Scene-and-Process-Dependent Spatial Image Quality Metrics

Edward W. S. Fry▲ **and Sophie Triantaphillidou**▲
Faculty of Science and Technology, The University of Westminster, UK
E-mail: e.fry@my.westminster.ac.uk

Robin B. Jenkin▲
NVIDIA Corporation, 2788 San Tomas Expressway, Santa Clara, CA 95050, USA

Ralph E. Jacobson▲ **and John R. Jarvis**
Faculty of Science and Technology, The University of Westminster, UK

Abstract. *Spatial image quality metrics designed for camera systems generally employ the Modulation Transfer Function (MTF), the Noise Power Spectrum (NPS) and a visual contrast detection model. Prior art indicates that scene-dependent characteristics of non-linear, content-aware image processing are unaccounted for by MTFs and NPSs measured by traditional methods. The authors present two novel metrics: the log Noise Equivalent Quanta (log NEQ) and Visual log NEQ. They both employ Scene-and-Process-Dependent MTF (SPD-MTF) and NPS (SPD-NPS) measures, which account for signal transfer and noise scene dependency, respectively. The authors also investigate implementing contrast detection and discrimination models that account for scene-dependent visual masking. Also, three leading camera metrics are revised to use the above scene-dependent measures. All metrics are validated by examining correlations with the perceived quality of images produced by simulated camera pipelines. Metric accuracy improved consistently when the SPD-MTFs and SPD-NPSs were implemented. The novel metrics outperformed existing metrics of the same genre.* © 2019 Society for Imaging Science and Technology.
[DOI: 10.2352/J.ImagingSci.Technol.2019.63.6.060407]

1. INTRODUCTION

Spatial image quality metrics (IQM) are proposed in this article for scene-dependent image capture systems that apply non-linear content-aware image signal processing (ISP). Subjective image quality is defined by Engeldrum as "the integrated set of perceptions of the overall degree of excellence of an image" [1]. It is often expressed as the multivariate combination of visuo-cognitive factors—the "nesses"—concerning image quality attributes such as resolution, sharpness, noisiness, contrast and colorfulness [1]. Spatial image quality relates to the intensity and distribution of two-dimensional (2D) luminance contrast signals and is associated with sharpness, resolution, noise and contrast.

Psychophysical evaluations are the only true means of measuring the overall subjective quality of images or the perceived magnitude of their individual quality attributes. Different methods involve categorical scaling [1], paired/triplet comparison [2] or comparison with a calibrated series of ruler images using the ISO 20462 [3] Image Quality Ruler. The last records observer ratings on a generalizable ratio Subjective Quality Scale (SQS₂) [3], with increments of just-noticeable difference (JND), and a zero point that refers to low-quality scenes that are difficult to identify perceptually.

Despite recent developments in speed and precision [2, 3], psychophysical image quality evaluations are slow, expensive and difficult to carry out accurately. IQMs save this time and expense by mapping the image information—or data concerning imaging system performance and the viewing conditions—to output scores that are intended to correlate with observer quality ratings (Figure 1).

IQMs are employed for online image quality control, ISP algorithm development and imaging systems design and optimization. A broad spectrum of metrics has evolved from several research areas including imaging systems engineering, information theory, signal/image processing, computer vision, machine learning, visual psychophysics and neural physiology [4]. These IQMs are tailored for different applications. They apply varying levels of calibration, or curve-fitting, to optimize their correlation with observer quality ratings from test image datasets that contain different types of artifacts.

This article focuses specifically on no-reference spatial metrics suited for image capture systems engineering. Suitable IQMs break image quality judgement down into components relating to the different attributes and characteristics of the imaging system and human visual system (HVS). A recent review [5] by the authors defines the following spatial IQM genres: *Computational IQMs, Image Fidelity Metrics, Signal Transfer Visual IQMs (STV-IQM)* and *Multivariate Formalism (MF-IQM)*. When each genre was evaluated from a capture system engineering perspective, the Computational IQMs and Image Fidelity Metrics were concluded to be least suitable for the purpose [5].

The STV-IQMs and MF-IQMs—referred to, in this article, as *engineering metrics*—employ standard spatial system performance measures such as the Modulation
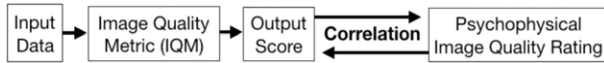
Figure 1. Generalization of image quality metric (IQM) characteristics [5].

Transfer Function (MTF), Noise Power Spectrum (NPS) and threshold Contrast Sensitivity Functions (CSF) describing visual spatial sensitivity. The Noise Equivalent Quanta (NEQ) signal-to-noise (SNR) measure is core to the most relevant STV-IQMs and is applied widely in capture system and sensor modeling [6–8]; it also uses the MTF and NPS.

Our recent evaluation of simulated camera pipelines, however, revealed that the currently employed MTF and NPS measures characterize systems that apply non-linear content-aware ISP with limited accuracy [9]. This is because signal transfer and noise in these systems is scene-dependent (i.e., dependent on the input signal). Novel *Scene-and-Process-Dependent MTF (SPD-MTF)* and *NPS (SPD-NPS)* measures have been developed to characterize such systems more suitably [9]. Likewise, contextual contrast detection [10] and discrimination [11] models, which account for each scene's contrast spectrum, should be more suitable visual models for image quality analysis than the currently used CSFs.

This article aims to revise current STV-IQMs and MF-IQMs to use these state-of-the-art input parameters and then validate them against their original incarnations. We also validate two novel STV-IQMs developed in this laboratory—the *log NEQ* and *Visual log NEQ*—that are based on a similarly revised *Scene-and-Process-Dependent NEQ (SPD-NEQ)* measure.

The following sections of the article introduce the STV-IQMs and MF-IQMs and discuss the limitations of their current input parameters. We then define the revised NPS, MTF and CSF parameters. The revised Square Root Integral with Noise (SQRIn), Perceived Information Capacity (PIC) and Camera Phone Image Quality (CPIQ) metrics and the novel log NEQ and Visual log NEQ metrics that implement these parameters are then defined. The validation methodology is then presented. This involved comparing each IQM's score with observer image quality scores for a number of test images. The latter were generated by two camera simulation pipelines that applied either linear or non-linear ISPs. Finally, we benchmark each IQM's accuracy and further analyze correlation plots describing typical metric behavior.

2. ENGINEERING METRICS

We define in this section the STV-IQMs and MF-IQMs and their input parameters, which are used in the engineering of image capture systems. These metrics express image quality as a function of the input signal, the system's performance and the observer's visual sensitivity under the viewing conditions. Their output scores are causally justified, relating to the imaging system and the HVS.

Univariate STV-IQMs, such as Acutance [12] (see later in Eq. (11)), model perceived image quality concerning the sharpness attribute. Multivariate STV-IQMs are of more relevance to this article. They account for both sharpness and noisiness, building upon signal-to-noise relationships from communications theory [13] and the founding work of Schade [14] and Nelson [15]. In this article, we revise the following STV-IQMs: Barten's [16] *Square Root Integral with Noise* and Töpfer and Jacobson's [17] *Perceived Information Capacity*. Their advantage is simplicity since they apply limited calibration and relate closely to the NEQ [18].

