^p}{b^q + \sum_{(u,f,\theta)\in S} (w(f,\theta)\, x(u, f, \theta))^q}, \quad (3)$$

where g is a gain factor, $w(f, \theta)$ represents an optional weight designed to take into account the CSF, b represents a saturation constant, p provides the pointwise nonlinearity to the current neuron, q provides the pointwise nonlinearity to the neurons in the inhibitory pool, and the set S indicates which other neurons are included in the inhibitory pool. The

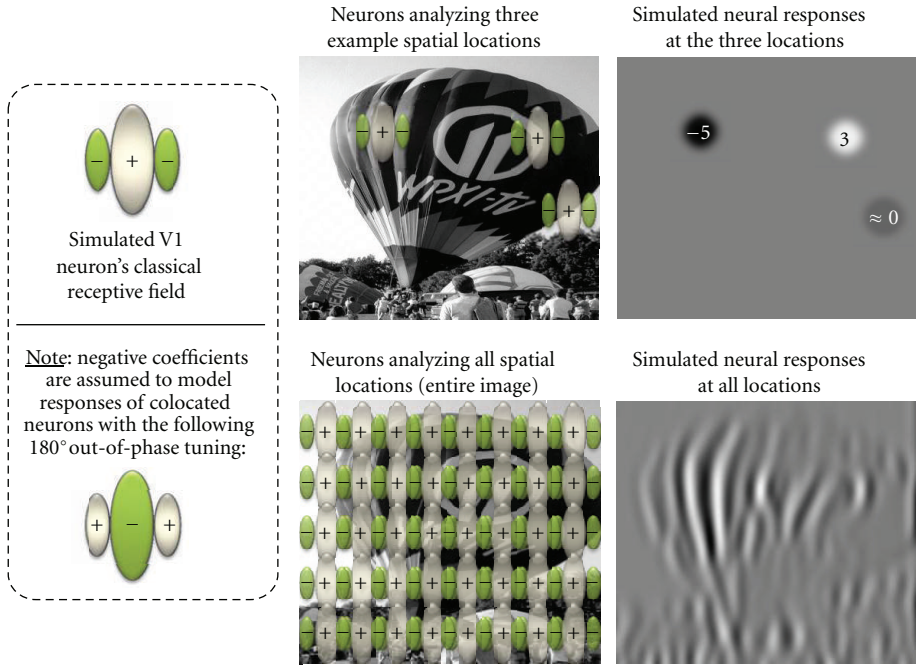


Figure 6: Demonstration of how spatial filtering (local frequency-based decomposition at a particular scale and orientation) mimics the initially linear responses of an array of similarly tuned neurons tiled across space. Note that actual V1 neurons cannot yield a negative response; however, negative values are still computed during the filtering and are assumed to model the responses of another array of neurons with a 180° of-out-phase tuning (i.e., with an inhibitory central region and excitatory flanking regions).

parameters b, p, q, and S are commonly adjusted to fit the experimental masking data. For example, model parameters have been optimized for detection thresholds measured using simple sinusoidal gratings [101], for filtered white noise [100], and for TvC curves of target Gabor patterns with sinusoidal masks [88, 99]. Typically, p and q are in the range $2 \leq q \leq p \leq 4$, and the inhibitory pool consists of neural responses in the same spatial-frequency band (f_0), at orientations within $\pm 45^\circ$ of θ_0 and within a local spatial neighborhood (e.g., 8-connected neighbors).

Equation (3) is applied to each coefficient of the decomposition of the reference image and to each coefficient of the decomposition of the distorted image. This operation results in two sets of simulated neural responses: (1) a set of neural responses to the reference image $\{r_{\text{ref}}(u, f, \theta)\}$ and (2) a set of neural responses to the distorted image $\{r_{\text{dst}}(u, f, \theta)\}$.

Summation of Responses. The final stage used in most V1 models entails comparing the two sets of simulated neural responses $\{r_{\text{ref}}(u, f, \theta)\}$ and $\{r_{\text{dst}}(u, f, \theta)\}$. When used as a masking model (Figure 4), to generate a map indicating the local visibility of the target, the responses at each location are compared and pooled across frequency and orientation as follows.

Distortions are visible at location u

$$= \begin{cases} \text{Yes}, & \left(\sum_{f,\theta} \left| r_{\text{ref}}(u, f, \theta) - r_{\text{dst}}(u, f, \theta) \right|^{\beta} \right)^{1/\beta} \geq T, \\ \text{No}, & \text{otherwise}, \end{cases} \quad (4)$$

where T is a predefined threshold which is typically held constant across images and where the summation exponent β is either chosen to match published results from summation studies or adjusted to fit published masking data. In an IQA setting (Figure 5), the comparison with T is often replaced with a sigmoid or logistic nonlinearity that maps the β-norm to an estimate of quality.

Numerous variations of (4) have been proposed in the literature; often the models are tuned to fit specific psychophysical data. The summation can also be applied across a local spatial neighborhood around u to determine a regional rather than a pointwise visibility. However, it is important to emphasize that the neural model is designed to mimic an array of visual neurons. This neurophysiological underpinning limits the choice of model parameters and operations to those which are biologically plausible.

In Section 3.1, I review several IQA algorithms which have employed variants of this V1-based model. It is also important to note that the vast majority of masking data have been obtained using simplistic, highly controlled targets (e.g., sine waves or Gabor patches) presented against unnatural masks (e.g., sine waves, Gabor patches, and noise). Consequently, most computational V1 models employ parameters which have been selected for such targets and masks. As I discuss later in Section 5.1, images can impose unique perceptual effects which cannot be fully captured by current V1 models.

2.2. Image Quality Databases. Another approach toward gaining insight into how humans judge quality is to directly collect quality ratings from a representative pool of human subjects on a database of altered images. Such ratings can also

be used to evaluate and refine IQA algorithms. *Image quality databases* provide this crucial ground-truth information. These databases typically contain a set of reference and altered images and average ratings of quality for each distorted image. The averages are generally taken across subjects, typically after z-score normalization and other adjustments (e.g., outlier tests) to attempt to account for individual biases; see [110]. The resulting averages are almost always reported in the form of mean opinion scores (MOS values) or differential mean opinion scores (DMOS values). For databases containing distorted images, a larger MOS (smaller DMOS) denotes greater quality, whereas a smaller MOS (larger DMOS) denotes lesser quality. Some databases further provide the standard deviations of the ratings across subjects.

Here, I first briefly summarize the existing publicly available image quality databases, and then I discuss techniques which are used to evaluate the performances of IQA algorithms on the databases.

2.2.1. List of Image Quality Databases. There are over 20 publicly available image quality databases, the details of which are described below and summarized in Table 1 (ordered by year of release). Many of these databases are listed as part of the extensive list of multimedia databases provided by the QUALINET consortium (European Network on Quality of Experience in Multimedia Systems and Services) [118]. Both Sheikh et al. [119] and Lin and Kuo [35] have provided analyses of the performances of various IQA algorithms on some of these databases. In addition, in [120], Winkler has provided quantitative comparisons of various aspects (source content, test conditions, and subjective ratings) of some of these databases. Note that 3D image quality databases are not listed here; see [118].

(i) *IRCCyN/IVC Image Quality Database (IVC).* The IRCCyN/IVC database [121, 122], developed at the Institut de Recherche en Communications et Cybernétique de Nantes (IRCCyN), France, contains 10 reference images and 185 distorted images in 24-bpp color BMP format at an image resolution of 512 × 512 pixels. There are five types of distortions in this database: JPEG compression (50 distorted images), JPEG compression of only the luminance component (25 distorted images), JPEG2000 compression (50 distorted images), locally adaptive-resolution coding (40 distorted images), and Gaussian blurring (20 distorted images). Each type of distortion was generated at five different amounts of distortion. The ratings were collected from 15 subjects.

(ii) *LIVE Image Quality Database.* The LIVE database [119, 123, 124], developed at the University of Texas at Austin, USA, contains 29 reference images and 779 distorted images in 24-bpp color BMP format at different image resolutions ranging from 634 × 438 to 768 × 512 pixels. There are five distortion types in this database: JPEG compression (169 distorted images), JPEG2000 compression (175 distorted images), additive Gaussian white noise (145 distorted images), Gaussian blurring (145 distorted images), and JPEG2000 with bit errors via a simulated Rayleigh fading channel (145 distorted images). Each type of distortion was generated at 5-6 different amounts of distortion. The ratings were collected from 29 subjects.

(iii) *A57 Image Quality Database.* The A57 database [125], developed at Cornell University, USA, contains three reference images and 54 distorted images in 8-bpp grayscale BMP format at a resolution of 512 × 512 pixels. The database contains six types of distortion: uniform quantization of the LH subbands of a 5-level discrete wavelet transform (DWT), additive Gaussian white noise, JPEG compression; JPEG2000 compression, custom JPEG2000 compression via the Dynamic Contrast-Based Quantization algorithm [126], and Gaussian blurring. Each type of distortion has three different amounts yielding nine distorted images for each distortion type. The ratings were collected from seven subjects.

(iv) *Tampere Image Quality (TID2008) Database.* The Tampere database [127, 128], developed at the Tampere University of Technology, Finland, contains 1700 distorted images generated from 25 reference images. The reference images were obtained from the Kodak Lossless True Color Image Suite. All of the images are stored in 24-bpp BMP format at a resolution of 384 × 512 pixels. There are 17 distortion types in the database (e.g., different types of noise, blur, denoising, JPEG and JPEG2000 compression, transmission of JPEG, JPEG2000 images with errors, local distortions, luminance, and contrast changes). Each type of distortion was generated at four different amounts. The ratings were obtained from 838 subjects.

(v) *Toyama Image Quality (MICT) Database.* The MICT database [129], developed at the University of Toyama, Japan, contains 14 reference images and 168 distorted images in 24-bpp color BMP format at a resolution of 768 × 512 pixels. There are two types of distortion in this database: JPEG compression (84 distorted images) and JPEG2000 compression (84 distorted images). Both types of distortion were generated at seven different amounts. The ratings were obtained from 16 subjects.

(vi) *IRCCyN/IVC Scores on the MICT Database.* Additional subjective ratings of the quality for the images from the MICT database were obtained at the Institut de Recherche en Communications et Cybernétique de Nantes (IRCCyN). The IRCCyN ratings were collected by using a different testing protocol, a different type of display, and different populations of subjects [130, 131]. The ratings were collected from 27 subjects.

(vii) *The Real Blur Image Database (RBID).* The RBID [111], developed at the Universidade Federal do Rio de Janeiro, Brazil, contains 585 blurred images in 24-bpp BMP format at resolutions ranging from 1280 × 960 to 2272 × 1704 pixels. The images in this database are

Table 1: Twenty-two publicly available image quality databases.

	Type	Year	No. of images	No. of reference	No. of distorted	No. of dist. types	Resolution	Format	No. of subjects
IRCCyN/IVC	Color	2005	195	10	185	5	512 × 512	BMP	15
LIVE	Color	2006	808	29	779	5	768 × 512	BMP	29
A57	Gray	2007	57	3	54	6	512 × 512	BMP	7
TID	Color	2008	1725	25	1700	17	384 × 512	BMP	838
MICT	Color	2008	196	14	168	2	768 × 512	BMP	16
IRCCyN/MICT	Color	2008	196	14	168	2	768 × 512	BMP	27
RBID	Color	2008	585	0	585	5	See [111]	BMP	20
WID/Enrico	Gray	2007	105	5	100	10	512 × 512	BMP	16
WID/BA	Gray	2009	130	10	120	2	512 × 512	PGM	17
WID/FSB	Gray	2009	215	5	210	6	512 × 512	BMP	7
WID/MW	Gray	2009	132	12	120	2	512 × 512	BMP	14
WIQ	Gray	2009	87	7	80	1	512 × 512	BMP	30
VAIQ	Color	2009	42	42	See [112]	See [112]	See [112]	BMP	15
CSIQ	Color	2010	896	30	866	6	512 × 512	PNG	35
TUD1	Color	2010	24	8	16	1	See [113]	BMP	12
TUD2	Color	2010	55	11	44	1	See [113]	BMP	20
IRCCyN/DIBR	Color	2011	96	3	96	3	1024 × 768	See [114]	43
JPEGXR	Color	2011	70	10	60	1	1280 × 1600	BMP	See [115]
HTI	Color	2011	72	12	60	1	512 × 768	See [116]	18
IBBI	Color	2011	72	12	60	1	321 × 481	See [116]	18
VCL@FER	Color	2011	575	23	552	4	See [117]	See [117]	118
DRIQ	Color	2012	104	26	78 enhanced	3 enhancements	512 × 512	PNG	9

categorized into five different blur classes: unblurred (204 images), out of focus (142 images), simple motion (57 images), complex motion (63 images), and others (119 images). The ratings were collected from 20 subjects.

(viii) *IRCCyN/IVC Watermarking Databases.* Four separate watermarking databases were developed by the Institut de Recherche en Communications et Cybernétique de Nantes (IRCCyN), France. The images were created by embedding watermarks with different algorithms: Enrico, Broken Arrows (BA), Fourier Subband (FSB), and Meerwald (MW).

(a) *IRCCyN/IVC Watermarking—Enrico Database.* This database [132] contains five reference images and 100 distorted images generated from 10 watermarking algorithms with two embedding strengths. All of the images are in 8-bpp grayscale BMP format at a resolution of 512 × 512 pixels. The ratings were obtained from 16 subjects.

(b) *IRCCyN/IVC Watermarking—Broken Arrows Database.* This database [133] contains 10 reference images and 120 distorted images with six different embedding strengths and either with or without CSF weighting. All of the images are in 8-bpp grayscale PPM/PGM format at a resolution of 512 × 512 pixels. The ratings were obtained from 17 subjects.

(c) *IRCCyN/IVC Watermarking—Fourier Subband Database.* This database [134] contains five reference images and 210 distorted images containing watermarks in six frequency subbands at seven embedding strengths for each subband. All of the images are in 8-bpp grayscale BMP format at a resolution of 512 × 512 pixels. The ratings were obtained from 7 subjects.

(d) *IRCCyN/IVC Watermarking—Meerwald Database.* This database [135] contains 12 reference images and 120 distorted images generated from five embedding strengths either in the DWT domain or in the dual-tree complex wavelet transform domain. All of the images are in 8-bpp grayscale BMP format at a resolution of 512×512 pixels. The ratings were obtained from 14 subjects.

(ix) *Wireless Imaging Quality (WIQ) Database.* The WIQ database [136, 137] was developed at the Radio Communication Group at the Blekinge Institute of Technology, Sweden. This database contains seven reference images and 80 images distorted via loss of JPEG data over a simulated wireless channel. The images are stored in 8-bpp BMP format at a resolution of 512 × 512 pixels. The ratings were obtained in two separate experiments from 30 subjects.

(x) *The Visual Attention Image Quality (VAIQ) Database.* The VAIQ database [112, 138], developed at the University of Western Sydney, Australia, contains

ground-truth visual gaze patterns of 42 reference images taken from the LIVE, IVC, and MICT image databases. Although this database is not strictly an image quality database, the visual gaze patterns can be useful for examining the effects of visual attention on quality. The visual gaze patterns were obtained from 15 subjects.

(xi) *Categorical Subjective Image Quality (CSIQ) Database.* The CSIQ database [27, 139], developed at Oklahoma State University, USA, contains 30 reference images and 866 distorted images in 24-bpp PNG format at a resolution of 512 × 512 pixels. There are six distortion types in this database: JPEG compression (150 distorted images), JPEG2000 compression (150 distorted images), additive Gaussian white noise (150 distorted images), additive Gaussian pink noise (150 distorted images), Gaussian blurring (150 distorted images), and global contrast decrements (116 distorted images). Each type of distortion was generated at 4-5 different amounts. The ratings were obtained from 35 subjects.

(xii) *TU Delft Perceived Ringing (TUD1 and TUD2) Datasets.* The TUD1 and TUD2 databases were developed at the Delft University of Technology, The Netherlands. The subjective ratings were collected from two experiments: (1) a ringing region experiment (TUD1) and (2) a ringing annoyance experiment (TUD2). In the ringing region experiment, 16 JPEG-compressed images were generated from eight reference images with two levels of compression. The results were collected from 12 subjects and are presented in the form of subjective ringing region maps. In the ringing annoyance experiment; 44 JPEG-compressed images were generated from 11 reference images with four different levels of compression. The ratings were obtained from 20 subjects.

(xiii) *IRCCyN/IVC DIBR Image Quality Database.* The DIBR database [114], developed by the Institut de Recherche en Communications et Cybernétique de Nantes (IRCCyN), France, contains 96 still images extracted from three different multiview-plus-depth sequences. All sequences have the same resolution of 1024 × 768 pixels but were captured with a variable number of cameras (16, 12, and 9) at different camera spacings (6.5 cm, 3.5 cm, and 5.0 cm). Each sequence was processed by seven depth-image-based rendering algorithms to generate four new viewpoints of each sequence. The ratings were obtained from 43 subjects.

(xiv) *MMSPG JPEG XR Image Compression Database.* The MMSPG JPEG XR image compression database [115], developed at the Swiss Federal Institute of Technology (EPFL), Switzerland, contains 60 compressed images generated from 10 reference images using JPEG XR compression. The images are stored in 24-bpp color BMP format at a resolution of 1280 × 1600 pixels. Six coding bitrates ranging from 0.25 to 1.50 bpp were used to generate the distorted images. The ratings were obtained from 16 subjects.

(xv) *HTI and IBBI Databases.* The HTI and IBBI Databases [116], jointly developed at TU Delft and IRCCyN, are designed to test the performances of the blurriness metrics. The highly textured images (HTI) database contains 12 reference images containing highly textured content and 60 blurred versions of the images; the images are stored at a resolution of 512×768 pixels. The intentionally blurred background images (IBBI) database contains 12 reference images and 60 blurred versions in which the background was intentionally blurred; the images are stored at a resolution of 321 × 481 pixels. In both databases, the blurring was performed via Gaussian filtering with five different levels of blur, and the ratings were obtained from 18 subjects.

(xvi) *VCL@FER Image Quality Database.* The VCL@FER database [117], developed at the University of Zagreb, Croatia, contains 23 reference images and 552 distorted images. There are four types of distortion in this database (148 distorted images for each type): additive Gaussian white noise, Gaussian blurring, JPEG compression, and JPEG2000 compression. Each distortion type was generated at six different amounts of distortion. The ratings were obtained from 118 subjects.

(xvii) *Digitally Retouched Image Quality (DRIQ) Database.* The DRIQ image quality database [140] is a full-reference enhanced-image database developed at Oklahoma State University, USA. This database contains 26 reference images and 78 enhanced images obtained via manual digital retouching. The images are stored in 24-bpp color PNG format. The ratings were obtained from 9 subjects. (Some images from DRIQ and additional details of the database are provided in Section 5.6.)

2.2.2. Quantifying the Predictive Performance. The image quality databases described in the previous section serve as crucial ground-truth information for evaluating IQA algorithms. Specifically, to quantify how well an IQA algorithm can predict the MOS or DMOS values from a particular database, it is customary to evaluate the algorithm in terms of three performance criteria recommended by the Video Quality Experts Group (VQEG) [141]: (1) prediction accuracy, (2) prediction monotonicity, and (3) prediction consistency.

The *prediction accuracy* can be quantified either by measuring how well an algorithm's predictions correlate with the MOS/DMOS values or by measuring the average error between the algorithm's predictions and the MOS/DMOS values. The Pearson correlation coefficient (CC) and the root-mean-squared error (RMSE), both recommended in [141], are most commonly used for quantifying correlation and average error, respectively.

Before computing CC or RMSE, it is customary to apply a nonlinear transformation to the predicted scores so as to bring the predictions on the same scale as the MOS/DMOS

values and to attempt to obtain a linear relationship between the predictions and opinion scores. Let y_i denote the DMOS or MOS value for the ith image, let x_i denote the corresponding predicted score from an IQA algorithm, and let $f(x_i)$ denote the corresponding transformed predicted score. In [141], three suggestions for the transform $f(x)$ are provided:

$$f(x) = \frac{\tau_1 - \tau_2}{1 + \exp\left((x - \tau_3)/\tau_4\right)} + \tau_5,$$

$$f(x) = \tau_1 x^3 + \tau_2 x^2 + \tau_3 x + \tau_4,$$

$$f(x) = g^{-1}(y), \quad \text{where } g(y) = \tau_1 y^3 + \tau_2 y^2 + \tau_3 y + \tau_4. \tag{5}$$

The parameters $\tau_1, \tau_2, \tau_3, \tau_4$, and τ_5 are chosen to minimize the MSE between the set of DMOS/MOS values $\{y_i\}$ (e.g., all DMOS/MOS values in a particular database) and the corresponding set of transformed predicted values $\{f(x_i)\}$. The minimization is conducted under the constraint that $f(x)$ must be a monotonic function of x over the range of predicted values.

It is important to note that because there is inherent variability across subjects and across different trials of the same subject, if the scores for a particular image demonstrate a large variability across subjects/trials, then the *mean* score (i.e., the MOS or DMOS) is not necessarily a good indication of what should be predicted. Instead, some leeway, determined based on the variability, should be given around the MOS/DMOS values. Some databases provide the standard deviation associated with each MOS/DMOS value, σ_{y_i}, which provides a measure of the variability across subjects/trials for the ith image. These standard deviations can be taken into account during the fitting procedure to determine the parameters $\tau_1, \tau_2, \tau_3, \tau_4$, and τ_5.

The *prediction monotonicity* specifies how well an algorithm predicts the rank-ordering of the opinion scores. Various rank-order correlation coefficients can be used to quantify the monotonicity (e.g., Spearman's ρ, Kendall's τ_a, Kendall's τ_b, Goodman and Kruskal's γ, and Somer's D; see [142]). The Spearman rank-order correlation coefficient (Spearman's ρ, SROCC) is recommended in [141] and is thus most commonly used.

However, it should be noted that the standard formula for computing SROCC must be adjusted for ties in the ranks [142], an adjustment which is rarely used in the IQA literature. The difficulty in accounting for ties stems from the fact that in order to determine ties, the variability (e.g., σ_{y_i}) associated with each score must be known. As I mentioned, not all databases provide such information, and thus ties are rarely taken into account.

The *prediction consistency* specifies how consistent is an IQA algorithm's quality predictions across the range of content provided in a database—for example, for different images, different distortion types (or other alteration types), and different amounts of each distortion/alteration type. Two measures of prediction consistency are the outlier ratio [141] and the outlier distance [139], both of which require the aforementioned standard deviations.

The most commonly used measure of prediction consistency, which is recommended in [141], is the outlier ratio. The outlier ratio, R_{out}, is defined as

$$R_{\text{out}} = \frac{N_{\text{false}}}{N_{\text{total}}}, \tag{6}$$

where N_{false} is the number of predictions outside two standard deviations, $2\sigma_{y_i}$, and N_{total} is the total number of scores. The range of $2\sigma_{y_i}$ was chosen in [141] because it contains 95% of all the subjective quality scores for a given image.

In addition to knowing if a predicted score is an outlier, it is also informative to know how far outside the error bars ($\pm 2\sigma_{y_i}$) the outlier falls. To quantify this, we proposed in [139] a new measure, termed the *outlier distance*. The outlier distance, d_{out}, is the distance from an outlier to the closest error bar; it is defined as

$$d_{\text{out}} = \sum_{i \in X_{\text{false}}} \min\left\{\left|f(x_i) - \left[y_i + 2\sigma_{y_i}\right]\right|, \left|f(x_i) - \left[y_i - 2\sigma_{y_i}\right]\right|\right\}, \tag{7}$$

where X_{false} is the set of all predicted scores outside $2\sigma_{y_i}$. Note that because d_{out} is dependent on the dynamic range of the MOS/DMOS values, it cannot be used to compare across databases.

3. Full-Reference Image Quality Assessment Algorithms

The vast majority of IQA algorithms are so-called *full-reference* algorithms, which take as input both a distorted image and a reference image and yield as output an estimate of the quality of the distorted image relative to the reference. The simplest approach to full-reference (FR) IQA is to measure local pixelwise differences and then to collapse these local measurements into a scalar which represents the overall quality difference, for example, the mean-squared error (MSE) or peak signal-to-noise ratio (PSNR), often measured in different domains, for example, [143]. More complete FR IQA algorithms have employed a wide variety of approaches ranging from estimating quality based on models of the HVS (see Section 3.1), to estimating quality based on image structure (see Section 3.2), to estimating quality by using various statistical and information-theoretic-based approaches (see Section 3.3) and many other techniques (see Section 3.4). Here, I provide a brief survey of these FR IQA algorithms.

3.1. Methods Based on HVS Models. Given a distorted image, a human can readily rate the quality of the image relative to the original image and relative to other distorted images. Accordingly, numerous IQA methods have been developed which employ computational models of the HVS [26, 30, 87, 88, 98, 100, 101, 105–107, 125, 130, 139, 144–157].

Most HVS-based IQA algorithms employ a variant of the V1 model described previously in Section 2.1. The images are typically processed through a set of spatial filters to obtain

oriented, spatial-frequency decompositions of the images designed to mimic the initially linear responses of neurons in V1. The CSF is taken into account either by adjusting the simulated linear neural responses based on the passbands of the filters or by using a prefiltering stage with a 2D CSF-based filter (see (1)). Masking is commonly taken into account by further adjusting the simulated neural responses via a divisive normalization (see (3)). Finally, the quality of the distorted image is estimated based on the extent to which the adjusted responses to the reference image differ from the adjusted responses to the distorted image. Typically, this final stage is performed by computing pointwise absolute differences between the original and distorted responses, and then collapsing these differences via an L_p norm (see, e.g., [87, 88, 99]).

Many HVS-based methods were originally designed to operate as predictors of visible image differences; that is, they have been designed to determine if changes are visible and accordingly operate best when the distorted images contain artifacts near the threshold of detection. Researchers have previously argued that the underlying V1 models need to be extended to take into account higher-level properties of human vision [85, 126, 156]. Unfortunately, although our current understanding of near-threshold vision for controlled stimuli is relatively mature from a modeling perspective, much less is known about how the HVS operates when the distortions are more complex and in the suprathreshold regime (which may invoke areas of visual cortex beyond V1). Nonetheless, recent HVS-based methods have begun to use improved models and/or models of mid- and higher-level vision [107, 125, 139, 151, 157, 158], and many of these methods have been shown to perform extremely well as general IQA algorithms.

For example, in [107], Damera-Venkata et al. augmented traditional models of contrast sensitivity and luminance and contrast masking with models of suprathreshold contrast perception. In [125], Chandler and Hemami presented a visual signal-to-noise ratio (VSNR), in which a wavelet-based model of low-level vision is combined with a model of how the HVS adaptively prefers different spatial frequencies depending on the amount of degradation. In [157], Laparra et al. presented an IQA algorithm which employs an improved divisive-normalization-based masking model. In [159], Cheng et al. supplemented a wavelet-based HVS model with measures of directional structural distortion and structural similarity [123]. Alternative image transforms have also been used to integrate properties of the HVS into IQA algorithms [160, 161].

In [139], Larson and Chandler presented an IQA algorithm, MAD (most apparent distortion), which explicitly models the adaptive nature of the HVS. MAD was one of the first algorithms to demonstrate that quality can be predicted by modeling two strategies employed by the HVS and by adapting these strategies based on the amount of distortion. For high-quality images, in which the distortion is less noticeable, the image is most apparent, and thus the HVS attempts to look past the image and look for the distortion—a detection-based strategy. For low-quality images, the distortion is most apparent, and thus the HVS attempts to look past the distortion and look for the image's subject matter—an appearance-based strategy. In MAD, local luminance and contrast masking are used to model the detection-based strategy for high-quality images, whereas changes in the local statistics of log-Gabor coefficients are used to model the appearance-based strategy for low-quality images.

Another recent trend in HVS-based IQA algorithms has aimed at incorporating aspects of visual attention and regions of interest (ROIs) during quality assessment (see, e.g., [60–62, 65, 162–164] for related psychophysical studies). Wang and Bovik [165] developed a model for adjusting contrast sensitivity based on foveation. Osberger et al. [152] developed a vision-based metric by using the CSF and masking, into which they incorporated a visual importance map by selecting effective ROIs based on higher-level visual properties such as size, shape, foreground/background, and the presence of people.