Keelan's MF-IQM [19] predicts the overall quality loss as the Minkowski combination [20] of several perceptually calibrated quality loss metrics for different attributes/artifacts. The recent IEEE P1858 Camera Phone Image Quality Standard [12] defines individual attribute metrics for texture loss, edge Spatial Frequency Response (SFR), local geometric distortion, visual noise, color uniformity, chroma level and lateral chromatic displacement. These metrics were combined to predict overall quality, validated in reference [21]. The CPIQ overall quality loss metric used in this article—referred to as *the CPIQ metric*—employs the texture loss and visual noise attribute metrics only, thus modeling the perceived quality concerning sharpness and noisiness. It is more complicated to implement and computationally intensive than multivariate STV-IQMs. It also applies higher levels of calibration including some curve fitting.

2.1 *Engineering Metric Input Parameters*

The accuracy of the engineering metrics is dependent on the capability of their MTF, NPS and CSF parameters to describe the system's performance in terms of sharpness and noise as well as the observer's visual sensitivity, respectively. In this section, we present sources of inaccuracy in the MTF and NPS measures currently employed. Such inaccuracies mainly relate to the MTF and NPS being used to characterize non-linear, content-aware systems. Both measures originate from linear system theory that requires systems to be linear, spatially invariant and homogeneous [18]. We also discuss the theoretical limitations of currently used CSFs.

The discrete 2D NPS (Eq. (1)) characterizes the power of the system's noise, $NPS(u, v)$, versus spatial frequency, (u, v), using the discrete Fourier transform (DFT) [22]. $I(x, y)$ is a luminance noise image of size $M \times N$, given by Eq. (2). $g(x, y)$ is the output image intensity and $\overline{g}(x, y)$ the expected intensity. If $I(x, y)$ is a scene, $NPS(u, v)$ is its DFT power spectrum, or $PS(u, v)$. The rotational average of $NPS(u, v)$ or $PS(u, v)$ yields the one-dimensional (1D) NPS, $NPS(u)$, or 1D scene power spectrum, $PS(u)$, respectively.

$$NPS(u, v) = \left| \sum_{x=\frac{M}{2}+1}^{M/2} \sum_{y=\frac{N}{2}+1}^{N/2} I(x, y)e^{-2\pi i(ux+vy)} \right|^2 \quad (1)$$

$$I(x, y) = g(x, y) - \overline{g}(x, y). \quad (2)$$

The STV-IQMs currently employ NPSs derived from captured uniform luminance patches, meaning $\overline{g}(x, y)$ can be conveniently assumed constant under certain conditions. The CPIQ visual noise attribute metric is also derived from

captured uniform patches [12]. For all capture systems, however, the amount of noise introduced to uniform patches and real scenes is not necessarily the same since photon noise is a function of intensity. Further inaccuracies result from the application of non-linear content-aware denoising and sharpening ISPs. These modify the intensity and spatial distribution of the noise dependent on the local image structure [9], rendering it both local-content-dependent and scene-dependent. Uniform patches provide the ideal input signal for content-aware denoising algorithms, in theory, and the derived noise image [5] or NPS [9] underestimates the average real-world noise level of such systems. More recently, noise measures [23] have been derived using the more suitable dead leaves target that simulates the "average scene" power spectrum and other natural scene statistics (NSS) [24]; but they have not been benchmarked.

The MTF—and the comparable SFR—characterize the system's signal transfer capabilities versus spatial frequency. The STV-IQMs currently employ MTFs derived from (i) the system's line spread function measured from a captured "perfect" edge [25], (ii) comparisons of output-to-input modulation of captured sinusoids [18] or (iii) output-to-input comparisons of (white) noise power spectra [26]. MTFs derived from such signals are not suitable for characterizing the average real-world performance of scene-dependent systems that apply non-linear content-aware ISPs. For example, their failure to account for the JPEG algorithm's non-linearity was found to reduce PIC and SQRIn metric accuracy when modeling the perceived quality of compressed images [26, 27].

Various MTF implementations employ the dead leaves target to attempt to trigger content-aware ISPs at similar levels to the "average natural scene," thus deriving a more suitable average real-world MTF. The intrinsic dead leaves implementation compares output-to-input dead leaves cross-spectra [28]. It is tolerant to noise but is a less suitable input parameter for IQMs since reversible ISPs are often unaccounted for, e.g., contrast stretching or sharpening [29]. The direct dead leaves implementation [30] (Eq. (3)) is used by the CPIQ [12] texture loss attribute metric. In Eq. (3), $PS_{Input}(u)$ and $PS_{Output}(u)$ are the input and output 1D dead leaves power spectra, respectively, and u is the spatial frequency. $NPS(u)$ is the uniform patch NPS (which is inaccurate for the reasons stated above).

$$MTF(u) = \sqrt{\frac{PS_{Output}(u) - NPS(u)}{PS_{Input}(u)}} \qquad (3)$$

Our recent camera pipeline simulations suggest that dead leaves signals may not trigger content-aware ISPs in the same fashion as the "average natural scene" [9]. Further, for scene-dependent systems, measurements from a single test target cannot accurately characterize the system's performance for any given input scene.

Finally, engineering metrics [12, 16, 17] generally employ the CSF that models thresholds of detection of unmasked narrow-band stimuli [31, 32] as a function of spatial frequency. Johnson and Fairchild's luminance CSF [31] is used by the CPIQ metric. It adapts Movshon's [33] black-box model and accounts for the stimulus spatial frequency only. The SQRIn and PIC metrics both use Barten's [32] luminance CSF that accounts for the stimulus spatial frequency, angular size and luminance. The IQMs use the CSF as a frequency domain weighting function. However, CSFs are not transfer functions and do not account for higher-level cognitive processes concerning image quality judgement [34]. Although other black-box visual models often take into account psychological functions [5], there is a lack of alternative mechanistic visual models that account for the psychophysical data and viewing conditions.

The debate is ongoing regarding whether image quality perception involves threshold or suprathreshold visual processes [11] or a combination of both [35]. What can be inferred, however, is that in image quality judgements, the detection and discrimination of visual signals, image attributes and artifacts are primarily contextual processes, which are affected by masking from other image contents as well as noise.

3. REVISED METRIC INPUT PARAMETERS
The imaging performance measures that are implemented by the revised and novel IQMs of this article are defined briefly in this section. They include various SPD-NPS and SPD-MTF measures [9] that were developed to characterize capture systems using content-aware ISPs (e.g., camera phones and autonomous vehicles). We also discuss the employed contextual CSF (cCSF) [10] and Visual Perception Function (cVPF) models [11] that account for visual masking. In theory, these parameters account for scene dependencies in spatial imaging system performance and HVS sensitivity, something that current engineering IQM parameters do not do.

3.1 *Scene-and-Process-Dependent Noise Power Spectra*
The SPD-NPS framework [5] derives the NPS of temporally varying noise in a Scene-and-Process-Dependent noise image. The latter is computed from several captured replicates of any scene (or test chart) signal. This article implements three SPD-NPS measures that apply this framework, defined below. They are summarized in Table I. They account for the effect of content-aware ISPs on temporally varying noise; Fixed Pattern Noise (FPN) is unaccounted for.

(i) The *dead leaves SPD-NPS* [9] implements the SPD-NPS framework using the dead leaves target to estimate the average real-world system performance. It is computed as the rotational average of $NPS(u, v)$ in Eq. (1), where $I(x, y)$ is given by Eq. (2) if $g(x, y)$ is a captured dead leaves signal and $\overline{g}(x, y)$ is the mean image of several captured replicates of this signal; all other parameters are as shown previously. Our simulations [9] show that it is a better noise measure for non-linear systems than the uniform patch NPS.