Le Callet et al. [155] presented an IQA algorithm which operates based on ten 1×1-degree areas containing the maximum perceived distortion, the implicit ROI assumes that the eye is drawn to regions of maximum error, and quality is estimated by spatially pooling the estimates of maximum perceived distortion. In [158], Carnec et al. combined low-level HVS properties with a measure of the structural information obtained via a stick-growing algorithm and estimates of visual fixation points. In [166], Moorthy and Bovik incorporated both visual-fixation-based weighting and quality-based weighting into an IQA algorithm.

In [167], Tong et al. employed saliency maps for IQA based on the observation that salient regions contribute more to the perceived image quality. Salient region information generated by the model of Itti and Koch [168] and a face detection model were used to generate weights in [167] to improve the performances of previous IQA algorithms. Similarly, in [169], Guo et al. assumed that humans often pay more attention to the image regions with important content. The authors incorporated saliency-based visual attention and visual-importance-based visual attention into the SSIM algorithm [123].

In [170], Wu et al. incorporated an "internal generative mechanism" (IGM) into the existing IQA algorithms. Their IGM advocates that the HVS actively predicts sensory information and tries to avoid the residual uncertainty for image perception and understanding. In [170], the images are decomposed into two parts: the predicted portion, consisting of the predicted visual content, and the disorderly portion, consisting of the residual content. SSIM [123] is employed to measure distortions in the predicted portion, PSNR is employed to measure distortions in the disorderly portion, and then the two results are adaptively combined to yield the overall quality prediction.

3.2. Methods Based on Image Structure. A recent thrust in image quality assessment has focused on measuring changes in an image's structure as a proxy for measuring image quality. The central assumption in this approach is that the HVS has evolved to extract structure from the natural environment.

Consequently, a higher-quality image is one whose structure closely matches that of the original image, whereas a lower-quality image exhibits less structural similarity to the original. Although a precise definition of "image structure" remains an open question, methods of this type have been shown to correlate quite highly with subjective ratings of quality.

The effects of distortion on image structure and the corresponding effects on image quality have been mentioned in the optics and engineering literature since the 1970s in the context of television pictures (e.g., [25, 171]). Eskicioglu and Fisher [172] were among the first to apply explicit measures of "structural content" and "correlation quality" (based on correlation and normalized cross-correlation) to IQA. In [173], Fränti presented a block-based IQA algorithm which included separate measures for structural errors (based on edge detection), quantization errors, and contrast errors. In [174], Wang and Bovik presented the universal image quality index (UQI), which employed cross-correlation and measures of luminance and contrast differences to estimate quality.

The use of cross-correlation-based measures of structural similarity was made popular by Wang et al. who proposed the Structural Similarity Index (SSIM) [123]. SSIM is an extended version of UQI in which the correlation, luminance, and contrast measures were modified by adding small constants to the numerator and the denominator of each measure. In [175], Wang et al. presented a multiscale version of SSIM in which the correlation, luminance, and contrast measures are also applied to filtered and downsampled versions of the images (MS-SSIM). In [176], Sampat et al. presented a complex version of SSIM which adds robustness to small affine transformations of the distorted image.

Over the last eight years, numerous variations of SSIM and other IQA algorithms which estimate quality based on structural similarity and/or structural degradation have been proposed.

3.2.1. SSIM-Based Methods. Various IQA algorithms have been developed which directly employ SSIM or MS-SSIM as a part of the IQA process.

(i) In [177], Yang et al. presented a modified version of MS-SSIM that operates by using the 9/7 DWT filters.

(ii) In [178], Ji et al. presented a modified version of SSIM using a discrete Haar wavelet transform (HWSSIM). A multiresolution version of HWSSIM was also defined in [178] by weighting/combining four HWSSIM values evaluated at four Haar wavelet levels with the CSF.

(iii) In [179], Cao et al. presented an IQA method which estimates quality based on the influence of both global and local distortions. The global distortion is measured via a rectified mean absolute difference; the local distortion is measured via SSIM. These two measures are combined using a weighting strategy to yield a final image quality estimate.

(iv) In [180], Shi et al. presented an IQA method for color images based on structural and color similarity indices. SSIM is applied to the luminance and the hue component of the image to yield two similarity indices which are combined into an overall structural similarity index.

(v) In [181], Rao and Reddy presented an IQA method in which the SSIM indices of local image regions are adjusted by perceptual weights defined from the regions' contrasts. A measure of the image's overall perceptual SSIM index (PSSIM_c) is calculated as the average of these weighted indices.

(vi) In [182], Zhang et al. employed image features from the coefficients of the 1st-order and 2nd-order Riesz transform. These two feature maps are masked by the image's edge locations before applying SSIM to compute the structural similarity.

(vii) As mentioned in Section 3.1, in [169], Guo et al. presented an IQA algorithm which incorporates the saliency-based the visual attention and visual-importance-based visual attention into SSIM.

(viii) As also mentioned in Section 3.1, Wu et al. [170] employ SSIM [123] to measure distortions in the predicted portion of their internal-generative-mechanism-based IQA model.

(ix) In [183], Chebbi et al. estimate quality of blurred images based on a combination of a perceptual blur measure and SSIM; the perceptual blur measure is derived from edge maps obtained via a Haar DWT.

(x) In [184], Fei et al. estimate the quality based on SSIM and visual masking. The contrast comparison in SSIM is augmented by measures of masking, and the structural comparison is modified by using the image's structure tensor.

3.2.2. Gradient-Based Methods. Another way to measure changes in structure is to compute changes in local image gradients. Several methods have been developed which take this approach.

(i) In [185], Kim and Park presented an IQA algorithm based on the Harris response. The authors observed that when an image is degraded or distorted, the image's gradient information is changed, causing the Harris response to change. Thus, in [185], the changes in the Harris response of the image, which is computed from the gradient information matrix and its eigenvalues, are used to measure image quality.

(ii) In [186], Zhu and Wang presented an IQA algorithm based on a three-stage multiscale visual gradient similarity (VGS) index. First, global contrast registration is applied for each scale. Second, the similarity of gradient directions and gradient magnitudes are combined to yield comparison maps. Finally, quality is estimated via intrascale and interscale pooling of the maps. Some parameters of VGS are trained on existing image quality databases to optimize performance.

(iii) In [187], Chen et al. argued that the structural information can be computed from the distribution of gradient magnitudes and edge directions. Accordingly, edge-direction-histogram (EDH) descriptors are extracted, and then quality is estimated based on the structural similarity between EDH descriptors of the reference and the distorted images.

(iv) In [188], Liu et al. compared gradient similarity between the reference and distorted images to evaluate quality. Their method first computes both luminance-based and contrast-based structural changes. Then, these changes are weighted by estimates of visibility thresholds. Finally, the weighted changes are adaptively integrated to yield the image's overall quality estimate.

3.2.3. Methods Based on Other Measures of Structure. Various alternative measures of structural similarity/- degradation have also been employed in several IQA algorithms.

(i) In [189], Zhai et al. presented an IQA algorithm which operates based on the notion of a "multiscale edge presentation." Their method measures structure via correspondences in wavelet magnitudes across spatial scales.

(ii) In [190], Zhang and Mou combined PSNR with a measure of structure based on differences in wavelet modulus maxima corresponding to low- and high-frequency bands.

(iii) In [191], Jin et al. presented a DCT-based IQA algorithm that considers contrast and brightness degradations as well as block-based structural similarity.

(iv) In [192], Chou and Hsu proposed an IQA algorithm which uses moment-preserving quantization for extracting geometric structural information. SSIM values of luminance and contrast are combined with a similarity measure of geometric structure to yield the final quality estimate.

(v) In [193], Zhang et al. proposed a feature-based similarity measure for IQA which operates based on phase congruency as the primary feature and the image gradient magnitude as a secondary feature. The phase congruency is also used as a weighting function to derive an overall image quality score from the local quality map obtained from the two features.

(vi) In [194], Narwaria et al. designed an IQA algorithm based on the phase and magnitude of the discrete Fourier transform. The algorithm compares the phase and magnitude of the Fourier coefficients of the reference and distorted images to compute image quality. Nonuniform binning of the frequency components and linear regression are employed to integrate the effects of the changes in phase and magnitude.

3.3. Methods Based on Image Statistics and/or Machine Learning. Other measures of image quality have been proposed which operate based primarily on statistical/information-theoretic measures, often supplemented by machine learning techniques. The methods have also demonstrated great success at predicting quality. See [36] for a recent thorough discussion of the use of natural-scene statistics for IQA.

In [195], Sheikh and Bovik presented the VIF (visual information fidelity) algorithm, which estimates image quality based on natural-scene statistics. VIF operates under the premise that the HVS has evolved based on the statistical properties of the natural environment. Accordingly, the quality of the distorted image can be quantified based on the amount of information the distorted image provides about the reference image. VIF models images as realizations of a mixture of marginal Gaussian densities of wavelet subbands, and quality is then determined based on the mutual information between the subband coefficients of the reference and distorted images.

In [196], Shnayderman et al. measured image quality based on a singular value decomposition (SVD). The Euclidean distance is measured between the singular values of an original image block and the singular values of the corresponding distorted image block; the collection of block-wise distances constitutes a local distortion map. An overall scalar value of image quality is computed as the average absolute difference between each block's distance and the median distance over all blocks.

Other SVD-based IQA algorithms have also been proposed. In [197], Mansouri et al. presented an IQA algorithm (RSVD) which estimates quality by factoring the image's SVD matrix into a matrix which captures luminance changes and two matrices which capture structural changes. In [198], Narwaria and Lin presented an IQA algorithm which uses SVD-based visual features and feature pooling via machine learning. In [199], Saha et al. presented an IQA algorithm in which approximate descriptions of the reference and distorted images are obtained at different scales by using an SVD filter; quality is estimated by computing the similarity between two pyramidal structures.

More direct machine-learning-based techniques have also been applied to IQA. In [7], Liu and Yang applied supervised learning to derive a measure of image quality based on decision fusion. A training step is used to determine an optimal linear combination of four IQA methods: PSNR, SSIM, VIF, and VSNR. Training is performed via canonical correlation analysis and images/subjective ratings from the LIVE [124] and A57 [125] image databases.

In [200], Peng and Li argued that IQA algorithms which operate based on individual features cannot accurately predict quality across different distortion types. To overcome this limitation, the authors proposed a two-stage scheme. First, the image distortion type is predicted by support-vector classifiers. Second, decision-level fusion of three existing algorithms (SSIM, VSNR, and VIF) is performed based on the k-nearest-neighbor regression where the acquired distortion-type knowledge is employed. More recent related work by Peng and Li can be found in [201].

In [202], Charrier et al. presented the Machine Learning-Based Image Quality Measure (MLIQM) which employs a learned classification process. The MLIQM method first

constructs a feature vector consisting of various measured image attributes. Then, a classification process is performed to assign the distorted image into a quality class. Finally, support-vector regression is performed based on the quality class to yield the estimate of quality.

Other statistical models of images have also been used for IQA. In [203], Wang and Li argued that the optimal perceptual weights for pooling the output of an IQA algorithm across space should be proportional to the local information content, which is estimated in [203] by using statistical models of natural scenes. In [204], Chang and Wang presented an IQA algorithm that relates image quality with the correlation between the sparse codes formed from the reference and distorted images. In [205], Pinto and Hemami provided an upper bound for the performances of IQA algorithms in the low-quality regime by using a family of IQA techniques based on VIF.

3.4. Methods Based on Other Techniques. In addition to the methods mentioned in the previous sections, numerous other techniques have been employed in IQA algorithms. For example, IQA algorithms have been developed to assess quality based on different color spaces [206], based on image segmentation and/or region-based analysis [207–209] and based on the use of additional features [210–217].

In [207], Xu and Hauske presented an IQA algorithm which operates based on error segmentation. The errors are divided into three types: those that affect objects' edges, those that affect other edges, and those that are most visible in smooth regions (e.g., blocking and noise). The segmented errors are used to compute distortion factors, which are combined into an overall estimate of quality via multiple linear regression.

In [210], Bianco et al. presented different computational strategies to improve the robustness and accuracy of SSIM and the S-CIELAB spatial-color model.

In [211], Okarma proposed an IQA algorithm which employs a combination of three previous methods: MS-SSIM [175], VIF [195], and RSVD [197].

In [212], Lahouhou et al. conducted an empirical study of several quality indicators (including PSNR, SSIM, and wavelet-based quality measures) and proposed a regularized regression model for combining the indicators to predict quality.

In [213], Xue and Mou proposed an IQA algorithm based on a ratio of non shift edges. First, the distorted image is filtered by a Laplacian of Gaussian, and then edge points are detected from the filtered image. Next, a binary "non-shift edge" map is computed to represent the strong edge structure present in the distorted image. Finally, quality is estimated based on the map.

In [214], Li et al. demonstrated that adaptively combining two quality measurements could improve quality predictions. The authors proposed an IQA algorithm that separately evaluates detail losses and additive impairments. The detail loss refers to the loss of useful visual information which affects the content visibility, and the additive impairment represents the redundant visual information which distracts attention from the useful content. Two quality measures corresponding to detail losses and additive impairments are computed, and then the outputs of the two quality measures are adaptively combined to yield the overall quality prediction.

In [215], Attar et al. presented the edge-based image quality assessment (EBIQA) algorithm. EBIQA employs four edge features computed from the reference and distorted images: edge orientation, average length of edges, primitive length of edges, and number of edge pixels.

In [216], Ponomarenko et al. presented an IQA algorithm which employs a parameter map that denotes the image's local self-similarity. Quality is estimated based on the mean-squared difference between the parameter maps for the reference and distorted images.

In [217], Solh and AlRegib developed an IQA method for multicamera systems by identifying and quantifying two types of visual distortions: photometric distortions and geometric distortions. Such distortions can be quantified by using three different indices: a luminance and contrast index, a spatial motion index, and an edge-based structure index. These indices are combined into one multicamera image quality measure (MIQM).

4. No-Reference and Reduced-Reference Image Quality Assessment

Although FR IQA provides a useful and effective way to evaluate quality differences, in many applications the reference image is not available. Although humans can often effortlessly judge the quality of a distorted image in the absence of a reference image, this task has proven to be quite challenging from a computational perspective. *No-reference* (NR) and *reduced-reference* (RR) IQA algorithms attempt to perform IQA with either no information (NR IQA) or only limited information (RR IQA) about the reference image. Here, I briefly survey existing NR and RR IQA algorithms.

4.1. No-Reference IQA. The vast majority of NR IQA algorithms attempt to detect specific types of distortion such as blurring, blocking, ringing, or various forms of noise. For example, algorithms for sharpness/blurriness estimation have been shown to perform well for NR IQA of blurred images. NR IQA algorithms have also been designed specifically for JPEG or JPEG2000 compression artifacts. Some NR algorithms have employed combinations of these aforementioned measures and/or other measures. Other NR IQA algorithms have taken a more distortion-agnostic approach.

4.1.1. Methods for Blurriness/Sharpness. Numerous algorithms have been developed to estimate the perceived sharpness or blurriness of images. Although the majority of these algorithms were not designed specifically for NR IQA, they have shown success at IQA for blurred images. Modern methods of sharpness/blurriness estimation generally fall into one of four categories: (1) those which operate via edge-appearance models, (2) those which operate in the spatial domain without any assumptions regarding edges, (3) those which operate by using transform-based methods, and

(4) hybrid techniques which employ two or more of these methods.

A common technique of sharpness/blurriness estimation involves the use of edge-appearance models. Methods of this type operate under the assumption that the appearance of edges is affected by blur, and accordingly these methods estimate sharpness/blurriness by extracting various properties of the image edges. For example, Marziliano et al. [218] estimate blurriness based on average edge widths. Ong et al. [219] estimate blurriness based on edge widths in both the edge direction and its gradient direction. Dijk et al. [220] model the widths and amplitudes of lines and edges as Gaussian profiles and then estimate sharpness based on the amplitudes corresponding to the narrowest profiles. Chung et al. [221] estimate sharpness based on a combination of the standard deviation and weighted mean of the edge gradient magnitude profile. Wu et al. [222] estimate blurriness based on the image estimated point spread function. Zhong et al. [223] estimate sharpness based on both edges and information from a saliency map. Ferzli and Karam [224] estimate sharpness based on an HVS-based model which predicts thresholds for just noticeable blur (JNB) the JNB for each edge block is used to estimate the block perceived blur distortions, and the final sharpness estimate is based on a probabilistic combination of these distortions. A related JNB-based method can be found in [225].

Other sharpness/blurriness estimators work in the spatial domain but do not attempt to locate edges. Wee and Paramesran [226] estimate sharpness based on the dominant eigenvalues of the covariance matrix of the image pixels. Zhu and Milanfar [227] estimate sharpness based on the SVD of the local image-gradient matrix. Roffet et al. [228] generate a blurred version of the input image and then estimate blurriness based on the variation between neighboring pixels in the input versus blurred images. Tsomko and Kim [229] estimate blurriness by using the variance of the prediction residue, which is computed as the difference between adjacent pixels. Debing et al. [230] measure blurring artifacts from H.264/AVC compression by averaging local blur values calculated at the boundaries of macroblocks.

A number of sharpness/blurriness estimators have also been developed based on transform-domain techniques. Marichal et al. [231] estimate sharpness based on the histogram of nonzero DCT coefficients among all 8×8 blocks of the transformed image. Caviedes and Gurbuz [232] estimate sharpness based on the kurtosis of DCT coefficients corresponding to edge profiles. Zhang et al. [233] estimate sharpness based on the peakedness of the image's energy spectrum. Shaked an Tastl [234] estimate sharpness based on the ratio of high-pass to low-pass frequency energy of the spatial derivative of each line/column. Kristan et al. [235] estimate sharpness based on the uniformity of the image spectrum. Hassen et al. [236] estimate sharpness based on the local phase coherence in the complex wavelet domain. Vu and Chandler [237] estimate sharpness based on a weighted average of the log energies of the image's DWT subbands.

Hybrid approaches have also been developed which employ a combination of edge-/pixel-based and transform-based methods. Hybrid approaches have generally proven to perform better than edge-only-based or transform-only-based methods, though at the expense of added computational complexity. Chen and Bovik [238] estimate blurriness based on the image gradient histogram and a wavelet-based edge map. Vu and Chandler [239] estimate sharpness based on a combination of spectral and spatial measures. The spectral measure uses the slope of the local magnitude spectrum, and the spatial measure uses the local total variation of pixel values; these two measures are then combined to generate an image sharpness map, which can be collapsed into a scalar indicating overall perceived sharpness.

4.1.2. Methods for JPEG Compression Artifacts. Numerous NR IQA algorithms have also been developed specifically for JPEG images. The general approach involves measuring the edge strength at block boundaries and then using this measure to estimate the visibility of the blocking, often based on masking. Quality is then determined based on this estimate of perceived blockiness.

In [240], Wang et al. presented an NR measure of blockiness which models blocky images as a nonblocky image corrupted by a pure blocky signal. Quality is estimated based on the energy of the blocky signal. In [241], Wang et al. propose a more efficient method that estimates image blockiness based on the average difference across block boundaries and the activity of the image signal.

In [242], Bovik and Liu presented an NR measure of blockiness which operates in the DCT domain. Blocking artifacts are first located via detection of 2D step functions, and then an HVS-based measurement of blocking impairment is employed.

In [243], Meesters and Martens presented an NR measure of blockiness which operates by detecting low-amplitude step edges and by estimating various edge parameters using a Hermite transform.

In [244], Pan et al. presented an NR measure of blockiness in images/video coded via the block discrete cosine transform. Quality is estimated based on directional information measured for edges. The authors demonstrate that their method does not require the exact location of the block boundary and is thus invariant to displacements, rotations, and scalings of the images.

In [245], Perra et al. exploited properties of the Sobel operator to generate an NR blockiness index based on two measures: one which quantifies the luminance variation of block boundary pixels and one which quantifies the luminance variation of the remaining pixels. Similarly, in [246], Zhang et al. presented a NR blockiness measure which calculates the image's luminance gradient matrix by using the Sobel operator. This matrix is used with HVS-based adjustments (luminance adaptation and texture masking) to estimate the severity of blocking artifacts and the annoyance of large flatness in low-rate images.

Park et al. [247] presented an NR measure for blocking artifacts by modeling abrupt changes between adjacent blocks in both the pixel domain and the DCT domain. Similarly, in [248], Chen et al. presented an NR measure of JPEG image quality by using selective gradient and plainness measures

followed by a boundary selection process that distinguishes the blocking boundaries from the true edge boundaries.

In [249], Suresh et al. presented a machine-learning-based NR approach for JPEG images. Their algorithm operates by estimating the functional relationship between several visual features (such as edge amplitude, edge length, background activity, and background luminance) and subjective scores. The problem of quality assessment is then transformed into a classification problem and solved via machine learning.

In [250], Suthaharan presented an NR technique for quantifying blocking artifacts via two units. The first unit measures the visibility of distortions as a combination of blocking artifacts and undistorted image edges. The second unit uses patterns of the least significant bits to identify image regions that are affected by JPEG compression. Both units are combined to form a normalized visually significant blocking artifact measure.

In [251], Chen and Bloom presented an NR DFT-based measure of blockiness. Given an image, the absolute difference between horizontally adjacent pixels is calculated, normalized, and averaged along each column. Then, a 1D DFT is employed to derive separate measures of horizontal and vertical blockiness. The overall blockiness is estimated based on a combination of these two directional measures.

4.1.3. Methods for JPEG2000 Compression Artifacts. Several NR IQA algorithms for JPEG2000 images have also been developed. The general approach involves measuring the amount of blurring or edge-spread by using edge-detection techniques. Other methods have also been developed based on natural-scene statistics.

In [252], Ong et al. presented a NR algorithm for JPEG2000 blurring which operates via four steps: (1) computing a gradient direction for each pixel; (2) edge detection by using a Canny edge detector; (3) measuring the edge-spread, that is, the extent of the slope of each edge along and perpendicular to the gradient; and (4) estimating quality based on the results from the previous steps.

In [253], Li et al. presented a principal-components-analysis-based NR method for JPEG2000 images. First, by viewing all edge points in JPEG2000 images as distorted or undistorted, local features are extracted at each of the detected edge points to indicate blurring and ringing. A model is then employed to map these local features to local distortion estimates through the probabilities of the edge points being distorted or undistorted. Quality is estimated based on the local distortion estimates. A similar method can also be seen in [254].

In [255, 256], Sazzad et al. presented an approach which uses pixel distortion and edge information for NR IQA of JPEG2000 images. Their technique operates under the assumption that human visual perception is very sensitive to edge information in images. Visual artifacts manifest as pixel distortions around these edges, and thus quality is estimated by measuring these pixel distortions.

Other algorithms for NR IQA of JPEG2000 images have used measures of the changes in the statistic regularities of DWT/DCT coefficients to estimate quality. For example, Sheikh et al. [257] reported that when JPEG2000 images are decomposed through a wavelet transform, the subband probabilities can indicate the loss of visual quality. Quality is estimated in [257] by first computing features based on these probabilities from all wavelet subbands and then applying a nonlinear combination of the features.

Zhou et al. [258] presented an NR algorithm to evaluate JPEG2000 images which employs three steps: (1) dividing the image into blocks, among which textured blocks are employed for quality prediction based on nature-scene statistics; (2) measuring positional similarity via projections of wavelet coefficients between adjacent scales of the same orientation; and (3) using a general regression neural network to estimate quality based on the features from the previous two steps.

Zhang et al. [259] utilized kurtosis in the DCT domain for NR IQA of JPEG2000 images. Three NR quality measures are proposed: (1) frequency-based 1D kurtosis, (2) basis-function-based 1D kurtosis, and (3) 2D kurtosis. The proposed measures were argued to be advantageous in terms of their parameter-free operation and their computational efficiency (they do not require edge/feature extraction).

4.1.4. Methods for Other/Multiple Artifacts. Researchers have also proposed NR methods for other distortion types/combinations, most commonly noise, blurring, blocking, and/or ringing.

In [260], Li presented an NR IQA method which combines NR measures for three different image distortion types (blur, noise, and block/ringing artifacts). Blur is characterized by a 2D parameterized edge model [261]. Impulse noise is measured by the percentage of noisy pixels, and Gaussian noise is estimated as the energy of noise after the application of a denoising algorithm [262]. Blocking is measured by the likelihood of detecting artificial horizontal or vertical edges, and ringing is measured by a ratio indicating the deviation of the spectrum of noise removed by an anisotropic diffusion filter.

In [263], Corner et al. [263] present an NR noise estimation technique based on data masking. Their method calculates a histogram of the local standard deviations over blocks after filtering and edge suppression using a gradient mask. Based on experimental data, they reported that the histogram median value supplied the most accurate final noise estimate.

In [56], Süsstrunk and Winkler presented an NR method for color images degraded by compression or transmission loss. The authors conducted a psychophysical experiment to obtain quality ratings, and they compared their proposed method with the ratings. Three measures for blockiness, blurriness, and colorfulness are proposed and shown to successfully predict the subjective ratings.

In [264, 265], Gabarda and Cristóbal presented an entropy-based NR IQA algorithm. Their method is based on a multiresolution analysis of images for which they conclude that the entropy per pixel is strictly decreasing with respect to decreasing resolution. Quality is estimated based

on anisotropy measures obtained via directional entropy estimates.

In [266], Brandão and Queluz presented an NR algorithm for estimating quantization artifacts due to lossy encoding such as JPEG or MPEG. Their method is based on the statistics of DCT coefficients whose distribution can be modeled by a Laplacian probability density function. The resulting coefficient distributions are used to estimate local errors, and these local error estimates are then used to estimate quality.

In [267], Cohen and Yitzhaky presented an NR method to identify and quantify the impact of noise and blur on quality. Their method operates by measuring common statistics of images obtained from their power spectra. The authors reported that manipulations of the distorted image's spectrum enhance the appearance of the distortion. Accordingly, their resulting method estimates the visual impact of the distortion based on deviations from the expected power spectrum.

4.1.5. Non-Distortion-Specific Methods. Researchers have also developed more general-purpose NR IQA algorithms which do not attempt to detect specific types of distortions. Methods of this type typically reformulate the IQA problem into a classification and regression problem in which the regressors/classifiers are trained using specific features. The relevant features are either discovered via machine learning or specified by using natural-scene statistics.

In [268], Tong et al. presented a learning-based NR algorithm which attempts to directly estimate quality via machine learning. First, some training examples are prepared for both high-quality and low-quality classes. Next, a binary classifier is built on the training set. Finally, the quality estimate of an unlabeled example is denoted by the extent to which it belongs to these two classes.

In [269], Tang et al. presented the LBIQ algorithm, another learning-based NR IQA method. LBIQ measures various low-level quality features derived from natural-scene and texture statistics. LBIQ estimates quality via a regression-based combination of the features.

In [270], Li et al. presented an NR algorithm which operates based on a general regression neural network (GRNN). Various features such as the mean value of a phase congruency map, the entropy of the phase congruency map, and the entropy and gradient of the distorted image are extracted. Quality is estimated by approximating the functional relationship between these features and subjective scores using a GRNN.

In [271, 272], Ye and Doermann presented the CBIQ-I and CBIQ-II algorithms which operate based on visual codebooks. The codebooks consist of Gabor-based features extracted from local image patches. The codebooks form a feature space, which is quantized, and then used to yield an estimate of quality via either example-based regression or support-vector regression.

Another popular approach to NR IQA is to use natural-scene statistics. The main idea in this approach is that natural images demonstrate certain statistical regularities that can be affected in the presence of distortion. Thus, quality can be estimated by extracting features which indicate the extent to which these statistics deviate in the distorted image. See [36] for a more thorough discussion of the use of natural-scene statistics for NR IQA.

Methods of this type usually contain two stages: (1) distortion identification and (2) distortion-specific quality assessment. Both stages require training: the classifier used to measure the probability that each distortion type exists in the distorted image requires training, and the regression model for each distortion type used to map the measured features to an associated quality score must also be trained.

In [273], Moorthy and Bovik presented the BIQI algorithm which estimates quality based on statistical features extracted using the 9/7 DWT. The subband coefficients obtained are modeled by a generalized Gaussian distribution, from which two parameters are estimated and used as features. The resulting 18-dimensional feature vectors (3 scales $\times$ 3 orientations $\times$ 2 parameters) are used to characterize the distortion and estimate quality via the aforementioned two-stage classification/regression framework.

In [274], Moorthy and Bovik presented the DIIVINE algorithm, which improves upon BIQI by using a steerable pyramid transform with two scales and six orientations. The features extracted in DIIVINE are based on statistical properties of the subband coefficients. A total of 88 features are extracted and used to estimate quality via the same two-stage classification/regression framework.

In [275, 276], Saad et al. presented the BLIINDS-I and BLIINDS-II algorithms which estimate quality based on DCT statistics. BLIINDS-I operates on 17×17 image patches and extracts DCT-based contrast and DCT-based structural features. DCT-based contrast is defined as the average of the ratio of the non-DC DCT coefficient magnitudes in the local patch normalized by the DC coefficient of that patch. The DCT-based structure is defined based on the kurtosis and anisotropy of each DCT patch. BLIINDS-II improves upon BLIINDS-I by employing a generalized statistical model of local DCT coefficients; the model parameters are used as features, which are combined to form the quality estimate.

In [277], Mittal et al. presented the BRISQUE algorithm, a fast NR IQA algorithm which employs statistics measured in the spatial domain. BRISQUE operates on two image scales; for each scale, 18 statistical features are extracted. The 36 features are used to perform distortion identification and quality assessment via the aforementioned two-stage classification/regression framework. Related work on the use of BRISQUE features and discriminatory latent characteristics for NR IQA can be found in [278].

4.2. Reduced-Reference IQA. Reduced-reference (RR) IQA methods provide a solution for cases in which the reference image is not fully accessible. Methods of this type generally operate by extracting a minimal set of parameters from the reference image, parameters which are later used with the distorted image to estimate quality.

An important question in RR research is how to determine effective parameters for the IQA task. In [279], Wang

and Simoncelli argued that the appropriate RR features should (1) provide an efficient summary of the reference images, (2) be sensitive to a variety of image distortions, and (3) be relevant to visual perception of image quality.

In [280], Maalouf et al. presented an RR algorithm based on the grouplet transform. Given a reference image and its distorted version, the grouplet transform is applied to both images in order to extract information regarding textures and gradients of the images. This information is then used with CSF filtering and thresholding to obtain sensitivity coefficients. Quality is estimated by comparing the sensitivity coefficients of the distorted image with the sensitivity coefficients of the reference image.

In [281], Guanawan and Ghanbari presented an RR algorithm which operates based on a local harmonic analysis for images degraded with blocking or blurring. Local harmonic amplitude information is computed from an edge-detected image, and this information is then used with the distorted image to estimate quality.

In [282], Rehman and Wang present an RR version of SSIM [123]. Instead of directly constructing an RR algorithm to predict subjective quality, this method extracts statistical features from a multiscale, multiorientation divisive normalization transform. The authors construct a distortion measure by following the philosophy analogous to that in the construction of SSIM. Based on the linear relationship between RR SSIM and FR SSIM given a fixed distortion type, a regression-by-discretization method is used to estimate quality. Related work can also be seen in [283].

In [284], Chono et al. presented an RR algorithm using distributed source coding for remotely monitoring image quality. In this scheme, an image server extracts a feature vector from the reference image and then transmits its Slepian-Wolf syndrome by using a low-density parity-check code. At the decoder, the feature vector and the received (distorted) image are used to estimate quality.

Other types of RR IQA methods operate based on natural-scene statistics. For example, in [279], Wang and Simoncelli presented an RR IQA method which operates based on a wavelet-domain statistical model of images. Quality is estimated based on the Kullback-Leibler divergence between the marginal probability distribution of wavelet coefficients of the reference and distorted images. Similar work can also be found in [285].

In [286], Xue and Mou presented an RR algorithm based on the steerable pyramid. At each pyramid scale, a strongest component map (SCM) is constructed by assembling coefficients with maximum amplitudes among different orientations. Several statistics of the SCM serve as the RR features, which are used with the distorted image to estimate quality.

In [287], Avanaki et al. presented an RR algorithm which operates by using watermarking to embed RR features into an image; these features can be extracted and used for IQA in the event that the image is distorted. The RR features used in [287] consist of approximation coefficients of a parameterized DWT of the image. At the receiver, the embedded features are extracted and compared to the corresponding features of the distorted image to estimate quality.

In [288], Li and Wang presented an RR algorithm based on a divisive normalization image representation. By using a Gaussian-scale mixture-based statistical model of wavelet coefficients, a divisive normalization transform (DNT) is applied to the images. Quality is estimated by comparing a set of RR statistical features extracted from DNT-domain representations of the reference and distorted images.

In [289], Ma et al. presented an RR algorithm based on DCT coefficient statistics. First, the DCT coefficients of image blocks are grouped into several representative subbands. Next, a generalized Gaussian distribution is employed to model the distribution of coefficients within each subband. Quality is then estimated based on the distance between distributions of the reference and distorted images. Similar work can also be seen in [290].

In [291], Soundararajan and Bovik presented a framework for RR IQA based on information-theoretic measures of differences between the reference and distorted images by using the entropies of wavelet coefficients. This algorithm differs from other approaches in terms of the amount of data needed for the entropy-difference calculations and in terms of the scalability in the amount of information that is needed from the reference image.

5. Seven Challenges in Image Quality Assessment

The previous sections of this paper have focused on reviewing the current knowledge and accomplishments in IQA research. In this section, the focus now shifts toward unsolved aspects of IQA. Here, I specifically discuss seven open challenges, all of which are critical for furthering IQA research and facilitating deployment and integration of IQA into existing and forthcoming applications.

It is important to stress again that these seven challenges do not represent an exhaustive list of important research topics in IQA. Other notable areas such as IQA of stereoscopic images, IQA of computer graphics, and video quality assessment are not discussed. The seven challenges described in this section were chosen to highlight some key limitations of current IQA knowledge and to point out areas which can begin to answer broader questions on IQA.

5.1. Challenge 1: HVS Models and Natural Images. As discussed in Section 2, knowledge of how the HVS analyzes visual input has played a pivotal role in IQA research. However, it must be stressed that our current understanding of the HVS, and thus the computational HVS modeling used in IQA, is far from complete. The vast majority of computational models do not model beyond primary visual cortex (V1), and many researchers have argued that even current V1 models are still incomplete. And, how visual stimuli are analyzed in V1 is only one contributor to visual perception, not to mention image quality.

A key difficulty in developing a more complete computational HVS model is the fact that visual neurons often respond quite differently to naturally occurring stimuli than they do to simple, controlled stimuli. Because neural responses are

highly nonlinear, it has proved difficult to predict neural responses to natural images based on responses to more simple stimuli. The same difficulty carries over to visual perception; psychophysical data collected using natural scenes can be difficult to predict based on data collected using more simple, controlled stimuli.

5.1.1. The Need for Improved V1 Models. While the characterization of V1 based on its responses to simple stimuli has proved useful, other researchers have suggested that in order to fully understand the response properties of visual cortex, one must first understand the signal that is to be encoded, that is, *natural scenes*. (As defined by Field [292], *natural scenes* refer to images from the natural environment which are devoid of man-made objects. However, the exclusion of man-made objects has since been relaxed in the visual psychology literature due, in part, to the presence of such objects in several popular natural-scene databases [293, 294]. In the image-processing literature, the term *natural images* is more commonly used and defined to be photographic images containing any naturally occurring subject matter that may occur during normal photopic or scotopic vision.) Field [292, 295, 296] postulated that cortical neurons are tuned to encode natural scenes in an efficient manner, and thus this special class of input has the potential to reveal properties of visual cortex beyond those invoked by using simple stimuli. Indeed, several studies have shown that neural networks trained with natural scenes under various sparse-coding objectives yield bases which possess striking similarities to simple and complex cortical cells [297, 298]. Other studies, focused on modeling the statistics of natural scenes, have revealed properties such as $1/f$ amplitude spectra [292, 299] (where f denotes spatial frequency) and the importance of phase [300, 301] and edge cooccurrences [109] in perception. More recently, natural images have been used in psychophysical studies [55, 302–305], revealing both supportive and confounding evidence for previous theories of V1.

In a recent, controversial essay, Olshausen and Field argued that as much as 85% of V1 has yet to be explained [306]. Much of what is known about the response properties of V1 neurons has come from neurophysiological studies employing single-cell recordings. During such recordings, many V1 neurons are not tested due to the fact that they yield weak extracellular action potentials, they yield low firing rates, or they are otherwise visually unresponsive. Olshausen and Field estimated that only 40% of V1 neurons have actually been tested. They further reported that, of these V1 neurons that have been tested, current computational models can explain only 30–40% of the response variance when the neurons are presented with natural images.

The lack of knowledge about how V1 operates when presented with naturally occurring stimuli is even more troubling for IQA because there is always an image present (usually a natural image), and thus our models must necessarily be equipped to handle such stimuli. In terms of visual masking, computational models have been not been extensively tested on thresholds measured using natural images as masks. Instead, as mentioned in Section 2.1, many of the models employed in IQA algorithms use parameters which have been optimized to fit thresholds measured for traditional targets placed upon unnatural masks (e.g., sine waves, Gabor patterns, or noise).

As an example of how natural images can affect masking results, Figure 7 shows relative detection thresholds measured for wavelet distortions presented against various image patches [64]. In this plot, the image patches, which are shown along the horizontal axis, have been ordered by eye based on the ability to learn and recognize changes to the content—from simplistic edges to complex textures (from left to right). All images were matched in RMS contrast (values of 0.32 and 0.64 were tested). The data points denote relative threshold elevation given by $\mathrm{CT}/\mathrm{CT}_{\mathrm{edge}}$, where CT denotes the contrast threshold for detecting the distortion and $\mathrm{CT}_{\mathrm{edge}}$ denotes the average contrast threshold for detecting the distortion in images from the edge category. The dashed lines denote average relative threshold elevations for each of the three image types.

Notice from Figure 7 that the thresholds generally increase as the images become more visually complex and thus become harder to learn and recognize, despite the fact that all images have the same RMS contrast. In [64], we reported that a computational model of masking (described in Section 2.1) performed well in predicting only the thresholds for the texture patches; these models generally failed for patches in the edges' and structures' categories. In fact, even after optimizing and extending the parameters of the model based on the actual thresholds measured for the edges and structures, the predictions remained quite subpar.

5.1.2. Ground-Truth Data for Natural Images. For IQA, one of the primary limitations when designing a computational neural model that can handle natural images is the lack of ground-truth data. For masking, which is crucial for IQA in the high-quality regime, there exists no database of local contrast detection thresholds for natural images (though, see [49, 58] for related studies). For lower quality images, there exists no database of ground-truth quality maps denoting local quality ratings for natural images. These types of local ground-truth data would be especially useful for training and testing purposes since they can provide insights into whether the local processing is correctly modeled or whether further adjustments are warranted.

Of course, the main difficulty in creating local masking and quality-rating data is the enormous time commitment required for the experiments. Nonetheless, even coarse maps would be a useful first step. For masking, we have begun to address this issue by measuring local masking maps for detection of vertically oriented 3.8 c/deg log-Gabor distortions in images from the CSIQ database [27]. Figure 8 shows preliminary results for 10 of the images.

In Figure 8, the first row shows the original images (masks), the second row shows maps of local detection thresholds in which brighter values denote greater thresholds (greater masking); these maps were averaged across subjects and trials. The third row shows the predicted masking maps

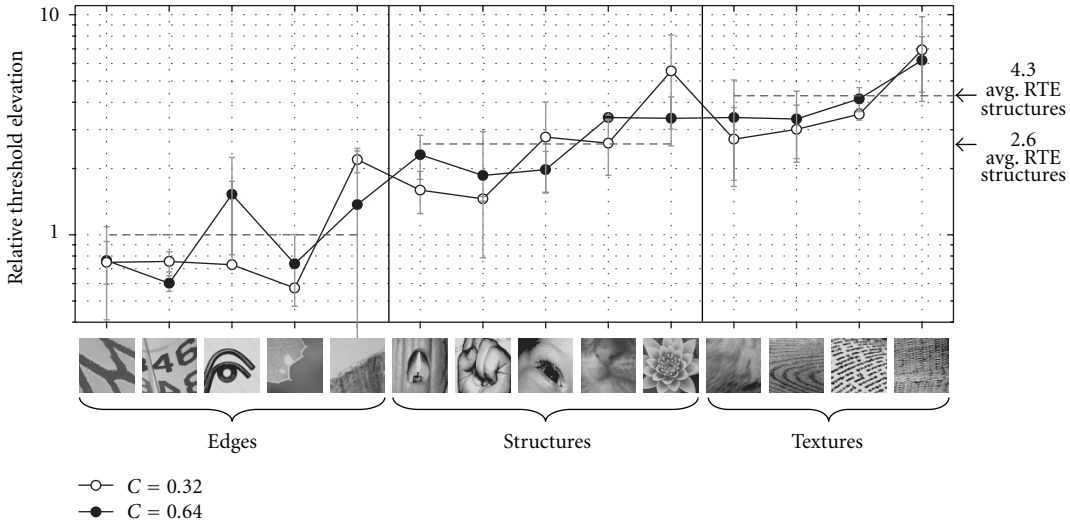


Figure 7: Relative threshold elevations for each of the 14 masks used in [64]. All masks were matched in RMS contrast (values of 0.32 and 0.64 were tested). The data points denote relative threshold elevations averaged over all subjects; error bars denote standard deviations of the means over subjects. The dashed lines denote average relative threshold elevations for each of the three image types. The images depicted on the horizontal axis have been ordered by eye to represent a general transition from simplistic edge to complex texture (from left to right). Indeed, notice that the data generally demonstrate a corresponding left-to-right increase in relative threshold elevation.

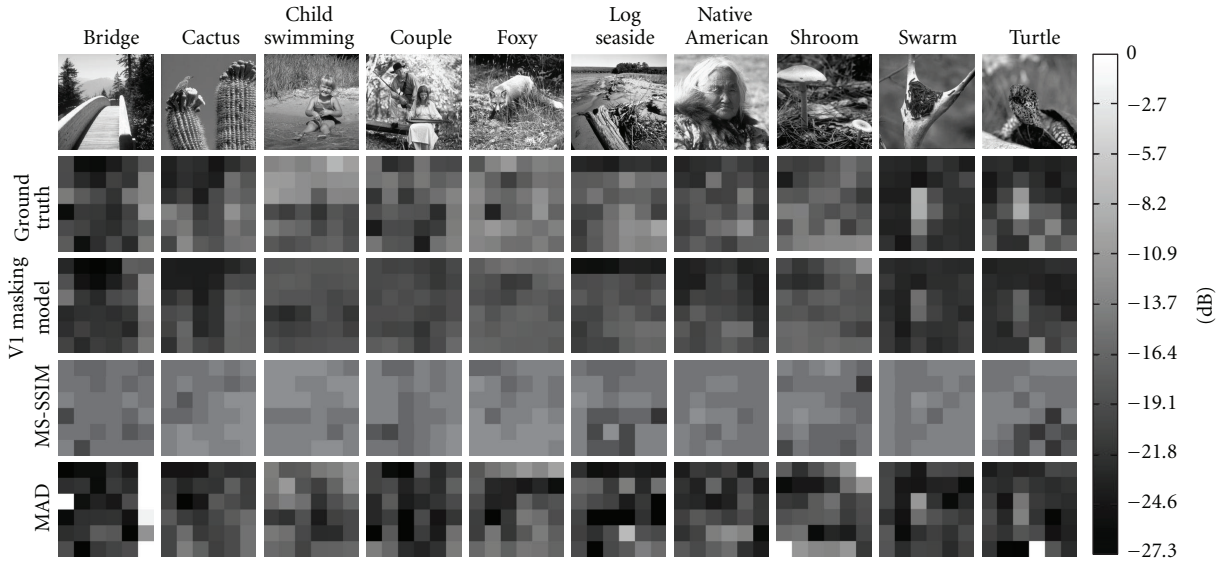


Figure 8: Local masked detection thresholds and corresponding predicted thresholds for 10 images from the CSIQ database. See text for details.

obtained from the computational V1 model of masking used in [64], which is based on the standard model described in Section 2.1. For comparison, the fourth and fifth rows show the predicted masking maps obtained from MS-SSIM [175] and the detection-based stage of MAD [139]. Because neither of these IQA algorithms was designed to predict thresholds, we selected a fixed index for each algorithm corresponding to an at-threshold amount of distortion (an MS-SSIM index of 0.995, a MAD index of 15). Then, for each patch of the image, we successively added log-Gabor distortions until the algorithm yielded that at-threshold index for the patch. Finally, the resulting masking map was computed as the RMS contrast of the distortion in each of these distorted patches.

Overall, the neural masking model yields the best predictions of the masking maps with an average Pearson correlation coefficient (CC) between the actual and predicted thresholds of 0.70. The best prediction from this model is on image *cactus* for which the CC is 0.95. However, there are many notable failure cases, particularly on more structured images. The worst prediction from this model is on image

couple for which the CC is 0.42. Both MS-SSIM and MAD perform considerably worse. MS-SSIM yields an average CC of 0.44, with the best and worst predictions on, respectively, images *foxy* (CC = 0.67) and *native American* (CC = 0.13). MAD yields an average CC of 0.64, with the best and worst predictions on, respectively, images *bridge* (CC = 0.86) and *shroom* (CC = 0.28).

Of course, MS-SSIM was never designed to estimate masking, and the masking model employed in MAD is a simplistic, spatial-domain-only local contrast measure. However, the performance of the neural masking model highlights the need for further research in this area.

5.1.3. Models of Areas Beyond V1. One possible explanation for the failure cases of V1 models in predicting masking in natural images is the fact that such masking is attributable to visual processing in areas beyond V1 [81, 307]. For example, by comparing EEG recordings obtained during unmasked versus masked detection, Fahrenfort et al. concluded that "*masking derives its effectiveness, at least partly, from disrupting reentrant processing, thereby interfering with the neural mechanisms of figure-ground segmentation and visual awareness itself*" [307]. Thus, even if a complete model of V1 for natural images was available, there still remains the question of how masking and image quality are influenced by visual processing in areas beyond V1.

Unfortunately, much less is known about the mechanisms and objectives of visual processing beyond V1 and the influences such processing might have on V1 itself. It is important to note that approximately half of V1's innervations come in the form of corticocortical feedback from higher levels [307]. Lee et al. [308] have proposed that the higher levels work in conjunction with V1 to perform complex tasks such as pattern analysis and object recognition. Rao and Ballard [309] have suggested that the higher levels function as predictive coders whose feedback connections to V1 carry the prediction and whose feedforward connections from V1 convey the prediction's error.

One generally accepted belief is that higher levels serve to efficiently encode the joint activity of V1 neurons [310–312]. Based on single-cell recordings in V2, Willmore et al. argued that V2 neurons integrate the outputs of V1 neurons across spatial frequency to enhance the representation of edges [313]. Related earlier work in visual psychophysics postulated similar theories that an image's features are integrated temporally across scale-space in a coarse-to-fine (global-to-local) fashion [314–317]. In terms of IQA, any distortions which disrupt this integration of neural responses/features could potentially lead to severe degradations in quality.

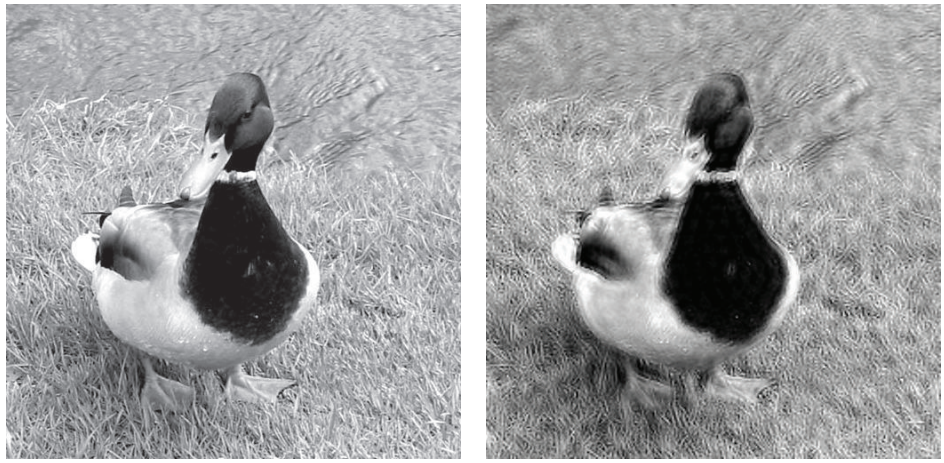
As an example, the *global precedence* theory of Navon [314] and the related *scale-space integration* theory of Hayes [317] both advocate that an image's edges are visually processed by combining information across spatial scales, beginning with the coarsest scale and ending with the finest available scale. Under this theory, eliminating or distorting content at an intermediate spatial scale should disrupt the HVS's ability to integrate coarse and fine information into a single percept. Instead, the result would be two percepts: a blurred version of the object and a separate, erroneous high-frequency pattern. This disruption of visual integration and its effects on image quality can be readily demonstrated, as shown in Figure 9. In this figure, the original image is shown on the left, and the distorted image is shown on the right. The distorted image was generated via notch filtering (discrete-space radial frequencies of $\pi/8$ to $\pi/4$ have been eliminated for all orientations). Because this filtering disrupts the HVS's ability to integrate information at different scales, the high-frequency content is perceived almost as additive noise on top of a second percept of a blurry image in the background.

The demonstration in Figure 9 highlights the need for further research on the role of higher-level visual processing on image quality. Although the mechanisms and functional objectives of these higher-level areas remain largely unknown, it is still possible to incorporate general principles of object perception and cognition derived from experiments or heuristics. For example, in the VSNR algorithm [125] (described previously in Section 3), we developed a basic model of global precedence specifically for IQA; although this model was based more on empirical observations rather than psychophysical data, it has proved quite effective for certain distortion types. Nonetheless, further research in this area, both in vision science and in IQA, is needed to help move IQA beyond its current capabilities to a level that can begin to capitalize on the properties of higher-level visual processing.

5.2. Challenge 2: Compound and Suprathreshold Distortions. Another challenge which designers face when incorporating psychophysical findings into IQA algorithms is the fact that distortions can be both compound and suprathreshold. The term *compound* is used to describe a visual target that stimulates more than one channel in the HVS's multichannel analysis. *Suprathreshold* refers to clearly visible targets that are at a contrast beyond the threshold of detection. In IQA, many of the distortions that are encountered meet both of these criteria.

5.2.1. Simple Targets versus Compound Distortions. In image processing applications, a wide variety of distortions are possible, and an IQA algorithm should ideally be able to handle such distortions. However, in visual psychophysics, the visual targets are generally much more simplistic, often consisting of sine-wave gratings, Gabor functions, or other highly controlled spatial patterns that are localized in space, frequency, and/or orientation. Highly controlled and localized targets are preferable in such experiments because they can be designed to stimulate only one channel of the HVS's multichannel analysis. However, this difference between simple targets and actual distortions poses a difficulty to designers who wish to incorporate psychophysical findings into an IQA algorithm.

In visual psychophysics, distortions would be considered *compound* targets that consist of multiple simple targets (e.g., multiple sine waves, Gabor functions, or wavelets). Due to the nonlinear behavior of the HVS, it is difficult to apply knowledge about the visibility of simple targets to predict how the HVS will respond to compound distortions.



(a) (b)

Figure 9: Demonstration of the visual effects of disrupting the HVS's ability to integrate features across scale space. The original image is shown on (a), and the distorted version is shown on (b). The distorted image was generated by notch-filtering discrete-space radial frequencies of $\pi/8$ to $\pi/4$. This filtering disrupts the HVS's ability to integrate information at different scales, and thus the high-frequency content is perceived as additive noise on top of a blurry image in the background.

Visual summation studies begin to address this issue by measuring thresholds for detecting compound targets and comparing them with thresholds measured for the individual components of the compound (the individual simple targets).

Traditionally, summation has been tested by using a visual detection paradigm, in which the detection threshold measured for a compound target (e.g., a plaid composed of two sine waves) is compared with detection thresholds measured separately for its individual components (the sine waves). Detection of the compound should be an easier task because the HVS has a greater of chance of detecting the target now that it contains two instead of just one component. Further, assuming that the components within the compound are detected by separate HVS channels, any changes in detection thresholds for compound versus simple targets would point toward interchannel cooperation, that is, *summation* of channel responses.

As mentioned in Section 2.1, summation is typically modeled via a Minkowski sum in which the amount of summation is controlled by the summation exponent β (see (4)). A value of $\beta = 1$ denotes complete summation, whereas a value of $\beta = \infty$ denotes no summation. The majority of psychophysical studies have generally found β to be in the range $\beta \approx 3$–5; that is, the compound target is only slightly more detectable than either of its components given that the components are analyzed by separate HVS channels.

However, as mentioned in Section 5.1, the presence of an image in the background can significantly change neural responses and psychophysical results, and visual summation is no exception. In [55], we measured detection thresholds for simple and compound distortions generated via quantization of individual DWT subbands (yielding simple distortions) and pairs of DWT subbands (yielding compound distortions). When the distortions were presented against a gray background (i.e., unmasked detection, as used in previous studies), we found results which were very consistent with previous studies: $\beta \approx 3$–5. However, when the distortions were presented against either of the two natural images tested in [55], much greater summation was found: $\beta \approx 1.3$–1.6. One possible explanation for our finding is the presence of intra-channel summation, wherein a single HVS channel is used to detect both components of the compound target. Intra-channel summation may result from the spatial correlations that exist between the distortions and the images and/or from off-frequency looking [83, 318].

It is also important to note that summation studies have tested compound targets consisting of relatively few components. However, in IQA, the distortions can be broadband in terms of radial frequency, orientation, and other dimensions. It remains unclear how summation is affected when such distortions serve as the targets. Furthermore, the distortions encountered in IQA may contain components which constructively interact to form salient visual patterns that might further affect summation. For example, if false contours or borders are formed, visual processing could be mediated by areas beyond V1 which attempt to integrate such contours [319], thus giving rise to different summation rules. Clearly, further research is needed in order to develop improved models of summation in the presence of natural images.

5.2.2. Perceived Contrast of Suprathreshold Distortions. Much of our current understanding of visual perception has resulted from research in visual detection in which the task is to gauge whether the distortions are visible. However, many applications generate images containing *suprathreshold* distortions whose contrasts are well beyond the threshold of detection. Thus, another challenge which IQA designers face is how to adjust existing models to handle suprathreshold distortions.

One key finding from studies employing suprathreshold stimuli is the fact that the perceived (or "apparent") contrast of a suprathreshold target depends much less on the target's spatial frequency than what is predicted by the CSF; that is, visual sensitivity at suprathreshold contrasts is relatively frequency independent. This finding, termed *contrast constancy* [320], was first reported by Georgeson and Sullivan [320] who attributed the effect to an intrachannel gain control mechanism that compensates for the CSF at suprathreshold contrasts.

In a similar study, Brady and Field [76] found contrast constancy using both Gabor patches and broadband noise patterns. Their data were successfully predicted via a model with equally sensitive octave-bandwidth spatial-frequency channels, which was reported to yield a constant response to the spatial scales of natural scenes. Brady and Field's study and a later study by Graham et al. [77] were the first to provide a theoretical account of why white noise is perceived as containing mostly high-frequency content despite the fact that the CSF peaks near 4–6 c/deg.

For IQA, not only can the distortions be suprathreshold, but the distortions are necessarily presented against an image. Thus, the perceived contrast of the distortions can be influenced by the image. In [54], we investigated whether contrast constancy is also observed in the presence of natural images by repeating Brady and Field's experiment using suprathreshold octave-bandwidth wavelet distortions presented against either a solid gray background or one of three natural images. In the experiment, subjects were asked to adjust the contrasts of 1.15, 2.3, 4.6, and 9.2 c/deg wavelet distortions such that they appeared to have the same contrast as 18.4 c/deg distortions, the latter of which was fixed in contrast at various suprathreshold values. Figure 10 shows the results of the experiment.

In Figure 10, the horizontal axis of each graph corresponds to the center spatial frequency of the wavelet distortions. The vertical axis of each graph denotes the physical RMS contrast of the distortions increasing in the downward direction. The topmost curve in each graph (square symbols) corresponds to average detection threshold values from a previous experiment [53]; this curve can be interpreted as a CSF for wavelet distortions. The lower four curves in each graph indicate how perceived contrast changes for increasingly suprathreshold distortions.

Notice in Figure 10 that the lower curves in each graph (the perceived contrast curves) are much flatter than the top curve (the CSF curve). These data indicate that, at increasingly suprathreshold contrasts, the perceived contrast becomes increasingly invariant with frequency; that is, contrast constancy is observed. The curves obtained when using the image backgrounds demonstrate a slight reduction in the perceived contrast for lower-frequency distortions, but the contrasts still demonstrate a significantly lesser dependence on spatial frequency than the top (CSF) curve.