(ii) The *pictorial image SPD-NPS* [9] executes the SPD-NPS framework with respect to a single pictorial scene. Like the dead leaves SPD-NPS, it is computed as the rotational

average of NPS(u, v) in Eq. (1). However, $I(x, y)$ is calculated by Eq. (2) with respect to the captured scene signal, $g(x, y)$, and the mean image of several replicate scene captures, $\overline{g}(x, y)$; all other parameters are as previously indicated. Our simulations suggest that it accounts extensively for scene-dependent variations in temporally varying system noise without suffering from significant bias, provided that ten or more replicates are employed [9].

(iii) The *mean pictorial image SPD-NPS* [9] measures the average real-world performance of the system with respect to a number of pictorial scene signals. Thus, it accounts for scene dependency more comprehensively than the dead leaves SPD-NPS. It is computed as the mean, $\overline{m}_P(u)$, of a large number (N) of pictorial image SPD-NPS measurements, $P_i(u)$, across a representative scene set (Eq. (4)); i denotes each scene and u is the spatial frequency. If the scene set represents the characteristics of commonly captured scenes and $P_i(u)$ is unbiased, the measure tends toward the average real-world system performance as N increases.

$$\overline{m}_P(u) = \frac{\sum_{i=1}^{N} P_i(u)}{N} \qquad (4)$$

3.2 *Scene-and-Process-Dependent Modulation Transfer Functions*

The SPD-MTF framework [5] is based on the direct dead leaves implementation (Eq. (3)). It derives the MTF of the system with respect to pictorial scene (or dead leaves) signals, accounting for system noise using the pictorial image (or dead leaves) SPD-NPS. This article employs the following three SPD-MTF measures that apply this framework and are summarized in Table II:

(i) The *dead leaves SPD-MTF* [9] characterizes system signal transfer with respect to a dead leaves signal to estimate the system's average real-world signal transfer. It is given by Eq. (3), where $PS_{Input}(u)$ and $PS_{Output}(u)$ are input and output dead leaves power spectra, respectively. It also implements the dead leaves SPD-NPS, NPS(u). Therefore, it is a more appropriate measure for non-linear systems than the direct dead leaves MTF (which uses the uniform patch NPS but is otherwise identical to this measure).

(ii) The *pictorial image SPD-MTF* [9] characterizes system signal transfer concerning a given pictorial scene, accounting most comprehensively for system scene dependency. It is computed by Eq. (3), where $PS_{Input}(u)$ and $PS_{Output}(u)$ are input and output scene power spectra, respectively. NPS(u) is the pictorial image SPD-NPS. The input scene should be windowed (i.e., edges are tapered to a neutral pixel value) to mitigate bias due to periodic replication artifacts resulting from DFT processing [9]. This measure builds upon the method of Branca et al. [36] but mitigates significantly biases from periodic replication artifacts and noise underestimation [9]. However, it still suffers from bias due to signal-to-noise limitations that mainly affect higher frequencies of lower-power scenes at lower signal-to-noise ratios [9].

(iii) The *mean pictorial image SPD-MTF* [9] is computed by Eq. (4) as the mean, $\overline{m}_P(u)$, of the pictorial image

SPD-MTFs for the system, $P_i(u)$, across a broad set of N typical scenes; u is the spatial frequency. Thus, it characterizes the system's average real-world signal transfer, accounting for system scene dependency. However, the signal-to-noise bias in (ii) also occurs in this measure.

3.3 *Contextual Contrast Sensitivity and Visual Perception Functions*

The cCSF [10] and cVPF [11] are based on Barten's contrast detection [32] and discrimination [35] models, respectively, and employ the Linear Amplification Model [37] to account for visual masking. These scene-dependent functions were validated against observer contrast detection/discrimination datasets, measured from band-limited images of pictorial scenes. They are summarized in Table III alongside the standard Barten and Johnson and Fairchild CSFs.

4. REVISED IMAGE QUALITY METRICS

The SQRIn, PIC, and CPIQ metrics are defined in this section. Each metric is then modified to implement the revised input parameters described previously. These include imaging performance measures and visual models that account for relevant camera and human visual system scene dependencies.

4.1 *Scene-and-Process-Dependent SQRIn and PIC*

Töpfer and Jacobson's PIC [17] and their reformulation of Barten's [16] SQRIn are defined by Eqs. (5) and (6), respectively. NPS$_{visual}(u)$ is the internal noise power of the eye [38], CSF(u) is the Barten CSF [32] and u and u_{max} are the spatial frequency and cut-off frequency, respectively. k_1 and k_2 are calibration constants. The original definitions of the displayed image spectrum, $S(u)$, and total imaging system noise, $N(u)$, are for analog systems [16, 17].

$$PIC = k_1$$
$$\frac{1}{\sqrt{\int_0^\infty \ln\left(1 + \frac{S(u)CSF^2(u)}{N(u)CSF^2(u) + NPS_{visual}(u)}\right) \frac{du}{u}}} + k_2 \qquad (5)$$

$$SQRIn = \frac{k_1}{\ln 2} \int_0^{u_{max}}$$
$$\left[\frac{S(u)CSF^2(u)}{N(u)CSF^2(u) + NPS_{visual}(u)}\right]^{0.25} \frac{du}{u} + k_2 \qquad (6)$$

We define the revised SQRIn and PIC metrics by substituting their current MTF, NPS, and CSF parameters with parameters that account for scene-dependent imaging system or HVS behavior.

The revised SQRIn and PIC are defined by Eq. (5) and Eq. (6), where CSF(u) refers to the Barten CSF, cCSF, or cVPF. Revised $S(u)$ and $N(u)$ parameters are used (Eqs. (7) and (8)), where MTF$_{SPD}(u)$ and NPS$_{SPD}(u)$ denote the SPD-MTF and SPD-NPS measures, respectively. PS$_{scene}(u)$ is the input scene 1D power spectrum. γ_{disp} and MTF$_{disp}(u)$ are the display's gamma and its modeled MTF [12], respectively. NPS$_{disp}(u)$ is the display's NPS; it is assumed negligible in

this implementation.

$$S(u) = PS_{\text{scene}}(u) . MTF^2_{SPD}(u) . \gamma^2_{\text{disp}} . MTF^2_{\text{disp}}(u)^{-1} \quad (7)$$

$$N(u) = NPS_{SPD}(u)\gamma^2_{\text{disp}}MTF^2_{\text{disp}}(u) + NPS_{\text{disp}}(u) \quad (8)$$

4.2 *Scene-and-Process-Dependent CPIQ metric*

The CPIQ metric is similarly revised by substituting its input parameters for parameters that account for capture system and HVS scene dependency. Keelan's [19] MF-IQM is defined by Eq. (9), where QL_m is the overall quality loss and QL_i is the quality loss with respect to each attribute metric, i. $n_{\max}$ is the power parameter (Eq. (10)) where $QL_{\max}$ is the maximum quality loss under the viewing conditions. The constants c_1 and c_2 are set to 2 and 16.9, respectively [21].

$$QL_m = \left(\sum_i (QL_i)^{n_{\max}} \right)^{\left(\frac{1}{n_{\max}} \right)} \quad (9)$$

$$n_{\max} = 1 + c_1 . \tanh \left(\frac{QL_{\max}}{c_2} \right) \quad (10)$$

The IEEE P1858 CPIQ Standard's [12] texture loss and visual noise attribute metrics map the imaging chain Acutance, Q_T, and the total visual noise metric, Ω, to JND units, respectively, using curve-fitting functions derived from correlations with observers' data.

Equation (11) defines the imaging chain Acutance [12], Q_T, where $MTF_{system}(u)$ is the capture system's direct dead leaves MTF [30]. $MTF_{\text{disp}}(u)$ is the display's modeled MTF. $CSF(u)$ is Johnson and Fairchild's luminance CSF [31] and u and $u_{\max}$ are the spatial frequency and the cut-off frequency, respectively.

$$Q_T = \frac{\int_0^{u_{\max}} MTF_{system}(u) . MTF_{\text{disp}}(u) . CSF(u) \, du}{\int_0^{\infty} CSF(u) \, du} . \quad (11)$$

The total visual noise metric (Ω) is computed by transforming captured uniform patch(es) through the linearized sRGB, CIEXYZ and AC_1C_2 color spaces. The AC_1C_2 images are Fourier transformed and filtered with (i) Johnson and Fairchild's luminance and chrominance CSFs [31], (ii) the modeled display MTF [12] and (iii) a high-pass filter. After being inverse Fourier transformed, they are converted to the CIELAB color space via CIEXYZ. Ω is computed from the L^*a^* covariance and the variances of L^*, a^* and b^*.

We calculate the revised CPIQ metric in three stages: (i) the texture blur attribute metric is computed from a *Scene-and-Process-Dependent Acutance* measure (Eq. (11)) where $MTF_{system}(u)$ is an appropriate SPD-MTF measure, $CSF(u)$ is one of the Barten CSF, cCSF, or cVPF and all other parameters are as previously described. (ii) The visual noise attribute metric is computed from a Scene-and-Process-Dependent noise image, derived using the SPD-NPS framework. Any one of the Barten CSF, cCSF or cVPF are used as the luminance CSF during spatial filtering. (iii) Keelan's MF-IQM [19] (Eq. (9)) is used to compute the quality loss with respect to (i) and (ii).

5. NOVEL SIGNAL-TO-NOISE-BASED METRICS

In this section, we present the two novel metrics of this article, the log NEQ and Visual log NEQ; they both employ the SPD-NEQ measure, which is also novel and is defined below.

The 2D NEQ (Eq. (12)) is the standard measure of capture system signal-to-noise performance versus spatial frequency (u, v) [8]. $MTF(u, v)$ and $NPS(u, v)$ are the 2D MTF and NPS of the system, respectively. μ_A is the mean linear signal. Utilizing 1D MTFs and NPSs yields the 1D NEQ. The aforementioned limitations of current standard MTFs and NPSs are carried into the NEQ.

$$NEQ(u, v) = \frac{MTF^2(u, v)}{NPS(u, v) / \mu_A^2} \quad (12)$$

The SPD-NEQ is computed by substituting $MTF(u, v)$ and $NPS(u, v)$ with the 1D SPD-MTF and SPD-NPS, respectively. Consequently, it accounts for the scene-dependent characteristics of non-linear system signal transfer and noise. The level to which system scene dependency is accounted for depends on which SPD-MTF and SPD-NPS measures are implemented.

5.1 *Log NEQ and Visual Log NEQ Metrics*

We propose the log NEQ (Eq. (13)) and Visual log NEQ (Eq. (14)) as simple, modular and adaptable spatial IQMs. Our motivation is to produce single figure metrics that implement the SPD-NEQ to account for scene-dependent imaging system signal transfer and noise. Both metrics relate directly to fundamental signal-to-noise relationships with perceived quality and apply minimal calibration compared to other spatial IQMs. Computing the logarithm of the integrated NEQ is justified since it relates to Fechner's law [39], Shannon's Channel Capacity [13], photographic Information Capacity [40], and the PIC [17]. It was also justified empirically since it improved the metrics' correlation with the subjective image quality dataset of this article.

The log NEQ and Visual log NEQ model perceived image quality as the logarithm of the integral of a weighted SPD-NEQ measure, $NEQ_{SPD}(u)$. $MTF^2_{\text{disp}}(u)$ is the display's modeled MTF [12] and u and $u_{\max}$ are the spatial frequency and the cut-off frequency, respectively. $CSF(u)$ refers to one of the Barten CSF, cCSF, or cVPF. k_1 and k_2 are calibration constants relating to the gain (or gradient) and offset (or intercept) of the metric scores, respectively. The calibration of these constants is discussed later on in the article.

$$LogNEQ = k_1 \log_{10} \left(\int_0^{u_{\max}} MTF^2_{\text{disp}}(u) NEQ_{SPD}(u) \frac{du}{u} \right) + k_2 \quad (13)$$

$$LogNEQ_{\text{Visual}} = k_1 \log_{10} \left(\int_0^{u_{\max}} CSF^2(u) MTF^2_{\text{disp}}(u) NEQ_{SPD}(u) \frac{du}{u} \right) + k_2 \quad (14)$$
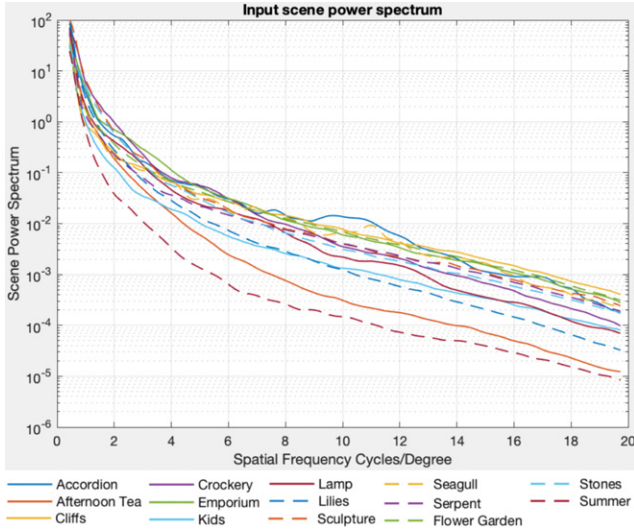
Figure 2. Power spectra for the input scenes to the image capture simulations.

6. IMAGE QUALITY METRIC VALIDATION METHODOLOGY

The methodology for generating all test images and the subjective image quality ratings dataset is detailed in this section. We also present our method for producing variants of each IQM and calibrating them to the SQS$_2$ scale.

6.1 *Test Image Dataset*

Image quality evaluations were carried out for a series of test images, generated by linear and non-linear image capture pipelines [9]. Test images were generated, starting with 14 original high-quality images, originating from a Canon DSLR camera equipped with professional lenses [41, 42]. These were downsized using bicubic interpolation and cropped to 512×512 pixels. Thumbnails of the original test scenes and their 1D power spectra are shown in Figure A1 and Figure 2, respectively. These scenes were subsequently processed by two simulated image capture pipelines that were tuned to replicate real camera phone artifacts at four exposure levels.