However, although contrast constancy can be used to estimate the perceived contrast of the distortion, contrast constancy has found little use in IQA. To demonstrate why, Figure 11 depicts images to which horizontally oriented 1.15–18.4 c/deg wavelet distortions have been added (these distortions were not generated via quantization and are thus spatially uncorrelated with the image). The RMS contrasts of the distortions in these images have been allocated in two different ways: For the image on the left, the contrasts have been proportioned according to the CSF (specified by the top curve in Figure 10); for the image on the right, the contrasts have been proportioned as specified by the middle curve (solid circles) of Figure 10 for the image *lena*. The distortions in both of these images exhibit a total RMS contrast of approximately 0.18.

Whereas the results of the contrast-matching experiments suggest that when distortions are suprathreshold, physical contrast is a better indicator of perceived contrast than predictions based on the CSF, Figure 11 clearly demonstrates that image quality is much better preserved when the contrasts of the distortions are proportioned according to the ratios specified by the CSF. One possible explanation for this effect is due to the total perceived contrast of the compound suprathreshold distortions. Specifically, by examining just the distortions in Figure 12, it is clear that the perceived contrast of the distortions when using contrast-matching-based proportions is much greater than the perceived contrast of the distortions when using CSF-based proportions. Thus, there appears to be an unexpected visual summation effect: although the perceived contrasts of the individual bandlimited distortions are relatively constant, when these distortions are combined and viewed together, they appear to visually interact in a way that affects the total perceived contrast.

Unfortunately, whereas visual summation at near-threshold contrasts has been extensively studied, visual summation of perceived contrast has received much less attention (see [321]). For IQA, visual summation of perceived contrast would further need to be tested in the presence of various images. Clearly, there is a need for more research in this area. Such research could prove particularly useful for IQA of images containing distortions which are perceived to be overlaid on top of the image as opposed to distortions which interact with the image's subject matter. In the following section, I delineate between these two types of distortions.

5.3. Challenge 3: Effects of Distortions on Image Appearance. Although the perceived contrast of the distortions can be used to estimate quality, the inherent assumption in this approach is that the viewer is looking for the distortion in the presence of the image. When the distortions are severe and spatially correlated with the image, viewers tend base quality judgments on the interaction between the distortions and the image's objects. Properly determining and modeling the perceptual effects of this interaction is yet another challenge in IQA research.

5.3.1. Capture and Transparency in IQA. For IQA, an important consideration is whether the distorted image is perceived as a single distorted image or whether it is perceived as distortions with an image in the background. In the visual psychophysics literature, these two scenarios would fall under

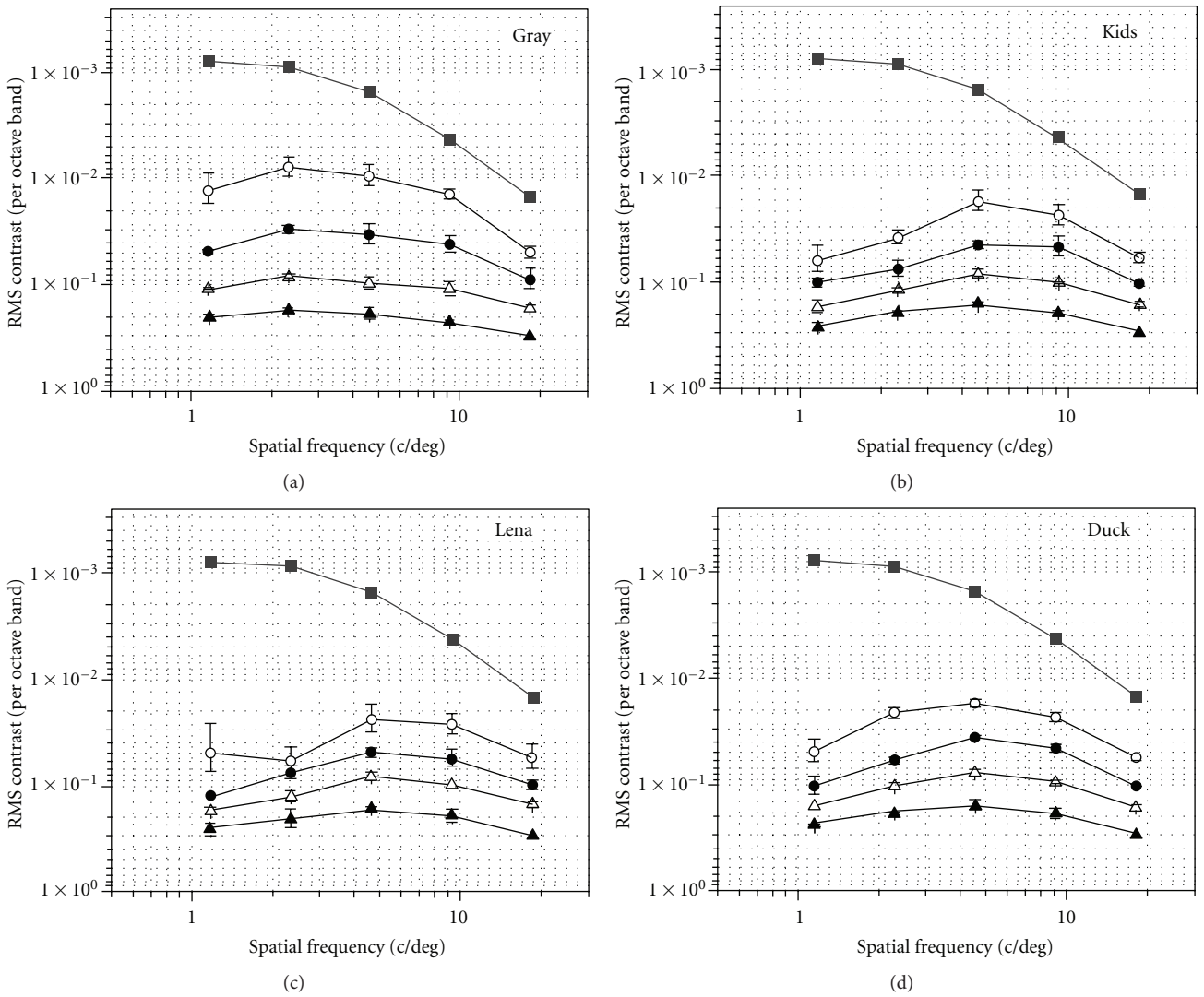


Figure 10: Contrast-matching results from [54] for octave-bandwidth wavelet distortions presented against a solid gray background and three images. See text for details.

the aegis of *capture* and *transparency* [322], which describe whether a target + background are perceived as one combined stimulus (captured) or whether they are perceived as two separate stimuli (transparent).

One of the earliest studies pointing to the need to distinguish between capture and transparency in IQA was performed by Goodman and Pearson [171] who used a multidimensional scaling (MDS) experiment. For image quality, an MDS experiment can be used to determine the number of, type of, and interactions between the perceptual attributes that underlie quality ratings. Goodman and Pearson used MDS to specifically investigate the quality of TV pictures impaired both by additive-type distortions (e.g., noise, echo) and by coding- and transmission-type distortions (DPCM quantization artifacts and blurring). Based on their MDS analysis, they reported that one of the multiple perceptual dimensions "*appears to be separating those impairments which cause the integrity of the objects in the picture to be destroyed from overlay types of impairment.*"

To demonstrate the effects of capture and transparency on image quality, Figure 13 shows two distorted versions of *lena*. The image on the left is repeated from Figure 11; this image contains additive wavelet distortions which are spatially uncorrelated with the image. The image on the right was generated via actual quantization of the wavelet subbands, thus resulting in distortions which are spatially correlated with the image (distortions which interact with the image's objects). For both images, the contrasts of the per-subband distortions have been proportioned as specified by the middle curve (solid circles) of Figure 10 for the image *lena*. The distortions in both images are at a total RMS contrast of approximately 0.18. Figure 14 shows the distortions in these images presented against a solid gray background.

In the experiment in [54], subjects were instructed to match the contrasts of the wavelet distortions, a task which involves examining the distortions. However, when the distortions are severe and spatially correlated with the image, judging the quality of the image involves attending

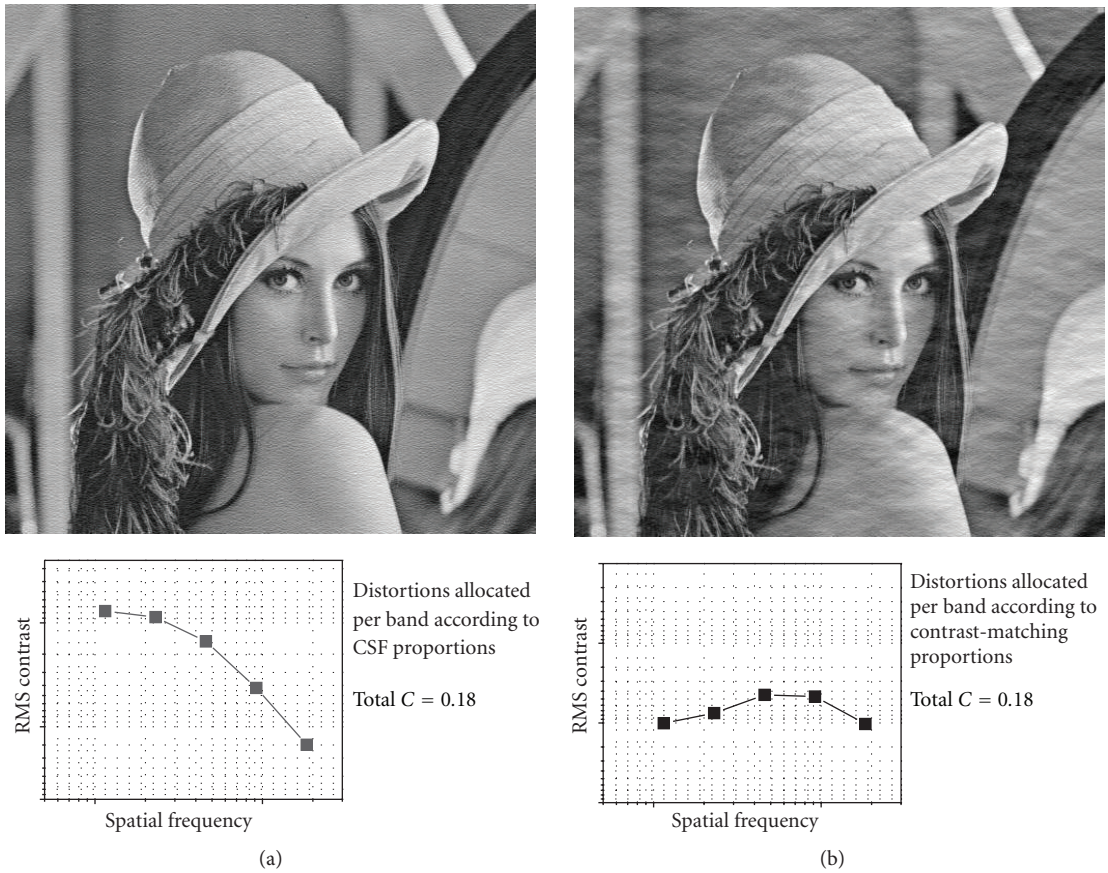


Figure 11: Images containing horizontally oriented wavelet distortions generated by adding random values to the LH subbands of a five-level DWT of the image *lena*. In the image on the left, the contrasts of the distortions have been proportioned according to ratios derived from the CSF. In the image on the right, the contrasts of the distortions have been proportioned according to the middle trend (filled circles) of Figure 10 for the image *lena* which specifies relatively constant contrasts across spatial frequency. Both images contain distortions at a total RMS contrast of approximately 0.18. These images were created assuming sRGB display characteristics and are meant to be viewed from approximately four picture heights.

to and looking at (capturing) the image. Indeed, the images in Figures 13 and 14 suggest that it is not just the perceived contrast of the distortions that determines the image's quality; notice from Figure 14 that the overall perceived contrasts of just the distortions are quite similar for both images. Rather, quality is also determined by visual interaction between the distortions and the image's subject matter. Nachmias [323] reported a similar observation in context of masked detection of sine-wave gratings. Namely, when a target is presented against a suprathreshold and spatially coherent mask, it is often easier to detect the target by examining its effect on the phenomenal appearance of the mask.

Thus, in addition to considering the perceived contrast of the distortions, for IQA, it is also important to take into account the effects these distortions impose on the appearance of the image. In particular, it would seem necessary to distinguish between the additive or "overlay" types of distortion and those which visually interact with the image's subject matter.

5.3.2. The Role of Visual Strategy in IQA. The effects of capture and transparency in IQA begin to address the broader issue of the adaptive nature of the HVS. Namely, the visual strategy that the HVS uses when judging image quality can change depending on both the amount of distortion and whether the distortion affects the phenomenal appearance of the image's objects. Numerous studies have shown the HVS to be a highly adaptive system, with adaptation occurring at multiple levels ranging from single neurons [69] to the entire cognitive processes [324]. It seems logical to assume that the visual strategy adapts based on many other factors related to the interaction between the distortion and the image.

In [139], we asked whether IQA could be improved by modeling this adaptive nature of the HVS via two separate computational models. For images containing near-threshold distortion (high-quality images), we assumed that transparency was in effect, and thus the HVS employs a detection-based strategy in an attempt to look for the distortions. For images containing suprathreshold distortion (low-quality images), we assumed that capture was in effect, and thus the HVS employs an appearance-based strategy in an attempt to recognize the image's content.

Figures 15 and 16 demonstrate the need to explicitly model these two separate strategies. As shown in Figure 15,

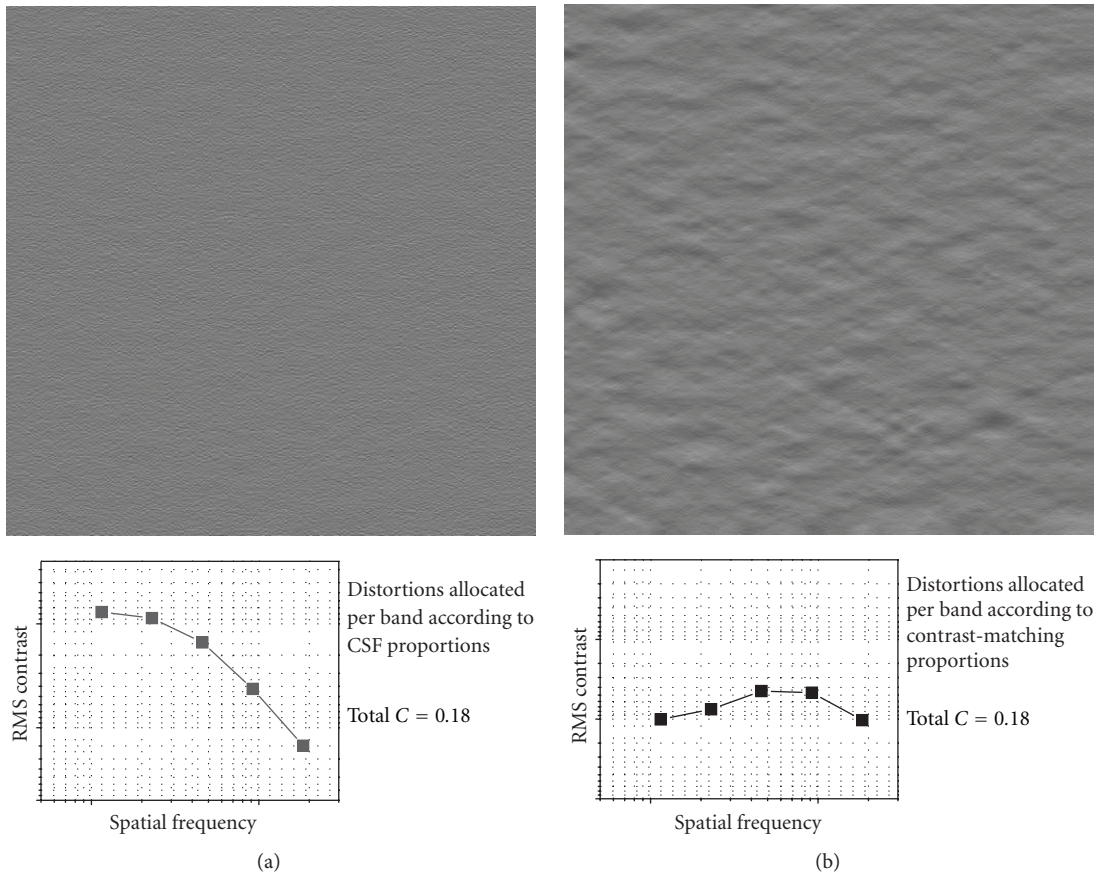


Figure 12: Distortions from Figure 11 presented against a solid gray background.

which contains high-quality images, when the distortions are not readily visible, our visual system seems to employ a detection-based strategy in an attempt to locate any visible differences. However, for the low-quality images shown in Figure 16, which contain suprathreshold distortions, the distortions dominate the overall appearance of each image, and thus visual detection is less applicable. Instead, for these latter images, quality is determined based primarily on our ability to recognize image content. We demonstrated in [139] that by using separate computational models for these two fundamentally different strategies and by estimating quality based on an adaptive combination of these modeled outputs, significant improvements in IQA could be achieved.

In [67], Rouse et al. performed a more specific study designed to investigate the role of recognition in determining image quality and image utility (usefulness). For high-quality images, Rouse et al. reported that the perceived utility scores do not correlate with the perceived quality scores. However, for low-quality images, a linear relationship between perceived utility and perceived quality was reported. These results and the later work in [205] suggest that the ability to recognize and utilize an image's content can play a crucial role in determining quality.

5.3.3. Higher-Level Effects of Distortion on Image Appearance. Although the efforts in [67, 139, 205] and the related work in [201] begin to address the issue of adaptive visual strategies, more generally IQA research could benefit from a better understanding of the interaction between the distortions and images. As an example, consider the two images shown in Figure 17, one of which has been compressed with JPEG and the other with JPEG2000, both at the same low bit-rate. Both images contain compression distortions which are clearly visible, since, at this low rate, there is no chance of hiding the distortions in the traditional sense. However, most people clearly prefer the JPEG2000 image over the JPEG image.

One explanation for the lower quality of the JPEG image is due to JPEG's blocking artifacts. However, there are several other, higher-level aspects which come into play: (1) the facial expressions are better preserved in the JPEG2000 image because the wavelet basis functions better capture the curvature of the eyes; (2) the objects in the image—water, skin, and hair—happen to be physically smooth in the real world, so JPEG2000's blurring is acceptable for these objects; (3) the object boundaries are less degraded in the JPEG2000 image, and therefore it is easier to recognize the image's subject matter. None of these higher-level perceptual aspects are considered in the current coding and IQA algorithms; they just happen to work in JPEG2000's favor for this particular image. An improved IQA algorithm which explicitly models higher-level perception can potentially lead to better quality estimates and thereby benefit compression and other image-processing applications.

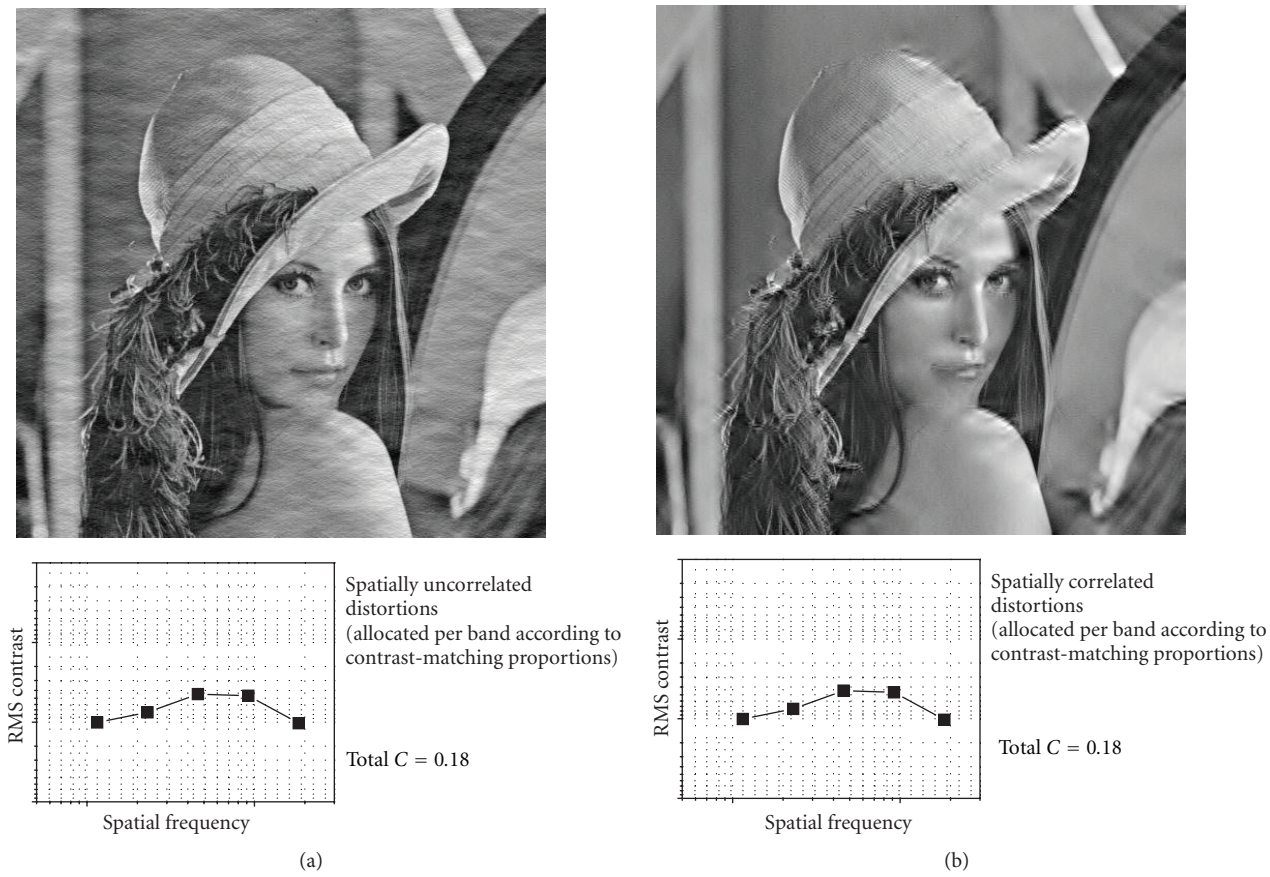


Figure 13: Images containing horizontally oriented wavelet distortions generated by adding random values to (a) or via actual quantization of (b) the LH subbands of a five-level DWT of the image *lena*. In both images, the contrasts of the distortions have been proportioned according to the contrast-matching proportions as described in Section 5.2. Both images contain distortions at a total RMS contrast of approximately 0.18. These images were created assuming sRGB display characteristics and are meant to be viewed from approximately four picture heights.

5.4. Challenge 4: Multiple Types of Distortions. Although IQA algorithms have been tested on images containing individual types of distortion, some applications can give rise to images which simultaneously contain *multiple types* of distortions. This scenario adds yet another level of difficulty for IQA. An IQA algorithm must not only consider the joint effects of these distortions on the image, but also consider the effects of these distortions on each other.

Consider, for example, a traditional model in signal processing in which the output signal y is given by $y = T(x) + n$, where $T(x)$ is a transformation of the input signal x (e.g., blurring or JPEG2000 coding-decoding) and n is additive noise. Here, the output image y may contain at least two distinct types of distortions: $T(x) - x$, which most often represents distortions that are spatially correlated with the image (e.g., blurring, ringing artifacts), and n, which is most often modeled as noise that is spatially uncorrelated with the image. Below, I describe several studies which have investigated the joint effects of these distortions on image quality.

5.4.1. Joint Effects of Blur and Noise. Much of the early work in investigating the effects of multiple types of distortions on image quality involved the use of multidimensional scaling (MDS) experiments. As described in Section 5.3, Goodman and Pearson [171] used MDS to investigate the joint effects additive-type distortions (e.g., noise, echo) and coding- and transmission-type distortions (DPCM quantization artifacts and blurring). They reported that four perceptual dimensions were used by subjects to rate quality: "(1)*overall picture clarity,* (2) *a distinction between overlay impairment and object impairment,* (3) *the amount of purely spatial or stationary overlay patterning, and* (4) *the amount of spatiotemporal or moving overlay patterning.*" [171].

In [325], Linde employed MDS to investigate the interactive effects of blur and noise. The images used in [325] were blurred in varying amounts, and then varying amounts of noise were added to these blurred images. Linde reported two key findings. First, when a fixed amount of noise was added to images subjected to varying amounts of blur, the perceived strength of the noise increased for the increasing amount of blur; that is, the noise is more pronounced when it is added to a more blurry image than when the noise is added to a less blurry image. This result would seem attributable, at least in part, to masking; that is, blurring an image reduces its ability to mask noise. Second, for a fixed amount of blur, the addition

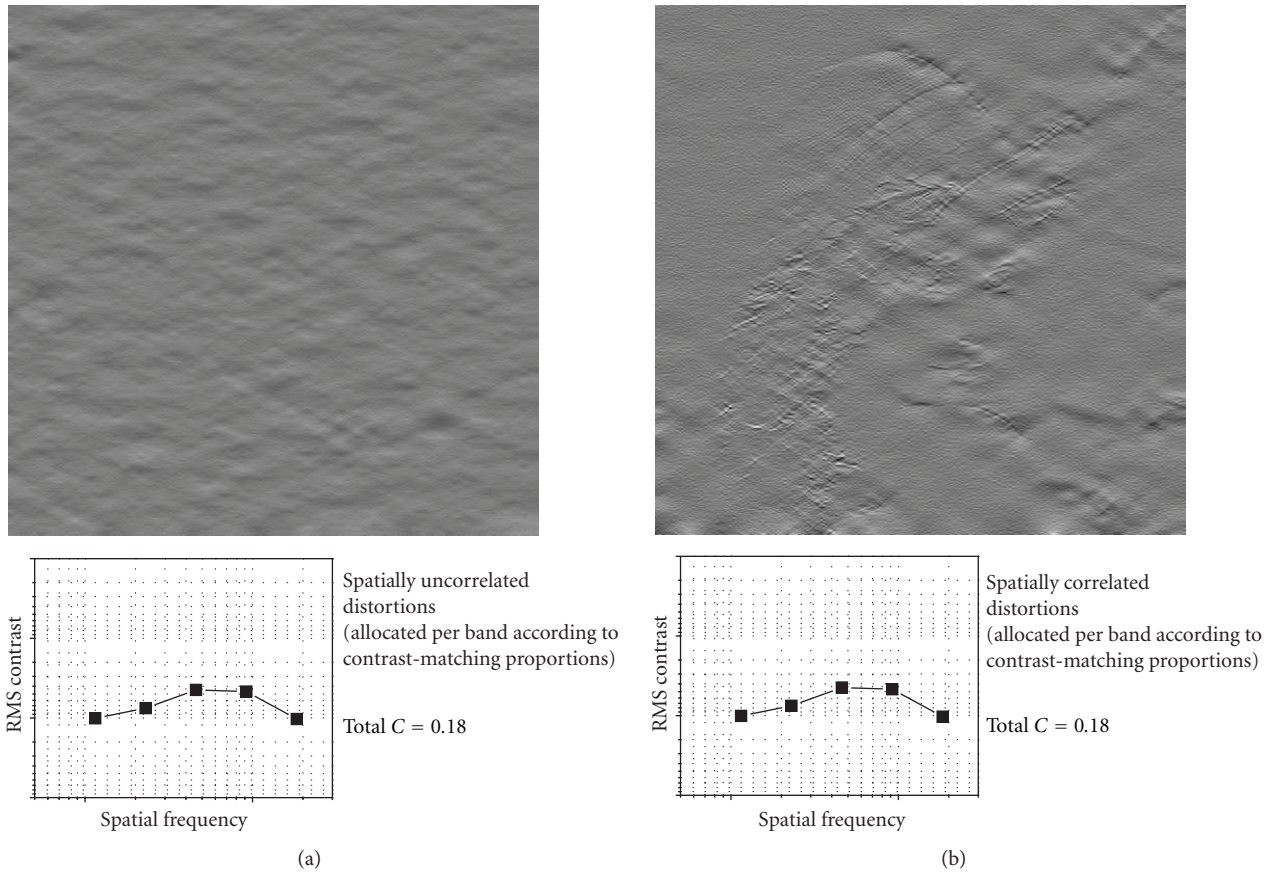


Figure 14: Distortions from Figure 13 presented against a solid gray background.