Both pipelines modeled the following processes identically: lens blur by convolution with a Gaussian model for a diffraction-limited lens's Airy disk; linear SNRs of 10, 20, 40, and 80 at saturation by 2D Poisson noise; read and dark noise by Gaussian noise of higher mean and standard deviation at lower SNRs; sensor quantum efficiency variations by scaling noise in the R, G and B channels by factors of 2, 1, and 3.3, respectively. Gain adjustments, noise floor removal, highlight recovery and Bayer color filter array sampling were also kept constant for both pipelines.

Further, the linear pipeline employed the following linear ISPs: demosaicing by Malvar et al. [43], denoising by 2D Gaussian filtering and sharpening by the MATLAB^{TM}*imsharpen* unsharp mask. The non-linear pipeline used the following non-linear content-aware ISPs: demosaicing by *One Step Alternating Projections* (OSAP) [44], denoising by *Block Matching and 3D Filtering*

Table I. Summary of noise measure parameters.

NPS measure	Summary of characterisation	Accounts for system noise scene-dependency	Sensitivity to system noise scene-dependency
Pictorial Image SPD-NPS	Characterises system noise with respect to a specific image	Yes	Highest
Mean Pictorial Image SPD-NPS	Characterises general system noise	Yes	↕
Dead Leaves SPD-NPS	Characterises general system noise	Partially	
Uniform Patch NPS	Characterises general system noise	No	Lowest

Table II. Summary of resolution measure parameters.

MTF measure	Summary of characterisation	Accounts for system signal transfer scene-dependency?	Accounts for system noise scene-dependency?	Sensitivity to system scene-dependency
Pictorial Image SPD-MTF	Characterises system signal transfer for an individual image	Yes	Yes	Highest
Mean Pictorial Image SPD-MTF	Characterises general system signal transfer	Yes	Yes	↕
Dead Leaves SPD-MTF	Characterises general system signal transfer	Partially	Partially	
Direct Dead Leaves MTF	Characterises general system signal transfer	Partially	No	Lowest

Table III. Summary of visual model parameters.

Contrast Sensitivity Function (CSF)	Scene-dependent Visual Behaviour Accounted For	Sensitivity to Visual Scene-Dependency
Johnson & Fairchild CSF	• N/A	Lowest
Barten CSF	• Adaptations with respect to global image luminance	↓
Contextual CSF (cCSF) / Contextual Visual Perception Function (cVPF)	• Adaptations with respect to global image luminance • Contrast masking	Highest

(BM3D) [45] and sharpening of individual color channels by the *Guided Image Filter (GIF)* [46].

The denoising and sharpening filter input parameters and the filters' opacities (Table IV) were tuned to optimize perceived output image quality after combined sharpening and denoising at each SNR under the experimental conditions. Reducing the opacity, P, below 100% lowered the intensity of filtering of the output image, $o(x, y)$, by blending the unfiltered, $g(x, y)$, and filtered images, $d(x, y)$ (Eq. (15)). This was the only way to optimize subjectively the intensity of certain ISP filters at higher SNRs. It also tested the robustness of the IQMs employing SPD-MTFs and SPD-NPSs that are designed for filtered image signals and noise, respectively [9].

$$o(x, y) = \frac{P}{100} . d(x, y) + \frac{100 - P}{100} . g(x, y). \quad (15)$$

Fifty-six test images were output from each pipeline after the demosaicing, denoising and sharpening ISP stages. They covered all permutations of the four SNRs for all 14 input scenes. Test images from the linear pipeline were selected to

Table IV. Optimal denoising and sharpening filter opacities.

Pipeline type	ISP type	Filter	Opacity (%)			
			SNR 10	SNR 20	SNR 40	SNR 80
Linear	Denoising	Gaussian	85	83	82	80
	Sharpening	USM	60	60	55	55
Non-linear	Denoising	BM3D	87	86	86	85
	Sharpening	GIF	60	70	65	60

represent both pipelines before denoising since the output images from both pipelines were nearly identical. A total of 280 test images were generated for evaluation.

6.2 *Psychophysical Image Quality Evaluations*

Psychophysical image quality ratings were recorded for each test image using the ISO 20462 Image Quality Ruler paradigm [3]. The image quality evaluations were carried out using a graphical user interface and a ruler image set, generated by Allen [41], following ISO 20462 Part 3 [3]. The subjective quality of each test image was rated by selecting the corresponding ruler image that matched its quality, from a series of 30 ruler images of the same scene. These ruler images differed in terms of sharpness only and ranged in quality from $SQS_2 = 3$ to $SQS_2 = 32$.

The experimental conditions (Figure 3) were similar to the requirements of ISO 20462 Part 3 [3] and the standard sRGB environment [47]. The EIZO ColorEdge CG245W display was employed with a pixel pitch of 0.27 mm. It was calibrated to the sRGB color space with white point luminance of 120 cd/m^2. The Nyquist frequency was 20 cycles/degree at the 60 cm viewing distance, which was restricted by a headrest. The ruler images were calibrated for these exact viewing conditions [41].

Twenty-seven observers participated, including 10 females and 17 males of various ethnicities, with approximate age range of 20–55. Six had prior experience in comparable image quality evaluations. Observers wore corrective spectacles/lenses if required for the viewing distance. Their visual acuity was confirmed for the given viewing conditions using a Snellen near vision test card [48].

6.3 *Metric Variant Computation and Calibration*

Variants of the revised and proposed IQMs were generated, and calibrated, to be benchmarked against the observer image quality ratings. A total of 332 variants were created, employing different permutations of the NPS, MTF, and CSF parameters in Tables I, II, and III, respectively. The sensitivity of these variants to imaging system and visual scene dependency varied considerably.

SPD-MTFs and SPD-NPSs were computed using ten replicates by adapting MATLAB^{TM} code for Burns's direct dead leaves MTF implementation [22]. The CPIQ visual noise metric was computed using Baxter and Murray's [49] MATLAB^{TM} implementation [50].
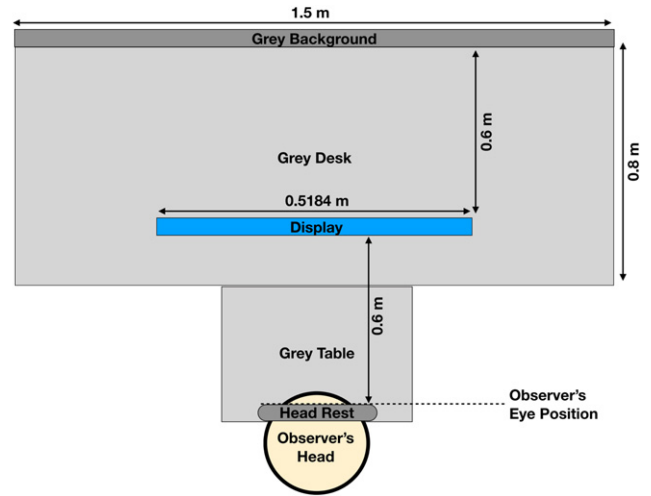

Figure 3. Layout of the laboratory equipment for the image quality evaluations.

Each variant of the log NEQ, Visual log NEQ, SQRIn and PIC was calibrated to the SQS_2 scale by (i) setting k_2 to zero following the method of [27] and (ii) by setting k_1 to an appropriate value, given by Eq. (16), by dividing the mean of the observer image quality ratings for all test images before denoising at SNR 80 (best quality), $\overline{m}_{SQS}$, by the mean of the metric's output scores for these test images when $k_1 = 1$, denoted by $\overline{m}_{output}$. When employed, the cCSF/cVPF were normalized to the same integrated area as the Barten CSF. The metric scores were thus affected by scene-dependent changes in the cCSFs/cVPFs shape but not changes in their magnitude.

$$k_1 = \frac{\overline{m}_{SQS}}{\overline{m}_{output}} \qquad (16)$$

Variants of the CPIQ metric were calibrated to SQS_2 units by subtracting the predicted quality loss (i.e., QL_m in Eq. (9)) from the SQS_2 value of 23 corresponding to the SQS_2 of the input scenes to the simulations. The cCSF, cVPF, and Barten CSF, when employed, were normalized to the same integrated area as Johnson and Fairchild's luminance CSF, as used by the IEEE P1858 CPIQ Standard [12].

7. RESULTS

The subjective image quality ratings dataset is first displayed. Variants of all IQMs are then benchmarked according to their capability to predict results from this dataset. Correlations between selected metric variants and this dataset are then presented, which demonstrate certain behavior of interest.

7.1 *Results from Quality Evaluations*

Figure 4 presents the observer ratings from the image quality evaluations. They were not calibrated according to the average scene relationship [3] since Allen [41] found this action removed virtually all scene dependency from the data. Such scene dependencies demanded consideration in this article since the purpose of the study was to test whether the scene-dependency of the capture system and
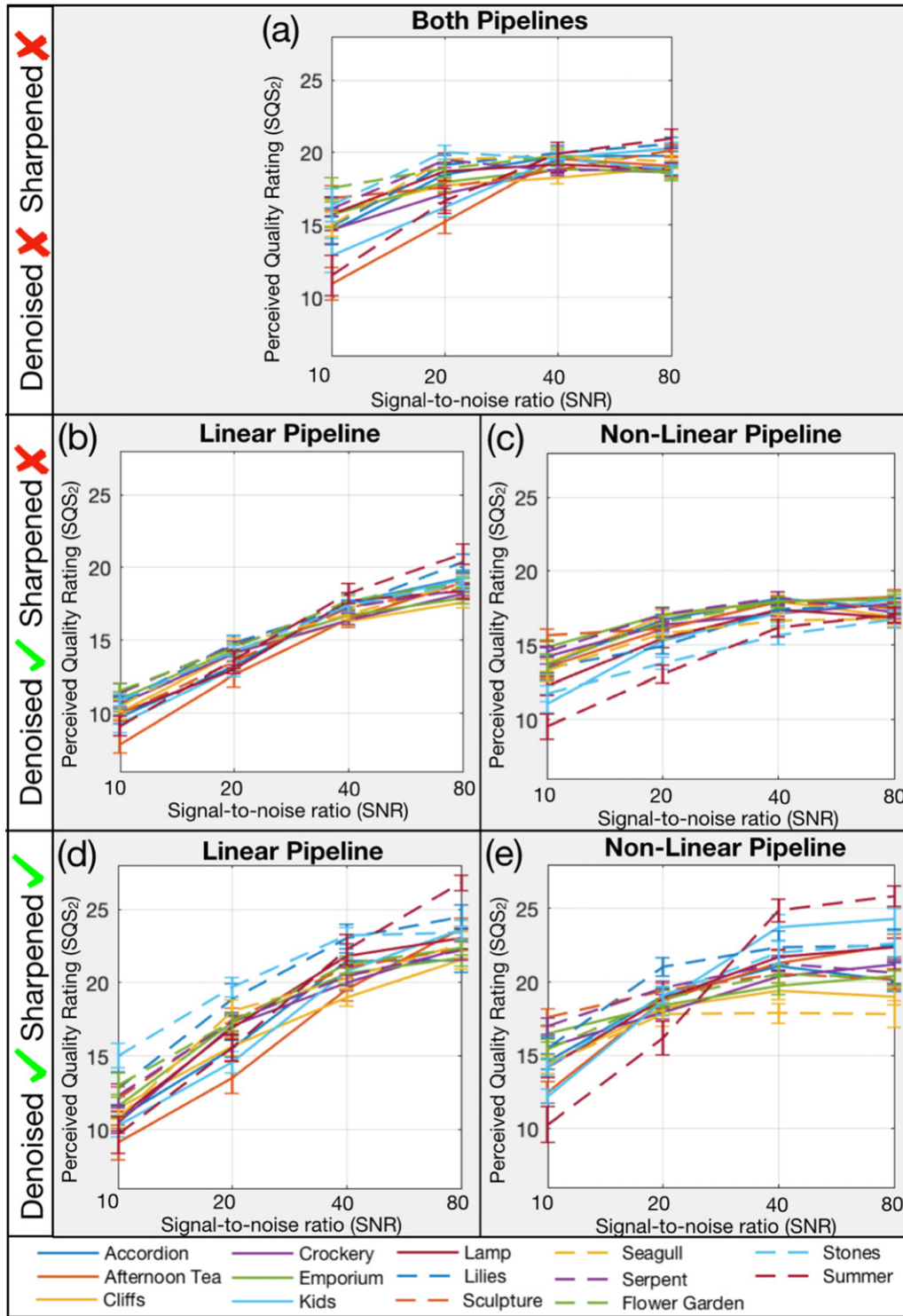
J. Imaging Sci. Technol.
IS&T International Symposium on Electronic Imaging 2020

060407-7

Nov.-Dec. 2019
Image Quality and System Performance

Figure 4. Observer image quality ratings (SQS_2) for each scene after different stages of linear and non-linear processing. Error bars show standard error.

HVS can be accounted for successfully in the selected metrics. Uncalibrated ratings were used successfully when evaluating predictions of non-linear JPEG and JPEG 2000 compressed image quality by IQMs [41].

The ratings were generally higher after combined denoising and sharpening, for which the tuning was optimized. As expected, non-linear ISPs caused the greatest scene-dependent variations. They also produced higher-quality images than the respective linear ISPs at lower SNRs because they preserved genuine image signal content or mitigated the amplification of noise. Variances of the scenes' susceptibility and observers' sensitivity to quality losses increased at lower SNRs. The former was lower than the latter, as observed previously by Keelan [19].
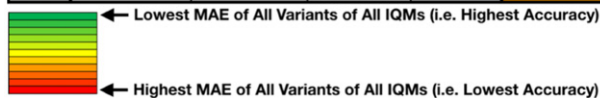
Table V. Benchmarking of highest and lowest performing variants of each metric in terms of Mean Absolute Error (MAE).