(a) Original image (b) JPEG-2000 compression DMOS = 17.7 (c) White noise DMOS = 23.5

Figure 15: When judging the quality of a distorted image containing near-threshold distortions, one tends to rely primarily on visual detection (often using point-by-point comparisons with the original image) in an attempt to locate any visible differences. (a) Closeup of original image *bikes*; (b) closeup of image *bikes* distorted via JPEG-2000 compression; (c) close-up of image *bikes* distorted by additive Gaussian white noise. DMOS values indicate differential mean opinion scores from the LIVE image database [124]. Figure from [139].

of increasing amounts of noise gave rise to images which appeared progressively less blurry. This latter result might be attributable to cross-masking whereby the noise serves to mask the blurring.

In a similar study, Kayargadde and Martens [326] employed MDS to investigate not just the potential interactions between blur and noise, but also their effects on the overall image quality. As in [325], Kayargadde and Martens reported that when a fixed amount of noise was added to images subjected to varying amounts of blur, the perceived strength of the noise was greater for the more blurry images. Similarly, they found that for highly blurred images, increasing the amount of noise served to make the images appear sharper. However, in contrast to [325],

(a) Gaussian blurring DMOS = 59.0

(b) JPEG-2000 compression DMOS = 67.1

(c) White noise DMOS = 74.6

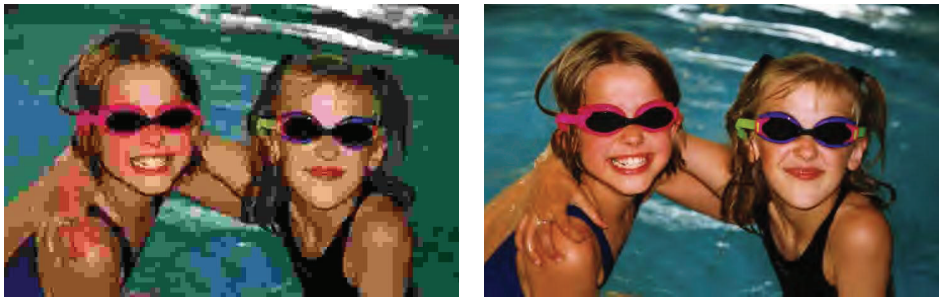
(d) JPEG-2000 compression DMOS = 82.7

FIGURE 16: When judging the quality of a distorted image containing clearly visible (suprathreshold) distortions, one tends to rely much less on visual detection and much more on overall image appearance in an attempt to recognize image content in the presence of the dominating distortions. (a) Closeup of image *bikes* distorted via Gaussian blurring; ((b) and (d)) close-up of image *bikes* distorted via JPEG-2000 compression; (c) close-up of image *bikes* distorted by additive Gaussian white noise. DMOS values indicate differential mean opinion scores from the LIVE image database [124]. Figure from [139].

Kayargadde and Martens reported that for small amounts of blur, the opposite trend was reported. For mildly blurred images, increasing the amount of noise served to make the images appear less sharp. And, for intermediate amounts of blur, the addition of noise had no effect on perceived sharpness. In terms of the effects of the blur and noise on image quality, Kayargadde and Martens reported that quality decreased for the increasing amount of blur and noise. For all of the images tested in [326], blur had a greater impact on quality than noise.

5.4.2. Joint Effects of Wavelet Distortions and Noise. As mentioned in Section 5.3, some distortions are perceived as more additive or "overlay" type distortions, whereas other distortions are more indirectly perceived based on how they affect the image's objects. In [59], we investigated how quality is affected when images were simultaneously subjected to both types of distortions. We specifically considered the case in which $e_{\text{struct}} \equiv T(x) - x$ represents distortion of the image's structure induced via disruption of the global precedence effect via quantization of wavelet subbands [54] and in which $e_{\text{noise}} \equiv n$ represents spatially uncorrelated additive white noise.

Figure 18 shows distorted versions of one of the images tested in [59]. Images in the top row contain only e_{noise}; three contrasts of e_{noise} are shown increasing from left to right (six contrasts were actually tested). Images in the leftmost column contain only e_{struct}; three contrasts of e_{struct} are shown increasing from top to bottom (six contrasts were actually tested). The remaining images contain combinations of e_{struct} and e_{noise} in the contrasts specified by their row and column headings. These combinations of distortions were tested on three natural images and—as a control condition—a solid gray image in which the distortions were generated from the image *horse* but were presented against a solid gray background (see Figure 19). For each of the distorted images, subjects were asked to rate the degradation in quality relative



(a) (b)

Figure 17: Images compressed with JPEG (a) and JPEG2000 (b), each to a compression ratio of 140 : 1.

Figure 18: Images tested in [59] containing various combinations of structural distortion and additive white noise. The structural distortion was generated via quantization of wavelet subbands with per-subband distortions selected to disrupt global precedence as described in Section 5.1 (cf. Figure 9).

to the corresponding original image (resulting in DMOS values).

Figure 20 depicts the results for each of the four images. In each graph, the horizontal axis denotes the total RMS contrast of the combined distortions and the vertical axis denotes DMOS. The data represented by the black circles in Figure 20 correspond to the condition in which the images contained only structural distortion; that is, $C_{noise} = 0$ and $C_{struct} \in [0.1, 0.3]$, where C_{noise} and C_{struct} denote the contrasts of e_{noise} and e_{struct}, respectively. The white circles in Figure 20 correspond to the condition in which the images contained only additive white noise; that is, $C_{struct} = 0$

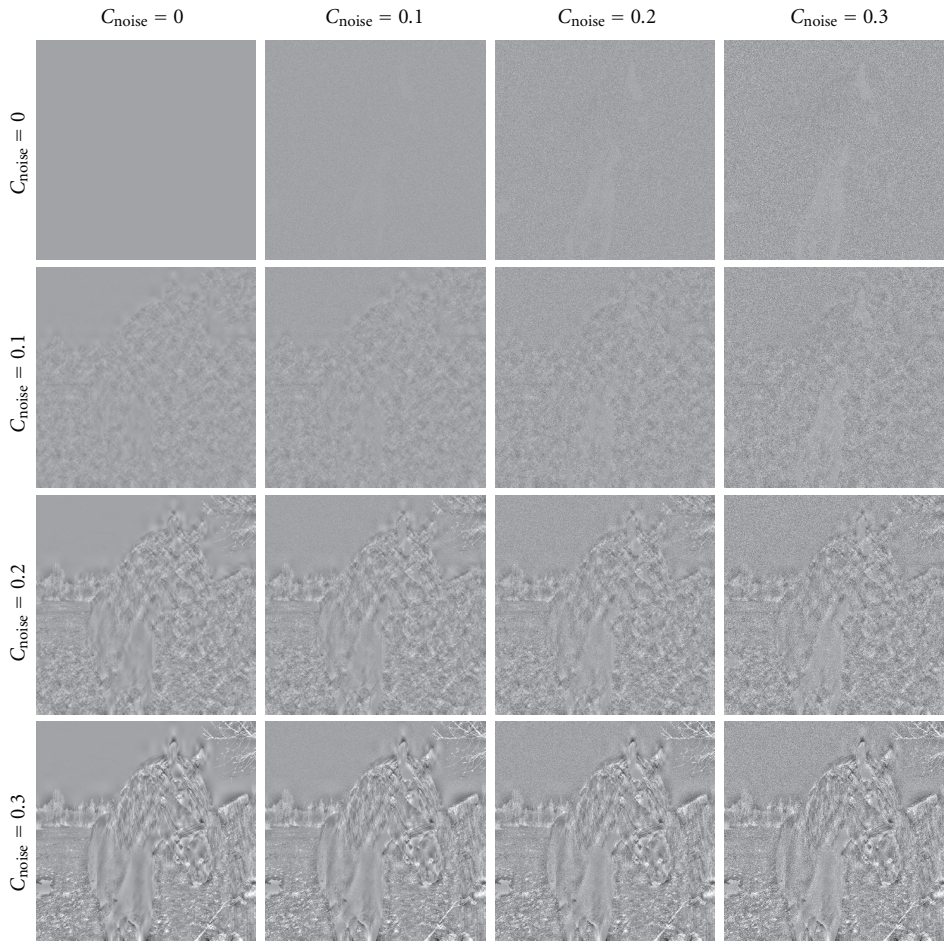


Figure 19: Images tested in [59] containing just the distortions from Figure 18 presented against a solid gray background.

and $C_{noise} \in [0.1, 0.3]$. The other symbols represent the conditions in which the images contained both e_{struct} and e_{noise}.

Notice from the trends in Figure 20 corresponding to the condition in which the images contained just e_{struct} (black circles) or just e_{noise} (white circles) that when an image contains just one type of distortion, increasing the contrast of the distortions also serves to increase the perceived distortion (decrease quality). Moreover, for the range of contrasts tested, e_{struct} gave rise to much greater perceived distortion than e_{noise}; for the three natural images, the perceived distortion induced by e_{struct} was on average 4-5 times greater than the perceived distortion induced by e_{noise}. These data provide further evidence that spatially correlated distortions which disrupt visual processing of the image's objects increase perceived distortion to a much greater extent than spatially uncorrelated additive white noise.

However, whereas RMS contrast and perceived distortion were monotonically related for images containing just e_{struct} or just e_{noise}, such a trend does not hold when both e_{struct} and e_{noise} were present. Rather, as reported in [325, 326], the trends in Figure 20 reveal that adding low-contrast e_{noise} to an image which already contains low- to mid-contrast e_{struct} actually serves to decrease perceived distortion; that is, adding small amounts of noise to a structurally distorted image serves to increase quality.

5.4.3. Modeling the Joint Effects of Multiple Distortions for IQA. One possible explanation for the interactive effects of the distortions observed in the above-mentioned studies is the *cross-masking*. Specifically, the noise might mask the structural distortion (or the blur/echo), and thus DMOS is decreased (quality increased) by adding e_{noise} to an image which also contains e_{struct} (see Figure 21). However, such cross-masking should also be present for the solid gray image tested in [59], yet the DMOS values for this image were largely unaffected by adding e_{noise} on top of e_{struct}. Another possible explanation is that the addition of noise serves to synthesize textures which were destroyed by e_{struct} (see Figure 22). This could explain why the greatest improvement in quality in [59] was observed for image *horse*, a significant portion of which is textured.

The results of the above-mentioned studies point toward an important implication for IQA: when multiple types of distortions are added to an image, the distortions can perceptually interact with each other, and with the image, in ways that may not be easily predicted based on their physical

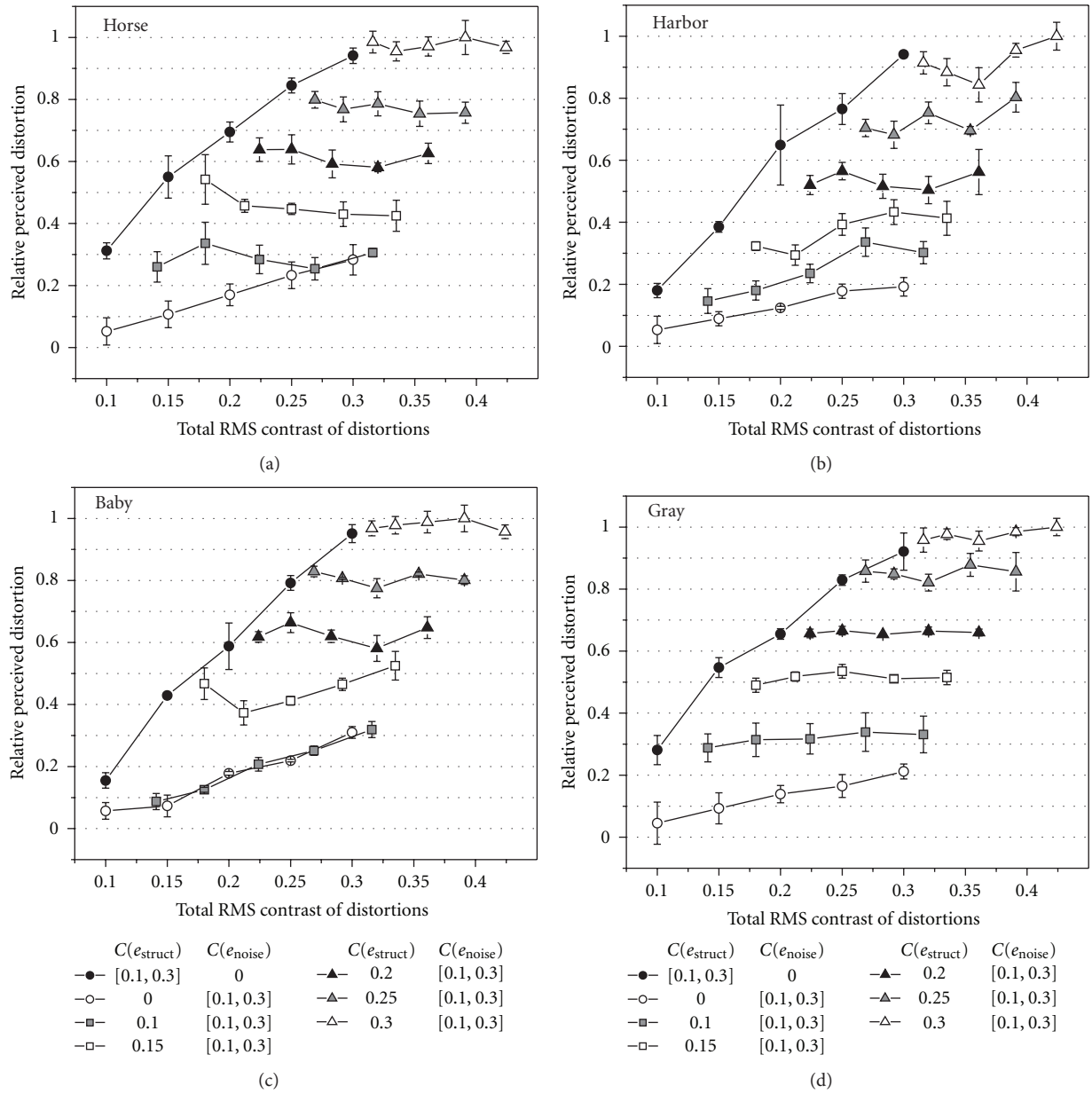


FIGURE 20: Subjective ratings of perceived distortion for images containing both structural distortion and additive white noise. In each graph, the horizontal axis denotes the total RMS contrast of the combined distortions and the vertical axis denotes perceived distortion relative to the corresponding original image; error bars denote standard errors of the means. *Black circles*: results for $C_{\text{struct}} \in [0.1, 0.3]$ and $C_{\text{noise}} = 0$. *White circles*: results for $C_{\text{struct}} = 0$ and $C_{\text{noise}} \in [0.1, 0.3]$. *Gray squares*: results for $C_{\text{struct}} = 0.1$ and $C_{\text{noise}} \in [0.1, 0.3]$. *White squares*: results for $C_{\text{struct}} = 0.15$ and $C_{\text{noise}} \in [0.1, 0.3]$. *Black triangles*: results for $C_{\text{struct}} = 0.2$ and $C_{\text{noise}} \in [0.1, 0.3]$. *Gray triangles*: results for $C_{\text{struct}} = 0.25$ and $C_{\text{noise}} \in [0.1, 0.3]$. *White triangles*: results for $C_{\text{struct}} = 0.3$ and $C_{\text{noise}} \in [0.1, 0.3]$. Note that, for the image *baby*, the contrasts are halved. Figure from [59].

combinations. IQA algorithms could certainly benefit from further research on these effects.

5.5. Challenge 5: Geometric Changes. One well-known shortcoming of the vast majority of IQA algorithms is their inability to handle geometric distortions such as translation, scaling, rotation, shearing, or changes in viewpoint. Such geometric changes, if they are not too drastic, usually have a minimal impact on visual quality. However, even slight geometric changes can give rise to massive pointwise changes in pixel intensities, and consequently, many IQA algorithms predict these geometrically modified images to be of much lower quality than indicated by the actual subjective ratings.

As an example, Figure 23 depicts five versions of an image which have been distorted via either geometric changes (spatial shift, rotation) or photometric changes (additive Gaussian white noise, change in brightness, change in contrast). All of

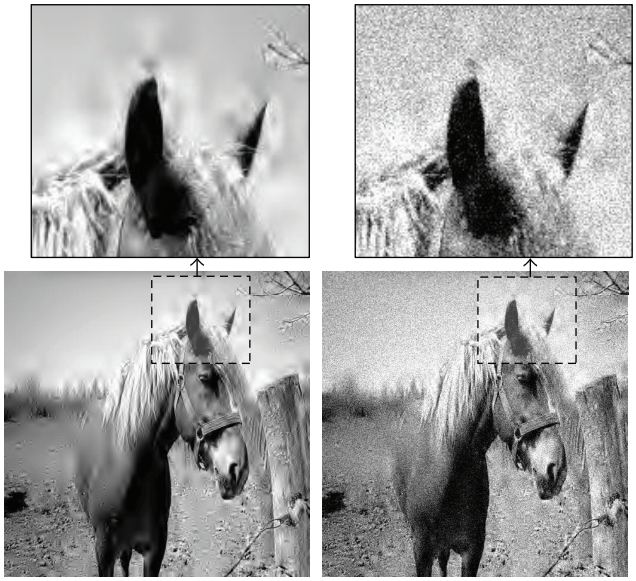


Figure 21: Closeups of two of the distorted images tested in [59] demonstrating the ability of white noise to mask the structural distortion.

Figure 22: Close-ups of two of the distorted images tested in [59] demonstrating the ability of white noise to simulate the appearance of texture that was destroyed by the structural distortion.

the images have approximately the same MSE relative to the original. Clearly, the geometric changes have a lesser impact on quality compared to the photometric changes, and the addition of white noise appears to have the greatest impact on quality.

Of course, one way to address geometric changes would be to employ a front-end stage which attempts to undo the geometric changes (e.g., via image registration). However, such an approach will not work in general, particularly when the type of geometric change is unknown, when multiple geometric changes are imposed, and when geometric changes are compounded with more traditional distortions such as noise or JPEG compression artifacts.

As IQA moves into more mainstream applications, the need to handle geometric changes has become increasingly apparent. One prime example is the comparison of photographs of the same scene taken with different cameras at slightly different viewpoints. Another example is when applying IQA on a frame-by-frame basis to assess the quality of video. If the frames of the distorted video become temporally unsynchronized with the frames of the reference video (e.g., delayed by a few frames), then the difference between the reference and distorted frames will typically manifest in the form of geometric changes due to movement of the subject matter and/or panning, zooming, and other viewpoint changes imposed by the camera.

5.5.1. Perception and IQA of Geometric Changes. One theory of why geometric changes have minimal impact on visual quality is that geometric changes are quite prevalent during normal vision, and thus the human visual system has adapted to become relatively insensitive to such changes. This theory was proposed in [327] by Kingdom et al. based on the results of a psychophysical experiment in which discrimination thresholds were measured for images containing various geometric distortions (affine transforms) and for images containing photometric distortions (luminance changes, contrast changes, and various forms of noise). Kingdom et al. found that subjects were 11–14x more sensitive to noise than they were to geometric distortions and were 2-3x more sensitive to brightness/contrast changes than they were to geometric distortions. Based on these findings, Kingdom et al. suggested that "*observers are least sensitive to those transformations most commonly experienced in the natural world.*" [327].

Some work in visual perception and IQA has specifically focused on addressing geometric distortion. For example, in [328], Chow et al. compared the detectability of local warping distortions in computer-generated scenes on monitors versus head-mounted displays. In [329], Rovamo et al. measured thresholds for detecting geometric distortion in faces as a function of retinal eccentricity. For IQA of watermarked images, Setyawan et al. [330] investigated the impact of particular forms of geometric distortion on the perceived quality of watermarked images. Setyawan et al. also presented in [330] an FR IQA algorithm which uses local estimates of affine transformations to estimate quality.

In [331], Wang and Simoncelli presented an FR IQA algorithm which can handle minor translations, rotations, and scalings. Their algorithm, CW-SSIM, is an extension of SSIM [123] which uses complex wavelets to achieve its invariance to these geometric changes. Quality is assessed via SSIM-type measures applied separately to the magnitudes and phases of the coefficients. Invariance to translation is afforded by the fact that translation in the spatial domain manifests in the coefficients as concerted shifts in phase. Wang and Simoncelli demonstrated the utility of CW-SSIM both for IQA of geometrically distorted images and for character recognition. In [176], CW-SSIM was also shown to perform

Figure 23: An original image and five versions of the image distorted via either geometric changes (spatial shift, rotation) or photometric changes (additive Gaussian white noise, change in brightness, change in contrast). All five distorted images have approximately the same MSE of 1500 relative to the original.

well on more general similarity tasks such as comparing segmentations and quantifying similarities between 3D facial surfaces.

In [332], D'Angelo et al. presented an FR IQA algorithm which uses an HVS model and displacement fields for QA of geometrically distorted images. Their technique applies a single-level, multiorientation Gabor decomposition with both even- and odd-symmetric filters to the reference and distorted images. The even- and odd-symmetric responses are combined via an L_2-norm to mimic the responses of complex cells in V1. These latter responses are then combined with local gradient information obtained from a displacement field, and the collection of modified responses are collapsed across orientation and space to arrive at a scalar estimate of quality. D'Angelo et al. demonstrated that, on a database of geometrically distorted images, their technique can yield quality predictions that correlate well with subjective ratings; Spearman's and Pearson's correlation coefficients of approximately 0.8 were reported.

5.5.2. More Radical Geometric Changes. Beyond basic affine transformations, it is also possible to generate images with more radial geometric changes and combinations of geometric and photometric changes. One particular area which researchers have begun to explore is IQA of textures. Given two samples of textures, humans can readily determine whether the two samples were taken from the same source texture (the same material). Or, given a database of actual and synthesized textures, humans can assign consistent quality ratings to each synthesized version relative to its corresponding original. From a computational standpoint, however, this task is extremely challenging. A major factor which complicates IQA of textures is the fact that point-by-point comparisons, a common approach used to some extent by most IQA algorithms, cannot be used to compare the visual similarity of two textures.

Although a great deal of human vision research has been conducted to investigate the perceptual and neural mechanisms which underlie the visual appearance of texture (see [333] for a review), further research is needed on how to actually apply these findings to IQA of textures. In [334], Bénard et al. investigated the effects of fractalization on the visual quality of synthesized textures; they reported that the average cooccurrence error between gray-level cooccurrence matrices measured for the original and fractalized textures can perform well in predicting the subjective ratings. In [335], Zujovic et al. addressed IQA of textures by designing a structural similarity index for texture retrieval. In [336], Zujovic et al. also demonstrated the utility of their index for texture-synthesis-based image compression.

Also on the topic of IQA for synthesized textures, in [337] we presented a preliminary database of original and synthesized textures and associated DMOS values. Forty-two textures from the Brodatz database [338] served as originals, and various texture-synthesis algorithms were used to generate the synthesized versions. Figure 24 shows some of the original and synthesized textures used in [337]. An examination of the quality ratings values revealed that the most detrimental artifacts were, (1) lack of structural details, (2) misalignment of the texture patterns, (3) blurring introduced in the texture patterns, and (4) tiling of the same pattern. In [337], we also demonstrated that a weighted geometric combination of KLD and parameters from Portilla and Simoncelli's parametric texture-synthesis algorithm

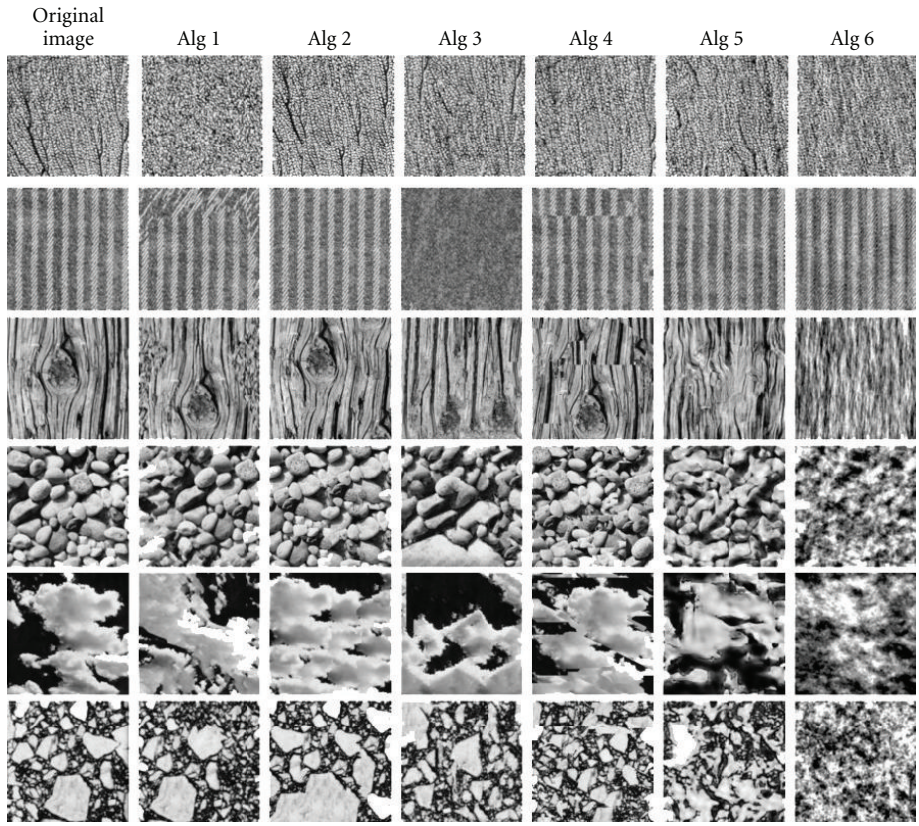


Figure 24: Examples of some of the original and synthesized textures (from six texture-synthesis algorithms) for which quality ratings were collected in [337].

[339] showed promise in predicting the ratings. However, there were many notable failure cases, particularly on textures containing more structured objects (e.g., flowers, stones).

With the emergence of applications which can potentially give rise to images containing such radical changes in geometric properties (e.g., texture-synthesis coding and object-based coding), there is clearly a need for further research in this area, both in terms of visual perception and in terms of associated IQA algorithms and databases.

5.6. Challenge 6: Enhanced Images and Aesthetic Quality. Today, the photo-editing software used in digital photography has replaced the role of the darkroom used in traditional photography. Through digital enhancement or "retouching," it is possible for an altered image to surpass the visual quality of the original image. For video, enhancements such as motion sharpening and anti-judder processing found in newer displays are known to make profound differences in quality.

Yet, the vast majority of IQA algorithms have been designed for *distorted* images, often operating under the assumption that a high-quality image is one which is most visually similar to the original (reference) image. However, for enhanced images, the notion of similarity is less applicable and a different QA tactic is needed. IQA of enhanced images remains an open research challenge.

5.6.1. Image Enhancement and Other Applications. Image enhancement is one of the most fundamental operations in image processing and digital photography. Enhancement/retouching is a required step that every professional photographer performs after capturing photos. Although there is no standard rule to follow when editing a scene, most photographers implement several steps such as (1) cropping the images for recomposition; (2) removing obstructions or unwanted objects; (3) applying noise-reduction techniques, if needed; (4) adjusting brightness and contrast; (5) white-balancing and color-correction; (6) sharpening, usually as the final step. These and other forms of processing (e.g., demosaicing, superresolution, computational photography) yield images that are dissimilar, but are most often of superior visual quality compared to the original images.

As mentioned previously, the vast majority of IQA methods have been designed for distorted images. IQA of enhanced images is challenging due to the fact that the changes can often be subtle and can affect the artistic impression of the image. Nonetheless, it is still possible to perform QA of enhanced images based on changes in low-level attributes. The work of Fairchild and Johnson [340] begins

to address this issue by using contrast- and color-appearance models. In addition, VIF [195] can yield a value larger than unity (denoting quality greater than the original) when the "distorted" image contains linear contrast enhancement.

Beside image enhancement, other image-processing applications can give rise to images of greater visual quality than the unprocessed images. For example, in denoising and/or artifact reduction, the primary objective is to yield an image of superior visual quality to the input (distorted) image. The performances of such algorithms are quantified by comparing the denoised/artifact-reduced image to the original, pristine image, where the comparison is made via PSNR or some FR IQA algorithm. However, such a comparison is not always valid because the denoised/artifact-reduced image often undergoes processing such as sharpening, contrast-stretching, and inpainting. Again, our current IQA algorithms have not been designed to handle such changes. A similar argument applies to dithering algorithms; see [107]. These and related applications could certainly benefit from IQA algorithms designed for such changes; proper IQA could not only be used to quantify performance, but they could also provide criteria for parameter optimization.

5.6.2. A Database for IQA of Enhanced Images. One of the main roadblocks in IQA of enhanced images is the lack of a database containing enhanced images and associated quality ratings. To address this issue, in [341], we presented such a database, the DRIQ (digitally retouched image quality) database. DRIQ contains 26 reference images of size 512×512 pixels obtained from the Kodak and CSIQ databases. Each of these 26 images was manually retouched to generate three enhanced versions spanning varying amounts of quality. The enhancements were made by editing either contrast, sharpness, brightness, color saturation, or combinations of these properties, both globally and locally. In total, the database contains 104 images (26 original images and $26 \times 3 = 78$ enhanced versions of these originals).

Obtaining reliable ratings of quality for enhanced images is more difficult than that for degraded images due to the fact that the changes are often subtle. Based on several pilot experiments, we employed a three-step procedure: (1) intraimage ranking via a pairwise-comparison procedure, then (2) Intraimage rating constrained by the ranks via a multiple-stimulus continuous quality evaluation (MSCQE) procedure, and then (3) across-image ratings constrained by the within-image ratings, again via an MSCQE paradigm. See [341] for further details of the experiments. Figure 25 shows some examples of the enhanced images and associated DMOS values; the entire database and ratings are available online [140].

Figure 26 shows the absolute best- and worst-rated images in the database along with their corresponding original reference images. The original image *flower* appears low in contrast, sharpness, and colorfulness. Its enhanced version, which received the highest rating, was enhanced in contrast, sharpened, and locally color-corrected (for the stamen of the flower and the leaves). However, the image *redwood* in the second row has little room for enhancement, and thus this image was enhanced only in terms of contrast; other enhanced versions of this image received similarly low DMOS values.

One very consistent finding which we observed when viewing an original image and its enhanced version is that the enhanced image makes the original image appear degraded (as long as the enhanced image is not overenhanced such that it appears artificial). Given this fact, in [341], we asked whether it is possible to perform IQA of enhanced images by operating existing distortion-based FR IQA methods in reverse and then reinterpreting the results. Specifically, given a reference image and its enhanced version, the reference image can be thought of as a distorted version of the enhanced image. Thus, existing FR IQA algorithms may be able to perform IQA of enhanced images by specifying the enhanced image as the reference image and by specifying the original image as the distorted image.

In [341], we showed that this reverse-mode-based approach, when supplemented with global measures of sharpness, contrast, and saturation, can yield quality estimates which correlate decently with the DMOS values (Pearson's CC of approximately 0.85). These preliminary findings suggest that a future IQA algorithm designed specifically for enhanced images may benefit from a strategy that involves both enhanced-feature measurements and statistical comparisons of local frequency coefficients.

Although the DRIQ database represents an important first step, it is also important to note that a correlation of approximately 0.85 still leaves much room for improvement. Furthermore, none of the images in DRIQ contain both distortions and enhancement, none of the images contain overenhancement (which will certainly result in lower quality ratings), and none of the images were enhanced via more sophisticated recomposition, superresolution, or computational photography techniques.

5.6.3. Image Enhancement and Aesthetic Quality. At the 2008 SPIE Human Vision and Electronic Imaging Conference, Scott Daly provided very insightful perspectives on IQA of enhanced images in his talk "On the Role of Artistic Intent of Image Quality." In his presentation, Daly clearly demonstrated how many standard enhancement techniques, which would normally improve image quality, can severely degrade aesthetic quality. Daly's main argument was based on the fact that artists and photographers often produce images which capture and convey specific visual impressions, images which make specific visual statements. Because traditional forms of enhancement can destroy the artist's intentions, even an IQA algorithm that can handle normal enhancement still has a long way to go in order to be successful in predicting the aesthetic quality.

Developing an IQA algorithm which can handle enhanced images, while considering the effects on artistic intent and aesthetic quality, certainly remains an unsolved research challenge.

5.7. Challenge 7: Runtime Performance. Although a great deal of research on IQA has focused on improving prediction

Figure 25: Examples of original and enhanced images from the DRIQ database [140, 341]. The DMOS values denote differences in perceived quality relative to each corresponding original and relative to all other enhanced images; larger DMOS denotes greater perceived quality.

accuracy, much less research has addressed algorithmic and microarchitectural efficiency. As IQA algorithms move from the research environment into more mainstream applications, issues surrounding efficiency—such as execution speed and memory bandwidth requirements—begin to emerge as equally important performance criteria. Many IQA algorithms which excel in terms of prediction accuracy fall short in terms of efficiency, often requiring relatively large memory footprints and runtimes on the order of seconds for even modest-sized images (e.g., <1 MPixels). As these algorithms are adapted to process frames of video (e.g., [342, 343]) or are used during optimization procedures (e.g., during RD optimization in a coding context), efficiency becomes of even greater importance.

From a signal-processing viewpoint, it would seem that the bulk of computation and runtime are likely to occur in two key stages employed by many IQA algorithms: (1) local frequency-based decompositions of the input image(s) and (2) local statistical computations on the coefficients. The first of these two stages can potentially require a considerable amount of computation and memory bandwidth, particularly when a large number of frequency bands are analyzed and when the decomposition must be applied to each image as a whole. The latter of these two stages would seem to require more computation, particularly when multiple statistical computations are required for each local region of coefficients. For example, in VIF [195], wavelet subband covariances can be computed via a block-based or overlapping block-based approach. In MAD [139], variances, skewness, and kurtoses of log-Gabor coefficients are also computed for overlapping blocks in each subband. As described in Section 3, many other HVS-based IQA methods have been designed to mimic the cortical processing in the HVS in which the local responses of neurons in V1 (modeled as coefficients) are computed and compared. Yet, unlike the HVS, most modern computing machines lack dedicated hardware for such computation.

5.7.1. Acceleration of Image Transforms and Local Statistics. Due to their extensive use in image compression and computer vision, a considerable amount of research has focused on accelerating two-dimensional image transforms which provide local frequency-based decompositions. For example, the discrete cosine transform (DCT) has been accelerated at the algorithm level by using variations of the same techniques used in the FFT (e.g., [344]) and by exploiting various other algebraic and structural properties of the transform, for example, via recursion [345], lifting [346], matrix factorization [347], cyclic convolution [348], and many other techniques (see [349] for a review). Numerous techniques for the hardware-based acceleration of the DCT have also

(a) (b)

(c) (d)

Figure 26: First row: the original image *flower* on the left and its enhanced version which received the absolute best subjective rating (DMOS = 1). Second row: the original image *redwood* on the left and its enhanced version which received the absolute worst subjective rating (DMOS = 0.17).

been proposed using general-purpose GPU (GPGPU) and FPGA implementations (e.g., [350–353]). Algorithm- and hardware-based acceleration has also been researched for the discrete wavelet transform (e.g., [354–356]) and the Gabor transform (e.g., [357–360]).

Techniques for accelerating the computation of local statistics in images have also been researched, though to a much lesser extent than the transforms. One technique, called *integral images*, which was originally developed in the context of computer graphics [361], has emerged as a popular approach for computing block-based sums of any two-dimensional matrix of values (e.g., a matrix of pixels or coefficients). The integral image, also known as the *summed area table*, requires first computing a table which has the same dimensions as the input matrix and in which each value in the table represents the sum of all matrix values above and to the left of the current position. Thereafter, the sum of values within any block of the matrix can be rapidly computed via addition/subtraction of three values in the table. A similar technique can be used to compute higher-order moments such as the variance, skewness, and kurtosis (see, e.g., [362, 363]).

5.7.2. Acceleration of Specific IQA Algorithms. Other researchers have investigated techniques for accelerating specific IQA algorithms. For example, in [364], Gordon et al. investigated the acceleration of PSNR by using GPGPU implementations in both CUDA and OpenGL. Via a performance analysis, they specifically investigated how the application and system performance is affected by utilizing GPGPU acceleration of PSNR in a model-based coding application (the primary bottleneck in model-based coding stems from the optimization procedure used to determine the model parameters from the input image). Gordon et al. concluded that because CUDA uses the CPU's store units to copy data between the graphics card and system memory and because data exchanged between GPU APIs travel through the main processor, a non-GPGPU implementation of the

PSNR computation runs faster than the same implementation using GPGPU programming methods.

In [365], Chen and Bovik presented the Fast SSIM and Fast MS-SSIM algorithms, which are accelerated versions of SSIM and MS-SSIM, respectively. Three modifications were used for Fast SSIM: (1) the luminance component of each block was computed by using an integral image, (2) the contrast and structure components of each block were computed based on 2×2 Roberts gradient operators, (3) the Gaussian-weighting window used in the contrast and structure components was replaced with an integer approximation. For Fast MS-SSIM, a further algorithm-level modification of skipping the contrast and structure computations at the finest scale was proposed. By using these modifications, Fast SSIM and Fast MS-SSIM were shown to be, respectively, 2.7x and 10x faster than their original counterparts on 768×432 frames from videos of the LIVE Video Quality database. Although algorithm-level modifications were used, the authors demonstrated that these modifications had negligible impact on predictive performance; testing on the LIVE Image Quality and Video Quality databases revealed effectively no impact on SROCC, CC, and RMSE. By further implementing the calculations of the contrast and structure components via Intel SSE2 (SIMD) instructions, speedups of approximately 5x for Fast SSIM and 14x for Fast MS-SSIM were reported. In addition, speedups of approximately 17x for Fast SSIM and 50x for Fast MS-SSIM were reported by further employing parallelization via a multithreaded implementation.

In [366], Okarma and Mazurek presented GPGPU techniques for accelerating SSIM, MS-SSIM, and CVQM (a video quality assessment algorithm developed previously by Okarma, which uses SSIM, MS-SSIM, and VIF to estimate quality). To accelerate the computation of both SSIM and MS-SSIM, the authors described a CUDA-based implementation in which separate GPU threads were used to compute SSIM or MS-SSIM on strategically sized fragments of the image. To overcome CUDA's memory-bandwidth limitations, the computed quality estimates for the fragments were stored in GPU registers and transferred only once to the system memory. Okarma and Mazurek reported that their GPGPU-based implementations resulted in 150x and 35x speedups of SSIM and MS-SSIM, respectively.

In [363], Phan et al. presented the results of a performance analysis and techniques for accelerating the MAD algorithm [139]. Although MAD is among the best in predictive performance, it is currently the one of the slowest IQA algorithms, requiring over 55 seconds for a 512×512 image when tested on several modern computers (Intel Core 2 and Xeon CPUs; see [363]). A performance analysis revealed that the main bottleneck in MAD stemmed from its appearance-based stage, which accounted for 98% of the total runtime. Within this appearance-based stage, the computation of the local statistical differences accounted for most of the runtime, and computation of the log-Gabor decomposition accounted for the bulk of the remainder. Phan et al. proposed four techniques of acceleration: (1) using integral images for the local statistical computations, (2) using procedure expansion and strength reduction, (3) using a GPGPU implementation of the log-Gabor decomposition, and (4) precomputation and caching of the log-Gabor filters. The first two of these modifications resulted in an approximately 17x speedup over the original MAD implementation. The latter two resulted in an approximately 47x speedup over the original MAD implementation.

5.7.3. The Need for Broader Performance Analyses. Although the aforementioned studies have successfully yielded more efficient versions of their respective algorithms, several larger questions remain unanswered. To what extent are the bottlenecks in IQA algorithms attributable to the decomposition and statistical computation stages versus more algorithm-specific auxiliary computations? To what extent are the bottlenecks attributable to computational complexity versus limitations in memory bandwidth? Are there generic implementation techniques or microarchitectural modifications that can be used to accelerate all or at least several IQA algorithms?

The answers to these questions cannot only provide important insights for furthering IQA research, but they can also facilitate deployment and integration of IQA algorithms into existing and forthcoming applications and platforms. Further research in this area can help guide, (1) the design of new IQA algorithms, which are likely to draw on multiple approaches used in several existing IQA algorithms, (2) efficient implementations of multiple IQA algorithms on a given hardware platform, (3) efficient integration of multiple IQA algorithms in specific applications, and (4) the selection and/or design of specific hardware which can efficiently execute multiple IQA algorithms.

6. Conclusion

In this paper, I have discussed the current state of the art in IQA research, summarized the IQA-related knowledge that has been gained from studies in visual psychophysics, and surveyed the progress that has been made through HVS modeling, non-HVS-based approaches, and other statistical modeling approaches. One main conclusion which can be drawn from this paper is that today's IQA algorithms can perform remarkably well at predicting human judgments of quality.

However, it should also be evident that the IQA problem is far from being solved; we have yet to reach the summit of this investigative ascent. Rather, it may seem that our current accomplishments have shown that we can design IQA algorithms to handle images generated by the more traditional applications—compression, watermarking, transmission errors, and camera/display processing—and other applications which have served as catalysts for IQA research. But, further research is needed to improve IQA based on the current challenges and to prepare IQA for future challenges.

Here, I have identified seven open challenges in IQA. The objective of this discussion was not only to highlight the limitations in our current knowledge of image quality, but to also emphasize the fact that there is substantial room for alternative theories and techniques beyond those surveyed

here. Budrikis' original 1972 statement that "full evaluations are as yet impossible" [23] still holds true today. However, IQA research continues to grow at an astounding rate, and these efforts will undoubtedly lead to improved evaluation techniques and further advancements in our understanding of image quality.

Acknowledgments

This material is based upon work supported by, or in part by, the National Science Foundation Awards 0917014 and 1054612 and by the U.S. Army Research Laboratory (USARL) and the U.S. Army Research Office (USARO) under Contract/Grant number W911NF-10-1-0015.

References

[1] C. J. B. Lambrecht, "Working spatio-temporal model of the human visual system for image restoration and quality assessment applications," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '96)*, pp. 2291–2294, May 1996.

[2] Z. Wang, A. C. Bovik, and L. Lu, "Wavelet-based foveated image quality measurement for region of interest image coding," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '01)*, pp. 89–92, grc, October 2001.

[3] K. Yang and H. Jiang, "Optimized-ssim based quantization in optical remote sensing image compression," in *Proceedings of the 6th International Conference on Image and Graphics (ICIG '11)*, pp. 117–122, 2011.

[4] J. Huang and Y. Q. Shi, "Adaptive image watermarking scheme based on visual masking," *Electronics Letters*, vol. 34, no. 8, pp. 748–750, 1998.

[5] M. Masry, D. Chandler, and S. S. Hemami, "Digital watermarking using local contrast-based texture masking," in *Conference Record of the 37th Asilomar Conference on Signals, Systems and Computers*, pp. 1590–1595, November 2003.

[6] I. G. Karybali and K. Berberidis, "Efficient spatial image watermarking via new perceptual masking and blind detection schemes," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 2, pp. 256–274, 2006.

[7] M. Liu and X. Yang, "A new image quality approach based on decision fusion," in *Proceedings of the 4th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD '08)*, pp. 10–14, October 2008.

[8] A. Koz and A. A. Alatan, "Oblivious spatio-temporal watermarking of digital video by exploiting the human visual system," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 3, pp. 326–337, 2008.

[9] T. T. Lam, L. J. Karam, and G. P. Abousleman, "Robust image coding using perceptually-tuned channel-optimized trellis-coded quantization," in *Proceedings of the IEEE 42nd Midwest Symposium on Circuits and Sistems*, vol. 2, pp. 1131–1134, August 1999.

[10] A. Rehman, M. Rostami, Z. Wang, D. Brunet, and E. R. Vrscay, "Ssiminspired image restoration using sparse representation," *EURASIP Journal on Advances in Signal Processin*, vol. 2012, p. 16, 2012.

[11] J. A. Ferwerda, "Fundamentals of spatial vision," in *Applications of Visual Perception in Computer Graphics*, V. Interrante, Ed., SIGGRAPH, pp. 1–27, 1998.

[12] B. Walter, S. N. Pattanaik, and D. P. Greenberg, "Using perceptual texture masking for efficient image synthesis," *Computer Graphics Forum*, vol. 21, no. 3, pp. 393–399, 2002.

[13] F. Ciaramello, A. Cavender, S. Hemami, E. Riskin, and R. Ladner, "Predicting intelligibility of compressed american sign language video with objective quality metrics," in *Proceedings of the International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, 2006.

[14] J. Horton, "The electrical transmission of pictures and images," *Proceedings of the Institute of Radio Engineers*, vol. 17, no. 9, pp. 1540–1563, 1929.

[15] L. Jesty and G. Winch, "Television images: an analysis of their essential qualities," *Translate Illum to English*, vol. 2, pp. 316–334, 1937.

[16] P. Goldmark and J. Dyer, "Quality in television pictures," *Proceedings of the Institute of Radio Engineers*, vol. 28, no. 8, pp. 343–350, 1940.

[17] G. Winch, "Colour television: some subjective and objective aspects of colour rendering," *Proceedings of the IEE*, vol. 99, no. 20, Part 3, pp. 854–860, 1952.

[18] L. Jesty, "Television as a communication problem," *Electrical Engineers, Journal of the Institution*, vol. 1953, no. 4, pp. 181–183, 1953.

[19] P. B. Fellgett and E. H. Linfoot, "On the assessment of optical images," *Philosophical Transactions of the Royal Society of London*, vol. 247, no. 931, pp. 369–407, 1955, Mathematical and Physical Sciences.

[20] C. Dean, "Measurements of the subjective effects of interference in television reception," *Proceedings of the Institute of Radio Engineers*, vol. 48, no. 6, pp. 1035–1049, 1960.

[21] H. Schmid, "Measurement of television picture impairments caused by linear distortions," *Journal of the SMPTE*, vol. 77, no. 3, pp. 215–220, 1968.

[22] D. Sakrison and V. Algazi, "Comparison of line-by-line and twodimensional encoding of random images," *IEEE Transactions on Information Theory*, vol. 17, no. 4, pp. 386–398, 1971.

[23] Z. Budrikis, "Visual fidelity criterion and modeling," *Proceedings of the IEEE*, vol. 60, no. 771, 779 pages, 1972.

[24] T. G. Stockham Jr., "Image processing in the context of a visual model," *Proceedings of the IEEE*, vol. 60, no. 7, pp. 828–842, 1972.

[25] O. Schade, *Image Quality: A Comparison of Photographic and Television Systems*, RCA Laboratories, 1975.

[26] J. L. Mannos and D. J. Sakrison, "The effects of a visual fidelity criterion on the encoding of images," *IEEE Transactions on Information Theory*, vol. 20, no. 4, pp. 525–536, 1974.

[27] E. C. Larson and D. M. Chandler, "Categorical subjective image quality CSIQ database," 2009, http://vision.okstate.edu/csiq/.

[28] R. J. Beaton, "Quantitative models of image quality," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 27, no. 1, pp. 41–45, 1983.

[29] A. J. Ahumada Jr., "Computational image quality metrics: a review," in *Proceedings of the SID International Symposium Digest of Technical Papers*, vol. 24, pp. 305–308, Playa del Rey, Calif, USA, 1993.

[30] M. P. Eckert and A. P. Bradley, "Perceptual quality metrics applied to still image compression," *Signal Processing*, vol. 70, no. 3, pp. 177–200, 1998.

[31] B. Keelan, *Handbook of Image Quality: Characterization and Prediction*, Optical engineering, Taylor & Francis, 2002.

[32] K. Seshadrinathan, T. Pappas, R. Safranek et al., "Image quality assessment," in *The Essential Guide to Image Processing*, A. Bovik, Ed., Electronics and Electrical, Academic Press, 2009.

[33] S. S. Hemami and A. R. Reibman, "No-reference image and video quality estimation: applications and human-motivated design," *Image Communication*, vol. 25, no. 7, pp. 469–481, 2010.

[34] K. Seshadrinathan and A. C. Bovik, "Automatic prediction of perceptual quality of multimedia signals-a survey," *Multimedia Tools and Applications*, vol. 51, no. 1, pp. 163–186, 2011.

[35] W. Lin and C. C. J. Kuo, "Perceptual visual quality metrics: a survey," *Journal of Visual Communication and Image Representation*, vol. 22, no. 4, pp. 297–312, 2011.

[36] A. C. Bovik, "Automatic prediction of perceptual image and video quality," *Proceedings of the IEEE*. In press.

[37] S. Winkler, "Video quality measurement standards: current status and trends," in *Proceedings of the 7th International Conference on Information, Communications and Signal Processing (ICICS '09)*, pp. 848–852, IEEE Press, Piscataway, NJ, USA, 2009.

[38] S. Chikkerur, V. Sundaram, M. Reisslein, and L. J. Karam, "Objective video quality assessment methods: a classification, review, and performance comparison," *IEEE Transactions on Broadcasting*, vol. 57, no. 2, pp. 165–182, 2011.

[39] M. Fairchild, *Color Appearance Models*, Addison-Wesley, 1998.

[40] R. Janssen, *Computational Image Quality*, Spie Press Monograph, SPIE Press, 2001.

[41] M. Fairchild, *Color Appearance Models*, Imaging Science and Technology, John Wiley & Sons, 2005.

[42] H. H. Barrett, J. Yao, J. P. Rolland, and K. J. Myers, "Model observers for assessment of image quality," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 90, no. 21, pp. 9758–9765, 1993.

[43] A. Ahumeda, "Simplified vision models for image-quality assessment," in *SID International Symposium Digest of Technical Papers*, vol. 27, pp. 397–402, Society for Information Display, 1996.

[44] A. B. Watson, M. Taylor, and R. Borthwick, "Image quality and entropy masking," in *Human Vision, Visual Processing, and Digital Display VIII*, vol. 3016 of *Proceedings of SPIE*, pp. 2–12, 1997.

[45] A. B. Watson, G. Y. Yang, J. A. Solomon, and J. Villasenor, "Visibility of wavelet quantization noise," *IEEE Transactions on Image Processing*, vol. 6, no. 8, pp. 1164–1175, 1997.

[46] D. Gur, D. A. Rubin, B. H. Kart et al., "Forced choice and ordinal discrete rating assessment of image quality: a comparison," *Journal of Digital Imaging*, vol. 10, no. 3, pp. 103–107, 1997.

[47] H. de Ridder, "Psychophysical evaluation of image quality: from judgment to impression," in *Human Vision and Electronic Imaging III*, B. E. Rogowitz and T. N. Pappas, Eds., vol. 3299 of *Proceedings of SPIE*, pp. 252–263, 1998.

[48] S. Winkler, "Issues in vision modeling for perceptual video quality assessment," *Signal Processing*, vol. 78, no. 2, pp. 231–252, 1999.

[49] M. J. Nadenau and J. Reichel, "Image compression related contrast masking measurements," in *Proceedings of the Human Vision and Electronic Imaging V*, B. E. Rogowitz and T. N. Pappas, Eds., vol. 3959, pp. 188–199, January 2000.

[50] M. G. Ramos and S. S. Hemami, "Suprathreshold wavelet coefficient quantization in complex stimuli: Psychophysical evaluation and analysis," *Journal of the Optical Society of America A*, vol. 18, no. 10, pp. 2385–2397, 2001.

[51] J. Martens and M. Boschmann, "The psychophysical measurement of image quality," in *Vision Models and Applications to Image and Video Processing*, 2001.

[52] S. Winkler, "Visual fidelity and perceived quality: towards comprehensive metrics," in *Human Vision and Electronic Imaging VI*, B. E. Rogowitz and T. N. Pappas, Eds., Proceedings of SPIE, 2001.

[53] D. M. Chandler and S. S. Hemami, "Additivity models for suprathreshold distortion in quantized wavelet-coded images," in *Human Vision and Electronic Imaging VII*, B. E. Rogowitz and T. N. Pappas, Eds., Proceedings of SPIE, pp. 105–118, San Jose, Calif, USA, January 2002.

[54] "Suprathreshold image compression based on contrast allocation and global precedence," in *Human Vision and Electronic Imaging VIII*, B. E. Rogowitz and T. N. Pappas, Eds., Proceedings of SPIE, Santa Clara, Calif, USA, 2003.

[55] "Effects of natural images on the detectability of simple and compound wavelet subband quantization distortions," *Journal of the Optical Society of America A*, vol. 20, no. 7, pp. 1164–1180, 2003.

[56] S. Süsstrunk and S. Winkler, "Color image quality on the internet," in *Internet Imaging V*, Proceedings of SPIE, pp. 118–131, January 2004.

[57] Z. Wang and E. P. Simoncelli, "Stimulus synthesis for efficient evaluation and refinement of perceptual image quality metrics," in *Human Vision and Electronic Imaging IX*, vol. 5292 of *Proceedings of SPIE*, pp. 99–108, January 2004.

[58] S. Winkler and S. Süsstrunk, "Visibility of noise in natural images," in *Human Vision and Electronic Imaging IX*, Proceedings of SPIE, pp. 121–129, January 2004.

[59] D. M. Chandler, K. H. Lim, and S. S. Hemami, "Effects of spatial correlations and global precedence on the visual fidelity of distorted images," in *Human Vision and Electronic Imaging XI*, B. E. Rogowitz, T. N. Pappas, and S. Daly, Eds., Proceedings of SPIE, San Jose, Calif, USA, January 2006.

[60] O. Le Meur, P. Le Callet, D. Barba, and D. Thoreau, "A coherent computational approach to model bottom-up visual attention," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 5, pp. 802–817, 2006.

[61] A. Ninassi, O. Le Meur, P. Le Callet, and D. Barba, "Does where you gaze on an image affect your perception of quality? Applying visual attention to image quality metric," in *Proceedings of the 14th IEEE International Conference on Image Processing (ICIP '07)*, pp. II169–II172, September 2007.

[62] C. T. Vu, E. C. Larson, and D. M. Chandler, "Visual fixation patterns when judging image quality: Effects of distortion type, amount, and subject experience," in *Proceedings of the IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI '08)*, pp. 73–76, March 2008.

[63] Y. J. Kim, M. R. Luo, W. Choe et al., "Factors affecting the psychophysical image quality evaluation of mobile phone displays: The case of transmissive liquid-crystal displays," *Journal of the Optical Society of America A*, vol. 25, no. 9, pp. 2215–2222, 2008.

[64] M. D. Gaubatz, D. M. Chandler, and S. S. Hemami, "A patch-based structural masking model with an application to compression," *Eurasip Journal on Image and Video Processing*, vol. 2009, Article ID 649316, 2009.

[65] U. Engelke, H. J. Zepernick, and A. Maeder, "Visual attention modeling: region-of-interest versus fixation patterns," in *Proceedings of 27th Conference on the Picture Coding Symposium (PCS '09)*, pp. 521–524, IEEE Press, Piscataway, NJ, USA, May 2009.

[66] K. Vilankar, L. Vasu, and D. M. Chandler, "On the visual perception of phase distortion," in *Human Vision and Electronic Imaging*, B. E. Rogowitz and T. N. Pappas, Eds., Proceedings of SPIE, San Francisco, Calif, USA, 2011.

[67] D. M. Rouse, S. S. Hemami, R. Pépion, and P. Le Callet, "Estimating the usefulness of distorted natural images using an image contour degradation measure," *Journal of the Optical Society of America A*, vol. 28, no. 2, pp. 157–188, 2011.

[68] F. Ciaramello and A. Reibman, "Systematic stress testing of image quality estimators," in *Proceedings of the Image Processing of 18th IEEE International Conference (ICIP '11)*, pp. 3101–3104, September 2011.

[69] R. L. DeValois and K. K. DeValois, *Spatial Vision*, Oxford University Press, 1990.

[70] D. Regan, *Human Perception of Objects: Early Visual Processing of Spatial Form Defined by Luminance, Color, Texture, Motion, and Binocular Disparity*, Sinauer Associates, 2000.

[71] O. H. Schade, "Optical and photoelectric analog of the eye," *Journal of the Optical Society of America*, vol. 46, no. 9, pp. 721–739, 1956.

[72] S. J. Daly, "Application of a noise-adaptive contrast sensitivity function to image data compression," *Optical Engineering*, vol. 29, pp. 977–987, 1990.

[73] E. Peli, L. E. Arend, G. M. Young, and R. B. Goldstein, "Contrast sensitivity to patch stimuli: effects of spatial bandwidth and temporal presentation," *Spatial Vision*, vol. 7, no. 1, pp. 1–14, 1993.

[74] S. Appelle, "Perception and discrimination as a function of stimulus orientation: the "oblique effect in man and animals," *Psychological Bulletin*, vol. 78, no. 4, pp. 266–278, 1972.