Highest Accuracy Variants of Each Metric

Metric Name	Input Parameters			Pipeline Type	MAE
	Noise Measure	MTF	CSF		
CPIQ	Uses Pictorial Image Replicates	Pictorial Image SPD-MTF	Barten	Non-Linear	1.82
				Linear	1.57
Visual Log NEQ	Mean Pictorial Image SPD-NPS	Dead Leaves SPD-MTF	Barten	Non-Linear	1.68
				Linear	2.25
Log NEQ	Pictorial Image SPD-NPS	Dead Leaves SPD-MTF	N/A	Non-Linear	2.16
				Linear	2.04
PIC	Pictorial Image SPD-NPS	Dead Leaves SPD-MTF	Any of Barten, cCSF or cVPF	Non-Linear	2.50
				Linear	2.47
SQRIn	Pictorial Image SPD-NPS	Pictorial Image SPD-MTF	Any of Barten, cCSF or cVPF	Non-Linear	5.46
				Linear	4.92

Lowest Accuracy Variants of Each Metric

Metric Name	Input Parameters			Pipeline Type	MAE
	Noise Measure	MTF	CSF		
CPIQ	Uses Dead Leaves Replicates	Direct Dead Leaves MTF	cVPF	Non-Linear	9.88
				Linear	8.82
Visual Log NEQ	Uniform Patch NPS	Pictorial Image SPD-MTF	cVPF	Non-Linear	2.48
				Linear	2.38
Log NEQ	Uniform Patch NPS	Pictorial Image SPD-MTF	N/A	Non-Linear	3.00
				Linear	2.13
PIC	Uniform Patch NPS	Direct Dead Leaves MTF	Any of Barten, cCSF or cVPF	Non-Linear	3.15
				Linear	2.53
SQRIn	Uniform Patch NPS	Direct Dead Leaves MTF	Any of Barten, cCSF or cVPF	Non-Linear	6.94
				Linear	5.34

← Lowest MAE of All Variants of All IQMs (i.e. Highest Accuracy)

← Highest MAE of All Variants of All IQMs (i.e. Lowest Accuracy)

Figure 5. The range of Mean Absolute Error (MAE) values across all variants of each metric.

7.2 Benchmarking of Image Quality Metrics

When evaluating the robustness of each metric to changes in its input parameters, the main factors we consider are the accuracy of its most accurate variant (Table V) and the range of accuracies across all variants of the metric (Figure 5).

Table V benchmarks the most and least accurate variants of each metric in terms of their Mean Absolute Error (MAE). The MAE describes specifically the mean difference in SQS_2 units between the metric scores and the ideal linear relationship with the observer quality ratings, shown by the pink line in Figures 6–8. Thus, if a variant has an MAE of 2, then it can be expected to predict the perceived quality of a given image with an accuracy of ± 2 JNDs.

The CPIQ metric produced the variants with the highest overall accuracy, followed closely by the new Visual log NEQ and log NEQ metrics of this article, and the PIC. However, the CPIQ metric was highly sensitive to changes in its input parameters (Fig. 5), in particular, the CSF. This led to it also producing the least accurate variants of all. We expect this
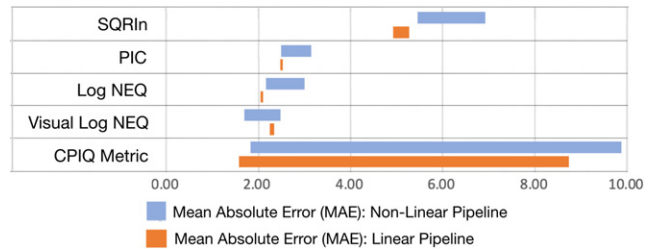
unpredictable behavior to be due to curve-fitting "forcing" the metric to its original input parameters.

The STV-IQMs—which include the PIC, SQRIn, log NEQ and Visual log NEQ—showed greater consistency when their input parameters were changed, especially for the linear pipeline (Fig. 5). The proposed log NEQ and Visual log NEQ metrics yielded the most accurate variants from this genre. Moreover, for the non-linear pipeline, the latter metric produced the most accurate variants out of all the metrics tested. This demonstrates the power of these simpler, more adaptable signal-to-noise-based metrics and the novel SPD-NEQ measure at their core.

Implementing the various SPD-NPSs and SPD-MTFs generally improved the metrics' accuracy for the non-linear pipeline. This was despite the reduced ISP filter opacities giving the current standard MTF/NPS measures an advantage. Tables benchmarking the MAE of every variant of each IQM demonstrated the consistency of these improvements. IQMs utilizing the SPD-NPSs and SPD-MTFs were often of comparable accuracy for the linear and non-linear pipelines.

The most accurate variants of each metric used SPD-NPSs or noise images derived from scenes. This demonstrates the robustness of the SPD-NPS framework and the benefits of accounting most comprehensively for system noise scene dependency. In contrast, the uniform patch NPS was used by the vast majority of the lowest accuracy variants, substantiating previous observations that the dead leaves SPD-NPS is a more generally appropriate measure [9].

When considering the appropriateness of the revised MTF, NPS and CSF parameters, evaluating their effect upon the accuracy of the log NEQ and Visual log NEQ is of particular relevance due to the simplicity and "purity" of these new metrics. The most accurate variants of these IQMs used the dead leaves SPD-MTF. This suggests that the negative effect of bias in the pictorial image SPD-MTF outweighed the benefit of accounting more comprehensively for system scene dependency. Nevertheless, the CPIQ metric and SQRIn performed most accurately with pictorial image SPD-MTFs. This demonstrates the potential benefits of the measure, which we suggest should be developed further to address measurement bias, especially when considering the success of the respective SPD-NPS measure.

The highest performing variants of each metric all used the Barten CSF. Implementing the cCSF or cVPF generally

affected negatively the metrics' accuracy. This was also the case when the cCSF/cVPF were not normalized. Changing the CSF parameter did not affect the accuracy of the PIC or SQRIn significantly. This was because the high display luminance reduced visual noise to a level where changes in the CSF parameter canceled themselves out.

Benchmarking with respect to the Root Mean Square Error (RMSE) and Spearman's Rank Order Correlation Coefficient (SROCC) displayed comparable trends to those in Table V.

7.3 *Further Analysis of Metric Correlations*
We analyze in this section correlations between output scores of the selected IQM variants and the observer quality ratings. The analysis demonstrates the typical IQM behavior and any significant changes resulting from revision of their input parameters.

We observe that when the log NEQ employed the more relevant dead leaves SPD-NPS (Fig. 6b) instead of the uniform patch NPS (Fig. 6a), it no longer overestimated perceived image quality after non-linear sharpening or denoising; interestingly, it also produced similar correlations to the linear pipeline. Utilizing SPD-NPSs derived from pictorial scenes led to further improvements (Fig. 6c). The above was also true for the PIC, SQRIn and Visual log NEQ. Respective variants of the latter metric performed similarly to those in Fig. 6 but showed improved accuracy.

The SQRIn overestimated perceived image quality at high SNRs, resulting in a curved distribution (Fig. 7). This corroborates Töpfer and Jacobson's observation [17] that under some conditions, the metric does not describe perceived image quality linearly with noise and in accurate JND units. This is expected to be due to the underestimation of the perceived intensity of noise at near-threshold levels [17]. The PIC displayed the same behavior to a lesser extent. We conclude that our simulations exceeded the SNR threshold to which the SQRIn and PIC apply, which is understandable since both metrics were developed and calibrated for analog systems with generally higher noise.

The CPIQ metric (Fig. 8) was linearly correlated with the observer ratings for all permutations of the input parameters. These correlations did not follow the "ideal" linear relationship with gain = 1 and offset = 0, however. The main cause of this gain in the original CPIQ metric's correlation (Fig. 8a) was the effects of content-aware ISPs, although its correlation still suffered from some offset for the linear pipeline. Employing the Barten CSF and the various SPD-MTF and SPD-NPS measures reduced generally both the gains and offsets of the linear regression and rendered the correlations for the linear and non-linear pipelines more similar. This was despite the limitations of changing the input parameters in a metric that employs pre-calibration with observers' ratings.