[75] D. O. Bowker, "Spatial frequency discrimination thresholds in different orientations," *Journal of the Optical Society of America*, vol. 70, no. 4, pp. 462–463, 1980.

[76] N. Brady and D. J. Field, "What's constant in contrast constancy? The effects of scaling on the perceived contrast of bandpass patterns," *Vision Research*, vol. 35, no. 6, pp. 739–756, 1995.

[77] D. J. Graham, D. M. Chandler, and D. J. Field, "Can the theory of "whitening" explain the center-surround properties of retinal ganglion cell receptive fields?" *Vision Research*, vol. 46, no. 18, pp. 2901–2913, 2006.

[78] P. G. J. Barten, "Formula for the contrast sensitivity of the human eye," in *Imaging Quality and System Performance*, Proceedings of SPIE, pp. 231–238, January 2004.

[79] G. E. Legge and J. M. Foley, "Contrast masking in human vision," *Journal of the Optical Society of America*, vol. 70, pp. 1458–1470, 1980.

[80] H. R. Blackwell, "Contrast thresholds of the human eye," *Journal of the Optical Society of America*, vol. 36, no. 11, pp. 624–632, 1946.

[81] B. Breitmeyer, "Visual masking: Past accomplishments, present status, future developments," *Advances in Cognitive Psychology*, vol. 3, no. 1-2, pp. 9–20, 2007.

[82] N. Graham, *Visual Pattern Analyzers*, Oxford University Press, New York, NY, USA, 1989.

[83] D. G. Pelli, *Effects of visual noise [Ph.D. thesis]*, Cambridge University Press, Cambridge, UK, 1981.

[84] P. W. Jones, S. J. Daly, R. S. Gaborski, and M. Rabbani, "Comparative study of wavelet and discrete cosine transform (DCT) decompositions with equivalent quantization and encoding strategies for medical images," in *Medical Imaging: Image Display*, Y. Kim, Ed., Proceedings of SPIE, pp. 571–582, February 1995.

[85] W. Zeng, S. Daly, and S. Lei, "An overview of the visual optimization tools in JPEG 2000," *Signal Processing: Image Communication*, vol. 17, no. 1, pp. 85–104, 2002.

[86] "Point-wise extended visual masking for JPEG-2000 image compression," in *International Conference on Image Processing (ICIP '00)*, pp. 657–660, September 2000.

[87] S. Daly, "Visible differences predictor: an algorithm for the assessment of image fidelity," in *Digital Images and Human Vision*, A. B. Watson, Ed., pp. 179–206, 1993.

[88] P. Teo and D. Heeger, "Perceptual image distortion," in *Proceedings of the IEEE International Conference Image Processing (ICIP '94)*, vol. 2, pp. 982–986, November 1994.

[89] Y. Zhang, B. Pham, and M. P. Eckstein, "Investigation of JPEG 2000 encoder options on model observer performance in signal known exactly but variable tasks (SKEV)," in *Medical Imaging: Image Perception, Observer Performance, and Technology Assessment*, A. K. D. P. Chakraborty, Ed., vol. 5034 of *Proceedings of SPIE*, pp. 371–382, February 2003.

[90] R. A. Smith and D. J. Swift, "Spatial-frequency masking and Birdsall's theorem," *Journal of the Optical Society of America A*, vol. 2, no. 9, pp. 1593–1599, 1985.

[91] F. W. Campbell and J. G. Robson, "Application of Fourier analysis to the visibility of gratings," *Journal of Physiology*, vol. 197, no. 3, pp. 551–566, 1968.

[92] C. R. Carlson, R. W. Cohen, and I. Gorog, "Visual processing of simple two dimensional sine wave luminance gratings," *Vision Research*, vol. 17, no. 3, pp. 351–358, 1977.

[93] M. B. Sachs, J. Nachmias, and J. G. Robson, "Spatial- frequency channels in human vision," *Journal of the Optical Society of America*, vol. 61, no. 9, pp. 1176–1186, 1971.

[94] N. Graham, "Visual detection of aperiodic spatial stimuli by probalility summation among narrowband channels," *Vision Research*, vol. 17, no. 5, pp. 637–652, 1977.

[95] A. B. Watson, "Summation of grating patches indicates many types of detector at one retinal location," *Vision Research*, vol. 22, no. 1, pp. 17–25, 1982.

[96] V. Manahilov and W. A. Simpson, "Energy model for contrast detection: spatial-frequency and orientation selectivity in grating summation," *Vision Research*, vol. 41, no. 12, pp. 1547–1560, 2001.

[97] G. Meinhardt, "Evidence for different nonlinear summation schemes for lines and gratings at threshold," *Biological Cybernetics*, vol. 81, no. 3, pp. 263–277, 1999.

[98] J. Lubin, "A visual discrimination model for imaging system design and evaluation," in *Vision Models for Target Detection and Recognition*, E. Peli, Ed., pp. 245–283, World Scientific, 1995.

[99] A. B. Watson and J. A. Solomon, "A model of visual contrast gain control and pattern masking," *Journal of the Optical Society of America A*, vol. 14, pp. 2378–2390, 1997.

[100] C. J. B. Lambrecht, "A working spatio-temporal model of the human visual system for image restoration and quality assessment applications," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '96)*, pp. 2291–2294, May 1996.

[101] S. Winkler, "Visual quality assessment using a contrast gain control model," in *Proceedings of the IEEE Signal Processing Society Workshop on Multimedia Signal Processing*, pp. 527–532, September 1999.

[102] D. J. Heeger, "Normalization of cell responses in cat striate cortex," *Visual neuroscience*, vol. 9, no. 2, pp. 181–197, 1992.

[103] J. M. Foley, "Human luminance pattern-vision mechanisms: masking experiments required a new model," *Journal of the Optical Society of America A*, vol. 11, no. 6, pp. 1710–1719, 1994.

[104] J. M. Foley and C. C. Chen, "Pattern detection in the presence of maskers that differ in spatial phase and temporal offset: threshold measurements and a model," *Vision Research*, vol. 39, no. 23, pp. 3855–3872, 1999.

[105] JNDMetrix Technology, "Sarnoff Corporation," http://www.sarnoff.com/.

[106] A. P. Bradley, "A wavelet visible difference predictor," *IEEE Transactions on Image Processing*, vol. 8, no. 5, pp. 717–730, 1999.

[107] N. Damera-Venkata, T. D. Kite, W. S. Geisler, B. L. Evans, and A. C. Bovik, "Image quality assessment based on a degradation model," *IEEE Transactions on Image Processing*, vol. 9, no. 4, pp. 636–650, 2000.

[108] P. Le Callet and D. Barba, "A robust quality metric for color image quality assessment," in *Proceedings of the International Conference on Image Processing (ICIP '03)*, pp. 437–440, September 2003.

[109] W. S. Geisler, J. S. Perry, B. J. Super, and D. P. Gallogly, "Edge co-occurrence in natural images predicts contour grouping performance," *Vision Research*, vol. 41, no. 6, pp. 711–724, 2001.

[110] J. B. Martens, "Multidimensional modeling of image quality," *Proceedings of the IEEE*, vol. 90, no. 1, pp. 133–153, 2002.

[111] L. de Processamento de Sinais, "RBID realistic blurred image database," http://www.lps.ufrj.br/profs/eduardo/ImageDatabase.htm.

[112] U. Engelke, A. J. Maeder, and H. -J. Zepernick, "Visual attention for image quality database," http://www.bth.se/tek/rcg.nsf/pages/vaiq-db.

[113] H. Liu, N. Klomp, and I. Heynderickx, "Tud image quality database: perceived ringing," http://mmi.tudelft.nl/iqlab/ringing.html.

[114] E. Bosc, R. Pepion, P. Le Callet et al., "Towards a new quality metric for 3-d synthesized view assessment," in *Proceedings of the IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 7, pp. 1332–1343, November 2011.

[115] F. De Simone, L. Goldmann, V. Baroncini, and T. Ebrahimi, "Subjective evaluation of JPEG XR image compression," in *Applications of Digital Image Processing*, vol. 7443 of *Proceedings of SPIE*, August 2009.

[116] H. Liu, J. Wang, J. Redi, P. Le Callet, and I. Heynderickx, "An efficient no-reference metric for perceived blur," in *Proceedings of the Visual Information Processing (EUVIP '11)*, 3rd European Workshop, pp. 174–179, July 2011.

[117] A. Zaric, N. Tatalovic, N. Brajkovic et al., "Vcl@fer image quality assessment database," in *Proceedings of the (ELMAR '11)*, pp. 105–110, September 2011.

[118] K. Fliegel, "QUALINET multimedia databases v3.0," 2012, http://dbq.multimediatech.cz/media/qo0206.doc.

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

[120] S. Winkler, "Analysis of public image and video databases for quality assessment," *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 6, pp. 616–625, 2012.

[121] A. Ninassi, P. L. Callet, and F. Autrusseau, "Pseudo No Reference image quality metric using perceptual data hiding," in *Human Vision and Electronic Imaging*, vol. 6057 of *Proceedings of SPIE*, pp. 146–157, January 2006.

[122] P. L. Callet and F. Autrusseau, "Subjective quality assessment irccyn/ivc database," http://www.irccyn.ec-nantes.fr/ivcdb/.

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

[124] H. Sheikh, Z. Wang, L. Cormack, and A. Bovik, "LIVE image quality assessment database Release 2," http://live.ece.utexas.edu/research/quality.

[125] D. M. Chandler and S. S. Hemami, "VSNR: a wavelet-based visual signal-to-noise ratio for natural images," *IEEE Transactions on Image Processing*, vol. 16, no. 9, pp. 2284–2298, 2007.

[126] D. M. Chandler and S. S. Hemami, "Dynamic contrast-based quantization for lossy wavelet image compression," *IEEE Transactions on Image Processing*, vol. 14, no. 4, pp. 397–410, 2005.

[127] N. Ponomarenko, F. Battisti, K. Egiazarian, J. Astola, and V. Lukin, "Metrics performance comparison for color image database," in *Proceedings of the 4th international workshop on video processing and quality metrics for consumer electronics*, January 2009.

[128] N. Ponomarenko, V. Lukin, A. Zelensky et al., "TID2008—a database for evaluation of full-reference visual quality assessment metrics," *Advances of Modern Radioelectronics*, vol. 10, pp. 30–45, 2009.

[129] Z. M. P. Sazzad, Y. kawayoke, and Y. Horita, "Image quality evaluation database," http://mict.eng.u-toyama.ac.jp/mictdb.html.

[130] A. Ninassi, O. L. Meur, P. L. Callet, and D. Barba, "Which semi-local visual masking model for wavelet based image quality metric?" in *Proceedings of the 15th IEEE International Conference on Image Processing (ICIP '08)*, pp. 1180–1183, October 2008.

[131] S. Tourancheau, F. Autrusseau, Z. M. P. Sazzad, and Y. Horita, "Impact of subjective dataset on the performance of image quality metrics," in *Proceedings of the 15th IEEE International Conference on Image Processing (ICIP '08)*, pp. 365–368, October 2008.

[132] E. Marini, F. Autrusseau, and P. L. Callet, *Evaluation of Standard Watermarking Techniques*, 2007, http://www.irccyn.ec-nantes.fr/spip.php?article487&lang=en.

[133] F. Autrusseau and P. Bas, *Subjective Quality Assessment of the Broken Arrows Watermarking Technique*, 2009, http://www.irccyn.ec-nantes.fr/~autrusse/Databases/Broken-Arrows/.

[134] M. Carosi, V. Pankajakshan, and F. Autrusseau, "Toward a simplified perceptual quality metric for watermarking applications," in *Society of Photo-Optical Instrumentation Engineers*, Proceedings of SPIE, January 2010.

[135] F. Autrusseau and P. Meerwald, *DT-CWT Versus DWT Watermark Embedding*, 2009, http://www.irccyn.ec-nantes.fr/spip.php?article487&lang=en.

[136] U. Engelke, M. Kusuma, H. J. Zepernick, and M. Caldera, "Reduced-reference metric design for objective perceptual quality assessment in wireless imaging," *Signal Processing: Image Communication*, vol. 24, no. 7, pp. 525–547, 2009.

[137] U. Engelke, H. -J. Zepernick, and M. Kusuma, *Wireless Imaging Quality Database*, http://www.bth.se/tek/rcg.nsf/pages/wiqdb.

[138] U. Engelke, A. Maeder, and H. J. Zepernick, "Visual attention modelling for subjective image quality databases," in *Proceedings of the IEEE International Workshop on Multimedia Signal Processing (MMSP '09)*, pp. 1–6, October 2009.

[139] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *Journal of Electronic Imaging*, vol. 19, no. 1, Article ID 011006, 2010.

[140] C. Vu, T. Phan, P. Singh, and D. M. Chandler, *Digitally Retouched Image Quality (DRIQ) Database*, 2012, http://vision.okstate.edu/driq/.

[141] VQEG, *Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment, Phase II*, 2003, http://www.vqeg.org.

[142] J. D. Long, *Rank Order Correlation*, John Wiley & Sons, 2010.

[143] B. Moulden, F. Kingdom, and L. F. Gatley, "The standard deviation of luminance as a metric for contrast in random-dot images," *Perception*, vol. 19, no. 1, pp. 79–101, 1990.

[144] F. X. Lukas and Z. L. Budrikis, "Picture quality prediction based on a visual model," *IEEE Transactions on Communications*, vol. 30, no. 7, pp. 1679–1692, 1982.

[145] N. B. Nill, "Visual model weighted cosine transform for image compression and quality assessment," *IEEE Transactions on Communications*, vol. 33, no. 6, pp. 551–557, 1985.

[146] C. Zetzsche and G. Hauske, "Multiple channel model for the prediction of subjective image quality," in *Human Vision, Visual Processing, and Digital Display*, B. E. Rogowitz, Ed., 1989.

[147] P. G. J. Barten, "Evaluation of subjective image quality with the squareroot integral method," *Journal of the Optical Society of America A*, vol. 7, no. 102024, 2031 pages, 1990.

[148] S. J. P. Westen, R. L. Lagendijk, and J. Biemond, "Perceptual image quality based on a multiple channel HVS model," in *Proceedings of the 20th International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, pp. 2351–2354, May 1995.

[149] C. C. Taylor, Z. Pizlo, J. P. Allebach, and C. A. Bouman, "Image quality assessment with a Gabor pyramid model of the human visual system," in *Human Vision and Electronic Imaging II*, N. B. . E. Rogowitz and T. N. Pappas, Ed., vol. 3016 of *Proceedings of SPIE*, pp. 58–69, February 1997.

[150] Y. K. Lai and C. C. J. Kuo, "Image quality measurement using the haar wavelet," in *Wavelet Applications in Signal and Image Processing V*, Proceedings of SPIE, pp. 127–138, July 1997.

[151] M. Miyahara, K. Kotani, and V. Ralph Algazi, "Objective picture quality scale (PQS) for image coding," *IEEE Transactions on Communications*, vol. 46, no. 9, pp. 1215–1226, 1998.

[152] W. Osberger, N. Bergmann, and A. Maeder, "Automatic image quality assessment technique incorporating higher level perceptual factors," in *Proceedings of the International Conference on Image Processing (ICIP '98)*, vol. 3, pp. 414–418, October 1998.

[153] S. Winkler, "Perceptual distortion metric for digital color images," in *Proceedings of the 1998 International Conference on Image Processing (ICIP '98)*, pp. 399–403, October 1998.

[154] J. Lubin et al., "Method and apparatus for assessing the visibility of differences between two image sequences," *US Patent*, vol. 5, pp. 974–159, 1999.

[155] P. Le Callet, A. Saadane, and D. Barba, "Frequency and spatial pooling of visual differences for still image quality assessment," in *Human Vision and Electronic Imaging*, Proceedings of SPIE, pp. 595–603, January 2000.

[156] T. N. Pappas, T. A. Michel, and R. O. Hinds, "Supra-threshold perceptual image coding," in *Proceedings of the 1996 IEEE International Conference on Image Processing (ICIP '96)*, pp. 237–240, September 1996.

[157] V. Laparra, J. Muñoz-Marí, and J. Malo, "Divisive normalization image quality metric revisited," *Journal of the Optical Society of America A*, vol. 27, no. 4, pp. 852–864, 2010.

[158] M. Carnec, P. Le Callet, and D. Barba, "An image quality assessment method based on perception of structural information," in *Proceedings: 2003 International Conference on Image Processing (ICIP '03)*, pp. 185–188, September 2003.

[159] G. Cheng, L. Cheng, and Y. Li, "Wavelet-based directional structural distortion model for image quality assessment," *Pattern Recognition and Image Analysis*, vol. 20, no. 3, pp. 286–292, 2010.

[160] W. Lu, X. Gao, D. Tao, and X. Li, "A waveletbased image quality assessment method," *International Journal of Wavelets*, vol. 06, pp. 541–551, 2008.

[161] Z. Haddad, A. Beghdadi, A. Serir, and A. Mokraoui, "Image quality assessment based on wave atoms transform," in *Proceedings of the 17th IEEE International Conference on Image Processing (ICIP '10)*, pp. 305–308, September 2010.

[162] G. A. Geri and Y. Y. Zeevi, "Visual assessment of variable-resolution imagery," *Journal of the Optical Society of America A*, vol. 12, no. 10, pp. 2367–2375, 1995.

[163] A. J. Maeder, "The image importance approach to human vision based image quality characterization," *Pattern Recognition Letters*, vol. 26, no. 3, pp. 347–354, 2005.

[164] J. Wang, D. M. Chandler, and P. Le Callet, "Quantifying the relationship between visual salience and visual importance," in *Human Vision and Electronic Imaging*, vol. 7527 of *Proceedings of SPIE*, January 2010.

[165] Z. Wang and A. C. Bovik, "Embedded foveation image coding," *IEEE Transactions on Image Processing*, vol. 10, no. 10, pp. 1397–1410, 2001.

[166] A. K. Moorthy and A. C. Bovik, "Visual importance pooling for image quality assessment," *IEEE Journal on Selected Topics in Signal Processing*, vol. 3, no. 2, pp. 193–201, 2009.

[167] Y. Tong, H. Konik, F. A. Cheikh, and A. Tremeau, "Full reference image quality assessment based on saliency map analysis," *Journal of Imaging Science and Technology*, vol. 54, no. 3, pp. 305031–3050314, 2010.

[168] L. Itti and C. Koch, "Computational modelling of visual attention," *Nature Reviews Neuroscience*, vol. 2, no. 3, pp. 194–203, 2001.

[169] A. Guo, D. Zhao, S. Liu, X. Fan, and W. Gao, "Visual attention based image quality assessment," in *Proceedings of the 18th IEEE International Conference on Image Processing (ICIP '11)*, pp. 3297–3300, September 2011.

[170] J. Wu, W. Lin, G. Shi, and A. Liu, "A perceptual quality metric with internal generative mechanism," in *Proceedings of the IEEE Transactions Image Processing*, no. 99, 2012.

[171] J. S. Goodman and D. E. Pearson, "Multidimensional scaling of multiply-impaired television pictures," *IEEE Trans Syst Man Cybern*, vol. 9, no. 6, pp. 353–356, 1979.

[172] A. M. Eskicioglu and P. S. Fisher, "Image quality measures and their performance," *IEEE Transactions on Communications*, vol. 43, no. 12, pp. 2959–2965, 1995.

[173] P. Fränti, "Blockwise distortion measure for statistical and structural errors in digital images," *Signal Processing: Image Communication*, vol. 13, no. 2, pp. 89–98, 1998.

[174] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 81–84, 2002.

[175] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," in *Proceedings of the Conference Record of the 37th Asilomar Conference on Signals, Systems and Computers*, pp. 1398–1402, November 2003.

[176] M. P. Sampat, Z. Wang, S. Gupta, A. C. Bovik, and M. K. Markey, "Complex wavelet structural similarity: a new image similarity index," *IEEE Transactions on Image Processing*, vol. 18, no. 11, pp. 2385–2401, 2009.

[177] C. L. Yang, W. R. Gao, and L. M. Po, "Discrete wavelet transform-based structural similarity for image quality assessment," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '08)*, pp. 377–380, October 2008.

[178] G. L. Ji, X. M. Ni, and H. Y. Bae, "A full-Reference image quality assessment algorithm based on haar wavelet transform," in *Proceedings of the International Conference on Computer Science and Software Engineering (CSSE '08)*, pp. 791–794, IEEE Computer Society, Washington, DC, USA, December 2008.

[179] G. Cao, L. Liang, S. Ma, and D. Zhao, "Image quality assessment using spatial frequency component," in *Proceedings of the 10th Pacific Rim Conference on Multimedia: Advances in Multimedia Information Processing*, pp. 201–211, 2009.

[180] Y. Shi, Y. Ding, R. Zhang, and J. Li, "Structure and hue similarity for color image quality assessment," in *Proceedings of the International Conference on Electronic Computer Technology (ICECT '09)*, pp. 329–333, February 2009.

[181] D. V. Rao and L. P. Reddy, "Contrast weighted perceptual structural similarity index for image quality assessment," in *Proceedings of the Annual IEEE India Council Conference (INDICON '09)*, December 2009.

[182] L. Zhang, L. Zhang, and X. Mou, "RFSIM: a feature based image quality assessment metric using Riesz transforms," in *Proceedings of the 17th IEEE International Conference on Image Processing (ICIP '10)*, pp. 321–324, September 2010.

[183] E. Chebbi, F. Benzarti, and H. Amiri, "Image quality assessment based on perceptual blur metric," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 1, no. 2, 2012.

[184] X. Fei, L. Xiao, Y. Sun, and Z. Wei, "Perceptual image quality assessment based on structural similarity and visual masking," *Signal Processing: Image Communication*, vol. 27, no. 7, pp. 772–783, 2012.

[185] D. O. Kim and R. H. Park, "New image quality metric using the harris response," *IEEE Signal Processing Letters*, vol. 16, no. 7, pp. 616–619, 2009.

[186] J. Zhu and N. Wang, "Image quality assessment by visual gradient similarity," in *Proceedings of the IEEE Transactions of Image Processing*, vol. 21, no. 3, pp. 919–933, March 2012.

[187] X. Chen, R. Zhang, and S. Zheng, "Image quality assessment based on local edge direction histogram," in *Proceedings of the International Conference on Image Analysis and Signal Processing (IASP '11)*, pp. 108–112, October 2011.

[188] A. Liu, W. Lin, and M. Narwaria, "Image quality assessment based on gradient similarity," in *Proceedings of the IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1500–1512, April 2012.

[189] G. Zhai, W. Zhang, X. Yang, and Y. Xu, "Image quality assessment metrics based on multi-scale edge presentation," in *Proceedings of the IEEE Workshop on Signal Processing Systems Design and Implementation (SiPS '05)*, pp. 331–336, November 2005.

[190] M. Zhang and X. Mou, "A psychovisual image quality metric based on multi-scale structure similarity," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '08)*, pp. 381–384, October 2008.

[191] L. Jin, N. Ponomarenko, and K. Egiazarian, "Novel image quality metric based on similarity," in *Proceedings of the 10th International Symposium on Signals, Circuits and Systems (ISSCS '11)*, pp. 1–4, July 2011.

[192] C. -H. Chou and Y. -H. Hsu, "Image quality assessment based on binary structure information," in *Proceedings of the 7th International Conference on Computational Intelligence and Security*, pp. 1136–1140, 2011.

[193] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: a feature similarity index for image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, 2011.

[194] M. Narwaria, W. Lin, I. McLoughlin, S. Emmanuel, and L. -T. Chia, "Fourier transform-based scalable image quality measure," in *Proceedings of the IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3364–3377, August 2012.

[195] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, 2006.

[196] A. Shnayderman, A. Gusev, and A. M. Eskicioglu, "An SVD-based grayscale image quality measure for local and global assessment," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 422–429, 2006.

[197] A. Mansouri, A. M. Aznaveh, F. Torkamani-Azar, and J. A. Jahanshahi, "Image quality assessment using the singular value decomposition theorem," *Optical Review*, vol. 16, no. 2, pp. 49–53, 2009.

[198] M. Narwaria and W. Lin, "Svd-based quality metric for image and video using machine learning," *Systems, Man, and Cybernetics B*, vol. 42, no. 2, pp. 347–364, 2012.

[199] A. Saha, G. Bhatnagar, and Q. Wu, "Svd filter based multiscale approach for image quality assessment," in *Proceedings of the IEEE International Conference on Multimedia and Expo Workshops (ICMEW '12)*, pp. 43–48, July 2012.

[200] P. Peng and Z. Li, "Image quality assessment based on distortion-aware decision fusion," in *Proceedings of the 2nd Sino-Foreign-Interchange Conference on Intelligent Science and Intelligent Data Engineering*, pp. 644–651, 2012.

[201] P. Peng and Z. -N. Li, "A mixture of experts approach to multi-strategy image quality assessment," in *Proceedings of the 9th International Conference on Image Analysis and Recognition (ICIAR '12)*, pp. 123–130, Springer, Berlin, Germany, 2012.

[202] C. Charrier, O. l'ezoray, and G. Lebrun, "Machine learning to design full-reference image quality assessment algorithm," *IEEE Transactions on Image Communication*, vol. 27, no. 3, pp. 209–219, 2012.

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

[204] H. W. Chang and M. H. Wang, "Sparse correlation coefficient for objective image quality assessment," *Signal Processing: Image Communication*, vol. 26, no. 10, pp. 577–588, 2011.

[205] G. O. Pinto and S. S. Hemami, "Image quality assessment in the low quality regime," in *Proceedings of the Human Vision and Electronic Imaging XVII*, 2012.

[206] R. N. Fonseca and M. A. Ramirez, "Using SCIELAB for image and video quality evaluation," in *Proceedings of 12th IEEE International Symposium on Consumer Electronics (ISCE '08)*, pp. 1–4, April 2008.

[207] W. Xu and G. Hauske, "Picture quality evaluation based on error segmentation," in *Visual Communications and Image Processing '94*, pp. 1454–1465, September 1994.

[208] B. Ghanem, E. Resendiz, and N. Ahuja, "Segmentation-based perceptual image quality assessment (SPIQA)," in *Proceedings*

of the 15th IEEE International Conference on Image Processing (ICIP '08), pp. 393–396, October 2008.

[209] K. Gu, W. Zhang, C. Wang, and G. Zhai, "Full-reference image quality assessment via region-based analysis," in *Proceedings of the 4th International Congress on Image and Signal Processing (CISP '11)*, vol. 3, pp. 1711–1715, October 2011.

[210] S. Bianco, G. Ciocca, F. Marini, and R. Schettini, "Image quality assessment by preprocessing and full reference model combination," in *The International Society for Optical Engineering*, vol. 7242 of *Proceedings of SPIE*, January 2009.

[211] K. Okarma, "Combined full-reference image quality metric linearly correlated with subjective assessment," in *Proceedings of the 10th International Conference on Artificial Intelligence and Soft Computing*, pp. 539–546, Springer, Berlin, Germany, 2010.

[212] A. Lahouhou, E. Viennet, and A. Beghdadi, "Selecting low-level features for image quality assessment by statistical methods," *Journal of Computing and Information Technology*, pp. 183–189, 2010.

[213] W. Xue and X. Mou, "An image quality assessment metric based on non-shift edge," in *Proceedings of the 18th IEEE International Conference on Image Processing (ICIP '11)*, pp. 3309–3312, September 2011.

[214] S. Li, F. Zhang, L. Ma, and K. N. Ngan, "Image quality assessment by separately evaluating detail losses and additive impairments," in *Proceedings of the IEEE Transactions on Multimedia*, vol. 13, no. 5, pp. 935–949, October 2011.

[215] A. Attar, R. Rad, and A. Shahbahrami, "Ebiqa: An edge based image quality assessment," in *Proceedings of the 7th Iranian Machine of Vision and Image Processing (MVIP '11)*, pp. 1–4, November 2011.

[216] N. Ponomarenko, L. Jin, V. Lukin, and K. Egiazarian, in *Proceedings of the 13th International Conference on Advanced Concepts for Intelligent Vision Systems*, pp. 459–470, Springer, Berlin, Germany, 2011.

[217] M. Solh and G. AlRegib, "Miqm: a multicamera image quality measure," in *Proceedings of the IEEE Transactions on Image Processing*, vol. 21, no. 9, pp. 3902–3914, September 2012.

[218] P. Marziliano, F. Dufaux, S. Winkler, T. Ebrahimi, and G. Sa, "A no-reference perceptual blur metric," in *International Conference on Image Processing (ICIP '02)*, pp. 57–60, September 2002.

[219] E. Ong, W. Lin, Z. Lu et al., "A no-reference quality metric for measuring image blur," in *Proceedings of the 7th International Symposium on Signal Processing and Its Applications*, vol. 1, pp. 469–472, 2003.

[220] J. Dijk, M. Van Ginkel, R. J. Van Asselt, L. J. Van Vliet, and P. W. Verbeek, "A new sharpness measure based on Gaussian lines and edges," *CAIP*, vol. 2756, pp. 149–156, 2003.

[221] Y. C. Chung, J. M. Wang, R. R. Bailey, S. W. Chen, and S. L. Chang, "A non-parametric blur measure based on edge analysis for image processing applications," in *Proceedings of the IEEE Conference on Cybernetics and Intelligent Systems*, pp. 356–360, December 2004.

[222] S. Wu, W. Lin, L. Jian, W. Xiong, and L. Chen, "An objective out-of-focus blur measurement," in *Proceedings of the 5th International Conference on Information, Communications and Signal Processing*, pp. 334–338, December 2005.

[223] S. H. Zhong, Y. Liu, Y. Liu, and F. L. Chung, "A semantic no-reference image sharpness metric based on top-down and bottom-up saliency map modeling," in *Proceedings of the 17th IEEE International Conference on Image Processing (ICIP '10)*, pp. 1553–1556, September 2010.

[224] R. Ferzli and L. J. Karam, "A no-reference objective image sharpness metric based on the notion of Just Noticeable Blur (JNB)," *IEEE Transactions on Image Processing*, vol. 18, no. 4, pp. 717–728, 2009.

[225] N. D. Narvekar and L. J. Karam, "A no-reference perceptual image sharpness metric based on a cumulative probability of blur detection," in *Proceedings of the International Workshop on Quality of Multimedia Experience (QoMEx '09)*, pp. 87–91, July 2009.

[226] C. Y. Wee and R. Paramesran, "Image sharpness measure using eigenvalues," in *Proceedings of the 9th International Conference on Signal Processing (ICSP '08)*, pp. 840–843, October 2008.

[227] X. Zhu and P. Milanfar, "A no-reference sharpness metric sensitive to blur and noise," in *Proceedings of the International Workshop on Quality of Multimedia Experience (QoMEx '09)*, pp. 64–69, July 2009.

[228] F. C. Roffet, T. Dolmiere, P. Ladret, and M. Nicolas, "The blur effect: perception and estimation with a new no-reference perceptual blur metric," in *Electronic Imaging Symposium Conference of Human Vision and Electronic Imaging*, vol. 12 of *Proceedings of SPIE*, pp. EI6492–EI6416, Departement Images et Signal Contrat CIFRE, San Jose, Calif, USA, January 2007.

[229] E. Tsomko and H. J. Kim, "Efficient method of detecting globally blurry or sharp images," in *Proceedings of the 9th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS '08)*, pp. 171–174, May 2008.

[230] L. Debing, C. Zhibo, M. Huadong, X. Feng, and G. Xiaodong, "No reference block based blur detection," in *Proceedings of the International Workshop on Quality of Multimedia Experience (QoMEx '09)*, pp. 75–80, July 2009.

[231] X. Marichal, W. Y. Ma, and H. Zhang, "Blur determination in the compressed domain using DCT information," in *Proceedings of the International Conference on Image Processing (ICIP '99)*, pp. 386–390, October 1999.

[232] J. Caviedes and S. Gurbuz, "No-reference sharpness metric based on local edge kurtosis," in *Proceedings of the International Conference on Image Processing (ICIP '02)*, pp. III/53–III/56, usa, September 2002.

[233] N. Zhang, A. Vladar, M. Postek, and B. Larrabee, "A kurtosis-based statistical measure for two-dimensional processes and its application to image sharpness," in *Proceedings of the Section of Physical and Engineering Sciences of American Statistical Society*, pp. 4730–4736, 2003.

[234] D. Shaked and I. Tastl, "Sharpness measure: Towards automatic image enhancement," in *Proceedings of the IEEE International Conference on Image Processing 2005 (ICIP '05)*, pp. 937–940, September 2005.

[235] M. Kristan, J. Perš, M. Perše, and S. Kovačič, "A Bayes-spectral-entropy-based measure of camera focus using a discrete cosine transform," *Pattern Recognition Letters*, vol. 27, no. 13, pp. 1431–1439, 2006.

[236] R. Hassen, Z. Wang, and M. Salama, "No-reference image sharpness assessment based on local phase coherence measurement," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '10)*, pp. 2434–2437, March 2010.

[237] P. Vu and D. Chandler, "A fast wavelet-based algorithm for global and local image sharpness estimation," *IEEE Signal Processing Letters*, vol. 19, no. 7, pp. 423–426, 2012.

[238] M. J. Chen and A. C. Bovik, "No-reference image blur assessment using multiscale gradient," in *2009 International Workshop on Quality of Multimedia Experience (QoMEx '09)*, pp. 70–74, July 2009.

[239] C. T. Vu and D. M. Chandler, "S3: a spectral and spatial sharpness measure," in *Proceedings of the 1st International Conference on Advances in Multimedia (MMEDIA '09)*, pp. 37–43, July 2009.

[240] Z. Wang, A. C. Bovik, and B. L. Evans, "Blind measurement of blocking artifacts in images," in *Proceedings of the International Conference on Image Processing (ICIP '00)*, pp. 981–984, September 2000.

[241] Z. Wang, H. R. Sheikh, and A. C. Bovik, "No reference perceptual quality assessment of JPEG compressed images," in *Proceedings of the International Conference on Image Processing (ICIP '02)*, pp. 477–480, September 2002.

[242] A. C. Bovik and S. Liu, "DCT-domain blind measurement of blocking artifacts in DCT-coded images," in *Proceedings of the IEEE Interntional Conference on Acoustics, Speech, and Signal Processing*, vol. 3, pp. 1725–1728, May 2001.

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

[244] F. Pan, X. Lin, S. Rahardja, E. P. Ong, and W. S. Lin, "Measuring blocking artifacts using edge direction information," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '04)*, pp. 1491–1494, June 2004.

[245] C. Perra, F. Massidda, and D. D. Giusto, "Image blockiness evaluation based on sobel operator," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '05)*, pp. 389–392, Genova, Italy, September 2005.

[246] H. Zhang, Y. Zhou, and X. Tian, "Weighted sobel operator-based no-reference blockiness metric," in *Proceedings of the Pacific-Asia Workshop on Computational Intelligence and Industrial Application (PACIIA '08)*, pp. 1002–1006, Huhan, China, December 2008.

[247] C. S. Park, J. H. Kim, and S. J. Ko, "Fast blind measurement of blocking artifacts in both pixel and DCT domains," *Journal of Mathematical Imaging and Vision*, vol. 28, no. 3, pp. 279–284, 2007.

[248] J. Chen, Y. Zhang, L. Liang, S. Ma, R. Wang, and W. Gao, "A no-reference blocking artifacts metric using selective gradient and plainness measures," in *Proceedings of the 9th Pacific Rim Conference on Multimedia: Advances in Multimedia Information Processing*, 2008.

[249] S. Suresh, R. Venkatesh Babu, and H. J. Kim, "No-reference image quality assessment using modified extreme learning machine classifier," *Applied Soft Computing Journal*, vol. 9, no. 2, pp. 541–552, 2009.

[250] S. Suthaharan, "No-reference visually significant blocking artifact metric for natural scene images," *Signal Processing*, vol. 89, no. 8, pp. 1647–1652, 2009.

[251] C. Chen and J. A. Bloom, "A blind reference-free blockiness measure," in *Proceedings of the 11th Pacific Rim conference on Advances in Multimedia Information Processing (PCM '10)*, Part 1, pp. 112–123, 2010.

[252] E. Ong, W. Lin, Z. Lu, S. Yao, X. Yang, and L. Jiang, "No-reference JPEG-2000 image quality metric," in *Proceedings of the International Conference on Multimedia and Expo*, vol. 1, pp. 6–9, 2003.

[253] C. -S. Z. M. Li and H. -J. Zhang, "No-reference quality assessment for JPEG2000 compressed images," in *Proceedings of the International Conference on Image Processing (ICIP '04)*, pp. 24–27, October 2004.

[254] P. M. Frederic, F. Dufaux, S. Winkler, and T. Ebrahimi, "Perceptual blur and ringing metrics: application to JPEG2000," *Signal Processing: Image Communication*, vol. 19, no. 2, pp. 163–172, 2004.

[255] Z. M. P. Sazzad, Y. Kawayoke, and Y. Horita, "Spatial features based no reference image quality assessment for JPEG2000," in *Proceedings of the 14th IEEE International Conference on Image Processing (ICIP '07)*, pp. III517–III520, September 2007.

[256] "No reference image quality assessment for JPEG2000 based on spatial features," *Signal Processing: Image Communication*, vol. 23, no. 4, pp. 257–268, 2008.

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

[258] J. Zhou, B. Xiao, and Q. Li, "A no reference image quality assessment method for JPEG2000," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN '08)*, pp. 863–868, June 2008.

[259] J. Zhang, S. H. Ong, and T. M. Le, "Kurtosis-based no-reference quality assessment of JPEG2000 images," *Signal Processing: Image Communication*, vol. 26, no. 1, pp. 13–23, 2011.

[260] X. Li, "Blind image quality assessment," in *Proceedings of the IEEE International Conference of Image Processing (ICIP2002)*, vol. 1, pp. 1449–1452.

[261] van Beek, *Edge-based image representation and coding [Ph.D. thesis]*, Delft University of Technology, 1995.

[262] A. I. El-Fallah and G. E. Ford, "The evolution of mean curvature for image filtering," in *Proceedings of the IEEE International Conference Image Processing (ICIP '94)*, vol. 1, pp. 298–302, 1994.

[263] B. R. Corner, R. M. Narayanan, and S. E. Reichenbach, "Noise estimation in remote sensing imagery using data masking," *International Journal of Remote Sensing*, vol. 24, no. 4, pp. 689–702, 2003.

[264] S. Gabarda and G. Cristóbal, "Blind image quality assessment through anisotropy," *Journal of the Optical Society of America A*, vol. 24, no. 12, pp. B42–B51, 2007.

[265] "No-reference image quality assessment through von mises distribution," *Journal of the Optical Society of America A*, vol. 29, pp. 2058–2066, 2012.

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

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

[268] H. Tong, M. Li, H. Zhang, and C. Zhang, "Learning no-reference quality metric by examples," in *Proceedings of the 11th International MultiMedia Modelling Conference (MMM '05)*, January 2005.

[269] H. Tang, N. Joshi, and A. Kapoor, "Learning a blind measure of perceptual image quality," in *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, 2011.

[270] C. Li, A. C. Bovik, and X. Wu, "Blind image quality assessment using a general regression neural network," *IEEE Transactions on Neural Networks*, vol. 22, no. 5, pp. 793–799, 2011.

[271] P. Ye and D. Doermann, "No-reference image quality assessment using visual codebook," in *Proceedings of the International Conference on Image Processing*, 2011.

[272] "No-reference image quality assessment using visual codebook," *IEEE Transaction on Image Processing*, vol. 21, no. 7, pp. 3129–3138, 2012.

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

[274] "Blind image quality assessment: from natural scene statistics to perceptual quality," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3350–3364, 2011.

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

[276] M. A. Saad and A. C. Bovik, "Blind image quality assessment: a natural scene statistics approach in the DCT domain," *IEEE Transactions on Image Processing*, vol. 21, pp. 3339–3352, 2012.

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

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

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

[280] A. Maalouf, M. C. Larabi, and C. Fernandez-Maloigne, "A grouplet-based reduced reference image quality assessment," in *Proceedings of the International Workshop on Quality of Multimedia Experience (QoMEx '09)*, pp. 59–63, July 2009.

[281] I. P. Guanawan and M. Ghanbari, "Reduced reference picture quality estimation by using local harmonic amplitude information," in *Proceedings of the London Communications Symposium*, pp. 137–140, September 2003.

[282] A. Rehman and Z. Wang, "Reduced-reference SSIM estimation," in *Proceedings of the 17th IEEE International Conference on Image Processing (ICIP '10)*, pp. 289–292, September 2010.

[283] "Reduced-reference image quality assessment by structural similarity estimation," *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3378–3389, 2012.

[284] K. Chono, Y. C. Lin, D. Varodayan, Y. Miyamoto, and B. Girod, "Reduced-reference image quality assessment using distributed source coding," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '08)*, pp. 609–612, June 2008.

[285] Z. Wang and X. Shang, "Spatial pooling strategies for perceptual image quality assessment," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '06)*, pp. 2945–2948, October 2006.

[286] W. Xue and X. Mou, "Reduced reference image quality assessment based on Weibull statistics," in *Proceedings of the 2nd International Workshop on Quality of Multimedia Experience (QoMEX '10)*, pp. 1–6, June 2010.

[287] A. N. Avanaki, S. Sodagari, and A. Diyanat, "Reduced reference image quality assessment metric using optimized parameterized wavelet watermarking," in *Proceedings of the 9th International Conference on Signal Processing (ICSP '08)*, pp. 868–871, October 2008.

[288] Q. Li and Z. Wang, "Reduced-reference image quality assessment using divisive normalization-based image representation," *IEEE Journal on Selected Topics in Signal Processing*, vol. 3, no. 2, pp. 202–211, 2009.

[289] L. Ma, S. Li, F. Zhang, and K. N. Ngan, "Reduced-reference image quality assessment using reorganized DCT-based image representation," *IEEE Transactions on Multimedia*, vol. 13, no. 4, pp. 824–829, 2011.

[290] L. Ma, S. Li, F. Zhang, and K. Ngan, "Reduced-reference image quality assessment using reorganized DCT-based image representation," *IEEE Transactions on Multimedia*, vol. 13, no. 4, pp. 824–829, 2011.

[291] R. Soundararajan and A. C. Bovik, "Rred indices: reduced reference entropic differencing for image quality assessment," *IEEE Transactions on Image Processing*, vol. 21, no. 2, pp. 517–526, 2012.

[292] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *Journal of the Optical Society of America. A*, vol. 4, no. 12, pp. 2379–2394, 1987.

[293] J. H. Van Hateren and A. Van der Schaaf, "Independent component filters of natural images compared with simple cells in primary visual cortex," *Proceedings of the Royal Society B*, vol. 265, no. 1394, pp. 359–366, 1998.

[294] A. Olmos and F. A. A. Kingdom, "Mcgill calibrated colour image database," http://tabby.vision.mcgill.ca.

[295] D. J. Field, *Scale-Invariance and Self-Similar "Wavelet" Transforms: An Analysis of Natural Scenes and Mammalian Visual Systems*, Oxford University press, 1993.

[296] "What is the goal of sensory coding?" *Neural Computation*, vol. 6, pp. 559–601, 1994.

[297] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: a strategy employed by V1?" *Vision Research*, vol. 37, no. 23, pp. 3311–3325, 1997.

[298] A. J. Bell and T. J. Sejnowski, "The 'independent components' of natural scenes are edge filters," *Vision Research*, vol. 37, no. 23, pp. 3327–3338, 1997.

[299] D. C. Knill, D. Field, and D. Kersten, "Human discrimination of fractal images," *Journal of the Optical Society of America. A*, vol. 7, no. 6, pp. 1113–1123, 1990.

[300] A. V. Oppenheim and J. S. Lim, "Importance of phase in signals," *Proceedings of the IEEE*, vol. 69, no. 5, pp. 529–541, 1981.

[301] M. G. A. Thomson, D. H. Foster, and R. J. Summers, "Human sensitivity to phase perturbations in natural images: a statistical framework," *Perception*, vol. 29, no. 9, pp. 1057–1069, 2000.

[302] T. Caelli and G. Moraglia, "On the detection of signals embedded in natural scenes," *Perception and Psychophysics*, vol. 39, no. 2, pp. 87–95, 1986.

[303] M. A. Webster and E. Miyahara, "Contrast adaptation and the spatial structure of natural images," *Journal of the Optical Society of America A*, vol. 14, no. 9, pp. 2355–2366, 1997.

[304] C. A. Párraga, T. Troscianko, and D. J. Tolhurst, "The human visual system is optimised for processing the spatial information in natural visual images," *Current Biology*, vol. 10, no. 1, pp. 35–38, 2000.

[305] P. J. Bex and W. Makous, "Spatial frequency, phase, and the contrast of natural images," *Journal of the Optical Society of America A*, vol. 19, no. 6, pp. 1096–1106, 2002.

[306] B. A. Olshausen and D. J. Field, "is the other 85% of V1 doing?" in *Problems in Systems Neuroscience*, T. Sejnowski and L. van Hemmen, Eds., Oxford University Press, 2004.

[307] J. J. Fahrenfort, H. S. Scholte, and V. A. F. Lamme, "Masking disrupts reentrant processing in human visual cortex," *Journal of Cognitive Neuroscience*, vol. 19, no. 9, pp. 1488–1497, 2007.

[308] T. S. Lee, D. Mumford, R. Romero, and V. A. F. Lamme, "The role of the primary visual cortex in higher level vision," *Vision Research*, vol. 38, no. 15-16, pp. 2429–2454, 1998.

[309] R. P. N. Rao and D. H. Ballard, "Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects," *Nature Neuroscience*, vol. 2, no. 1, pp. 79–87, 1999.

[310] Y. Karklin and M. S. Lewicki, "Learning higher-order structures in natural images," *Network*, vol. 14, no. 3, pp. 483–499, 2003.

[311] P. O. Hoyer and A. Hyvärinen, "A multi-layer sparse coding network learns contour coding from natural images," *Vision Research*, vol. 42, no. 12, pp. 1593–1605, 2002.

[312] H. Neumann and W. Sepp, "Recurrent V1-V2 interaction in early visual boundary processing," *Biological Cybernetics*, vol. 81, no. 5-6, pp. 425–444, 1999.

[313] B. Willmore, R. J. Prenger, and J. L. Gallant, "Principles of neural shape coding in area V2," *Journal of Vision*, vol. 5, no. 8, p. 82, 2005.

[314] D. Navon, "Forest before trees: the precedence of global features in visual perception," *Cognitive Psychology*, vol. 9, no. 3, pp. 353–383, 1977.

[315] D. Marr and E. Hildreth, "Theory of edge detection," *Proceedings of the Royal Society of London*, vol. 207, no. 1167, pp. 187–217, 1980.

[316] P. G. Schyns and A. Oliva, "Dr. Angry and Mr. Smile: when categorization flexibly modifies the perception of faces in rapid visual presentations," *Cognition*, vol. 69, no. 3, pp. 243–265, 1999.

[317] A. Hayes, *Representation by images restricted in resolution and intensity range [Ph.D. thesis]*, University of Western Australia, Perth, Australia, 1989.

[318] J. A. Solomon, "Channel selection with non-white-noise masks," *Journal of the Optical Society of America A*, vol. 17, no. 6, pp. 986–993, 2000.

[319] D. J. Field, A. Hayes, and R. F. Hess, "Contour integration by the human visual system: Evidence for a local 'association field'," *Vision Research*, vol. 33, no. 2, pp. 173–193, 1993.

[320] M. A. Georgeson and G. D. Sullivan, "Contrast constancy: deblurring in human vision by spatial frequency channels," *Journal of Physiology*, vol. 252, no. 3, pp. 627–656, 1975.

[321] M. A. Georgeson and T. M. Shackleton, "Perceived contrast of gratings and plaids: Non-linear summation across oriented filters," *Vision Research*, vol. 34, no. 8, pp. 1061–1075, 1994.

[322] M. C. Morrone and D. C. Burr, "Capture and transparency in coarse quantized images," *Vision Research*, vol. 37, no. 18, pp. 2609–2629, 1997.

[323] J. Nachmias, "Masked detection of gratings: the standard model revisited," *Vision Research*, vol. 33, no. 10, pp. 1359–1365, 1993.

[324] J. Schulkin, *Cognitive Adaptation*, Cambridge University Press, 1st edition, 2008.

[325] L. Linde, "Similarity of distorted pictures: on the interaction between edge blur and random noise," FOA Rep. C 53004-H9, Swedish National Defense Research Institute, 1981.

[326] V. Kayargadde and J. B. Martens, "Perceptual characterization of images degraded by blur and noise: Experiments," *Journal of the Optical Society of America A*, vol. 13, no. 6, pp. 1166–1177, 1996.

[327] F. A. A. Kingdom, D. J. Field, and A. Olmos, "Does spatial invariance result from insensitivity to change?" *Journal of Vision*, vol. 7, no. 14, article 11, 2007.

[328] Y.-W. Chow, R. Pose, M. Regan, and J. Phillips, "Human visual perception of region warping distortions," in *Proceedings of the 29th Australasian Computer Science Conference (ACSC '06)*, vol. 48, pp. 217–226, 2006.

[329] J. Rovamo, P. Mäkelä, R. Näsänen, and D. Whitaker, "Detection of geometric image distortions at various eccentricities," *Investigative Ophthalmology and Visual Science*, vol. 38, no. 5, pp. 1029–1039, 1997.

[330] I. Setyawan, D. Delannay, B. Macq, and R. L. Lagendijk, "Perceptual quality evaluation of geometrically distorted images using relevant geometric transformation modeling," in *Security and Watermarking of Multimedia Contents V*, Proceedings of SPIE, pp. 85–94, January 2003.

[331] Z. Wang and E. P. Simoncelli, "Translation insensitive image similarity in complex wavelet domain," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05)*, pp. 573–576, March 2005.

[332] A. D'Angelo, L. Zhaoping, and M. Barni, "A full-reference quality metric for geometrically distorted images," *IEEE Transactions on Image Processing*, vol. 19, no. 4, pp. 867–881, 2010.

[333] M. S. Landy and N. Graham, "Visual perception of texture," in *The Visual Neurosciences*, pp. 1106–1118, MIT Press, 2004.

[334] P. Bénard, J. Thollot, and F. Sillion, "Quality assessment of fractalized NPR textures: a perceptual objective metric," in *proceedings of the Symposium on Applied Perception in Graphics and Visualization (APGV '09)*, pp. 117–120, October 2009.

[335] J. Zujovic, T. N. Pappas, and D. L. Neuhoff, "Perceptual similarity metrics for retrieval of natural textures," in *Proceedings of the IEEE International Workshop on Multimedia Signal Processing (MMSP '09)*, October 2009.

[336] J. Zujovic, T. N. Pappas, and D. L. Neuhoff, "Structural similarity metrics for texture analysis and retrieval," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '09)*, pp. 2225–2228, November 2009.

[337] D. S. Swamy, K. J. Butler, D. M. Chandler, and S. S. Hemami, "Parametric quality assessment of synthesized textures," in *Human Vision and Electronic Imaging XVI*, Proceedings of SPIE, January 2011.

[338] P. Brodatz, *Textures: A Photographic Album for Artists and Designers*, Dover Publications, 1999.

[339] E. P. Simoncelli and J. Portilla, "Texture characterization via joint statistics of wavelet coefficient magnitudes," in *Proceedings of the 1998 International Conference on Image Processing (ICIP '98)*, pp. 62–66, October 1998.

[340] M. D. Fairchild and G. M. Johnson, "iCAM framework for image appearance, differences, and quality," *Journal of Electronic Imaging*, vol. 13, no. 1, pp. 126–138, 2004.

[341] C. Vu, T. Phan, P. Banga, and D. Chandler, "On the quality assessment of enhanced images: a database, analysis, and strategies for augmenting existing methods," in *Proceedings of the IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI '12)*, pp. 181–184, April 2012.

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

[343] P. Vu, C. Vu, and D. Chandler, "A spatiotemporal most-apparentdistortion model for video quality assessment," in *Proceedings of the 18th IEEE International Conference on Image Processing (ICIP '11)*, pp. 2505–2508, September 2011.

[344] W. H. Chen, C. H. Smith, and S. C. Fralick, "Fast computational algorithm for the discrete cosine transform," *IEEE Transactions on Communications*, vol. 25, no. 9, pp. 1004–1009, 1977.

[345] H. Hou, "fast recursive algorithm for computing the discrete cosine transform," in *Proceedings of the IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 35, no. 10, pp. 1455–1461, October 1987.

[346] J. Liang and T. D. Tran, "Fast multiplierless approximations of the DCT with the lifting scheme," *IEEE Transactions on Signal Processing*, vol. 49, no. 12, pp. 3032–3044, 2001.

[347] Y. Wenjia, H. Pengwei, and X. Chao, "Matrix factorization for fast DCT algorithms," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '06)*, pp. III948–III951, May 2006.

[348] C. Cheng and K. K. Parhi, "Hardware efficient fast DCT based on novel cyclic convolution structures," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4419–4434, 2006.

[349] V. Britanak, P. Yip, and K. Rao, *Discrete Cosine and Sine Transforms: General Properties, Fast Algorithms and Integer Approximations*, Academic, 2007.

[350] D. W. Trainor, J. P. Heron, and R. F. Woods, "Implementation of the 2D DCT using a XILINX XC6264 FPGA," in *Proceedings of the 1997 IEEE Workshop on Signal Processing Systems, SiPS 97: Design and Implementation*, pp. 541–550, November 1997.

[351] G. Kiryukhin and M. Celenk, "Implementation of 2D-DCT on XC4000 series FPGA using DFT-based DSFG and DA architectures," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '01)*, pp. 302–305, October 2001.

[352] B. Fang, G. Shen, S. Li, and H. Chen, "Techniques for efficient DCT/IDCT implementation on generic GPU," in *Proceedings of the IEEE International Symposium on Circuits and Systems 2005 (ISCAS '05)*, pp. 1126–1129, May 2005.

[353] S. Tokdemir and S. Belkasim, "Parallel processing of DCT on GPU," in *Proceedings of the Data Compression Conference (DCC '11)*, p. 479, March 2011.

[354] T. T. Wong, C. S. Leung, P. A. Heng, and J. Wang, "Discrete wavelet transform on consumer-level graphics hardware," *IEEE Transactions on Multimedia*, vol. 9, no. 3, pp. 668–673, 2007.

[355] C. Tenllado, J. Setoain, M. Prieto, L. Piñuel, and F. Tirado, "Parallel implementation of the 2D discrete wavelet transform on graphics processing units: filter Bank versus lifting," *IEEE Transactions on Parallel and Distributed Systems*, vol. 19, no. 3, pp. 299–310, 2008.

[356] J. Franco, G. Bernabé, J. Fernández, and M. E. Acacio, "A parallel implementation of the 2D wavelet transform using CUDA," in *Proceedings of the 17th Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP '09)*, pp. 111–118, February 2009.

[357] M. Unser, "Fast Gabor-like windowed Fourier and continuous wavelet transforms," *IEEE Signal Processing Letters*, vol. 1, no. 5, pp. 76–79, 1994.

[358] L. Tao and H. K. Kwan, "Fast parallel approach for 2-D DHT-based real-valued discrete Gabor transform," *IEEE Transactions on Image Processing*, vol. 18, no. 12, pp. 2790–2796, 2009.

[359] X. Wang and B. E. Shi, "GPU implemention of fast gabor filters," in *Proceedings of the IEEE International Symposium on Circuits and Systems: Nano-Bio Circuit Fabrics and Systems (ISCAS '10)*, pp. 373–376, June 2010.

[360] L. Tao and H. K. Kwan, "Multirate-based fast parallel algorithms for 2-d DHT-based real-valued discrete Gabor transform," in *IEEE Transactions on Image Processing*, vol. 21, no. 7, pp. 3306–3311, July 2012.

[361] F. C. Crow, "Summed-area tables for texture mapping," *Computer Graphics*, vol. 18, no. 3, pp. 207–212, 1984.

[362] F. Shafait, D. Keysers, and T. M. Breuel, "Efficient implementation of local adaptive thresholding techniques using integral images," in *Society of Photo-Optical Instrumentation Engineers*, vol. 6815 of *Proceedings of SPIE*, January 2008.

[363] T. Phan, S. Sohoni, D. M. Chandler, and E. C. Larson, "Performanceanalysis-based acceleration of image quality assessment," in *Proceedings of the IEEE Southwest Symposium on Image Analysis and Interpretation*, April 2012.

[364] B. Gordon, S. Sohoni, and D. Chandler, "Data handling inefficiencies between CUDA, 3D rendering, and system memory," in *Proceedings of the IEEE International Symposium on Workload Characterization (0IISWC '10)*, December 2010.

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

[366] K. Okarma and P. Mazurek, *GPGPU Based Estimation of the Combined Video Quality Metric*, vol. 102 of *Advances in Intelligent and Soft Computing*, Springer, Berlin, Germany, 2011.