8. CONCLUSIONS
Two novel metrics—the log NEQ and Visual log NEQ—were presented in this article. Leading engineering IQMs,
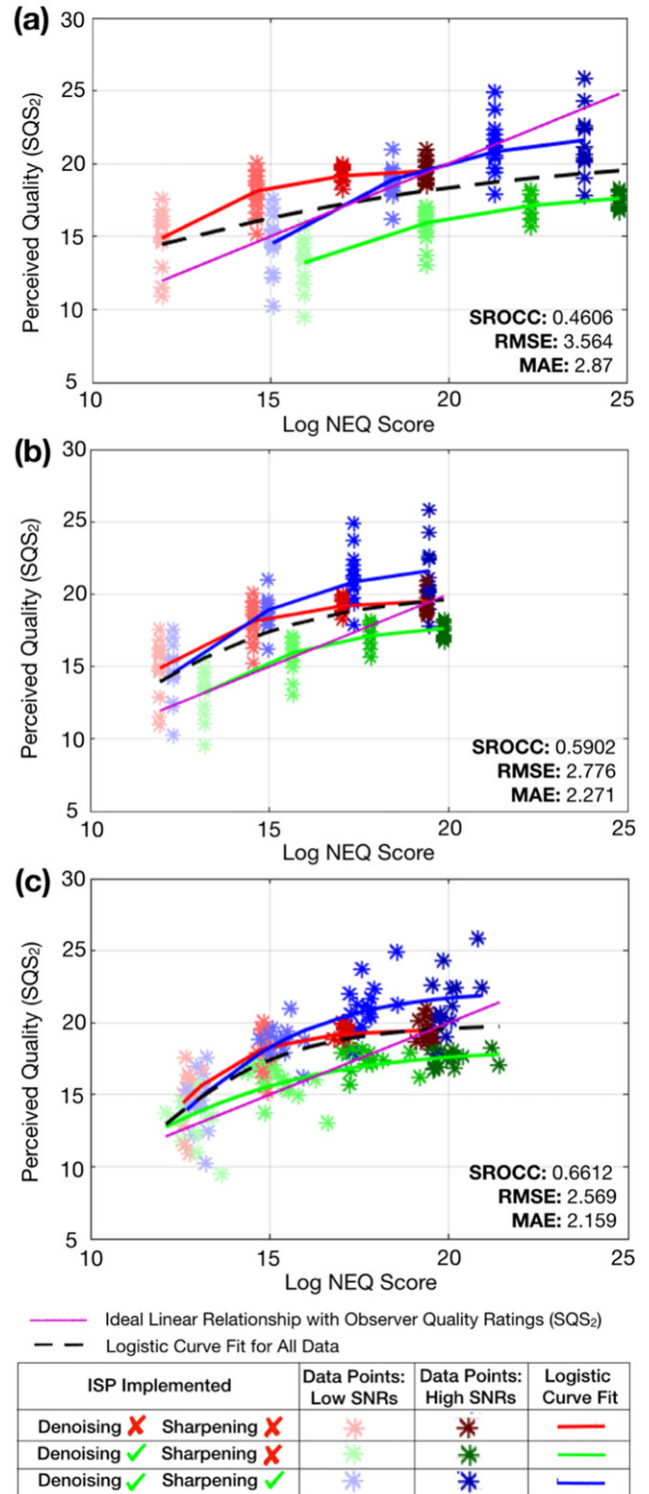

Figure 6. Observer image quality ratings versus output scores of log NEQ variants employing (a) the direct dead leaves MTF [30] and uniform patch NPS, (b) the direct dead leaves MTF [30] and dead leaves SPD-NPS and (c) the pictorial image SPD-NPS and dead leaves SPD-MTF. (c) is the most accurate log NEQ variant. All test images were generated by the non-linear pipeline.

modeling spatial image quality, were also revised, including the IEEE P1868 CPIQ metric [12], SQRIn [16] and
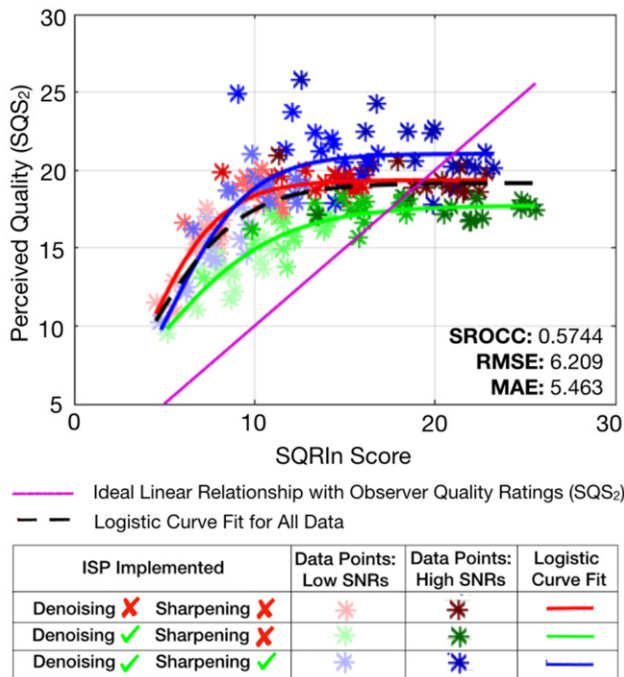
Figure 7. Observer image quality ratings versus metric scores for the highest performing SQRIn variant, which employed the pictorial image SPD-MTF and SPD-NPS. All test images were generated by the non-linear pipeline.

Figure 8. Observer image quality ratings versus output scores of CPIQ metric variants employing (a) input parameters specified in the IEEE P1858 CPIQ Standard [12] including the direct dead leaves MTF [30], uniform patch noise image and Johnson and Fairchild luminance CSF [31]; (b) the pictorial image SPD-MTF, noise images computed from pictorial image replicates and the Barten CSF [32]. (b) is the highest performing CPIQ metric variant. All test images were generated by the non-linear pipeline.

PIC [17]. Substituting the MTF, NPS, and CSF parameters with equivalent scene-dependent measures (i.e., SPD-MTF, SPD-NPS, cCSF, and cVPF) created variants of each metric. All metric variants were benchmarked using different permutations of these measures.

The log NEQ and Visual log NEQ relate directly to the novel SPD-NEQ signal-to-noise measure and apply minimal calibration, which is well prescribed and depends on the given experimental conditions only. Thus, revising their input parameters does not, in theory, offer any violation to these metrics. Variants of these IQMs were generally more accurate than those of the comparable SQRIn and PIC metrics and almost as accurate as those of the CPIQ metric. This not only demonstrates the relevance of the fundamental NEQ and SPD-NEQ measures to quality modeling, but it is also a very exciting outcome, considering the metrics' simplicity and "purity."

The CPIQ metric produced the most accurate variants of all, but the other variants ranged widely in terms of accuracy. We expect that the latter was due to the CPIQ metric's pre-calibration and suggest that the accuracy of the log NEQ and Visual log NEQ variants informs better regarding the appropriateness of the various input parameters.

Implementing the SPD-NPSs consistently improved the accuracy of all IQMs. This was particularly the case for SPD-NPSs derived from scenes. This indicates that the SPD-NPS framework is more suitable for quality modeling than the uniform patch NPS and corroborates previous conclusions that it is robust, and accounts for scene dependencies in temporally varying system noise [9].
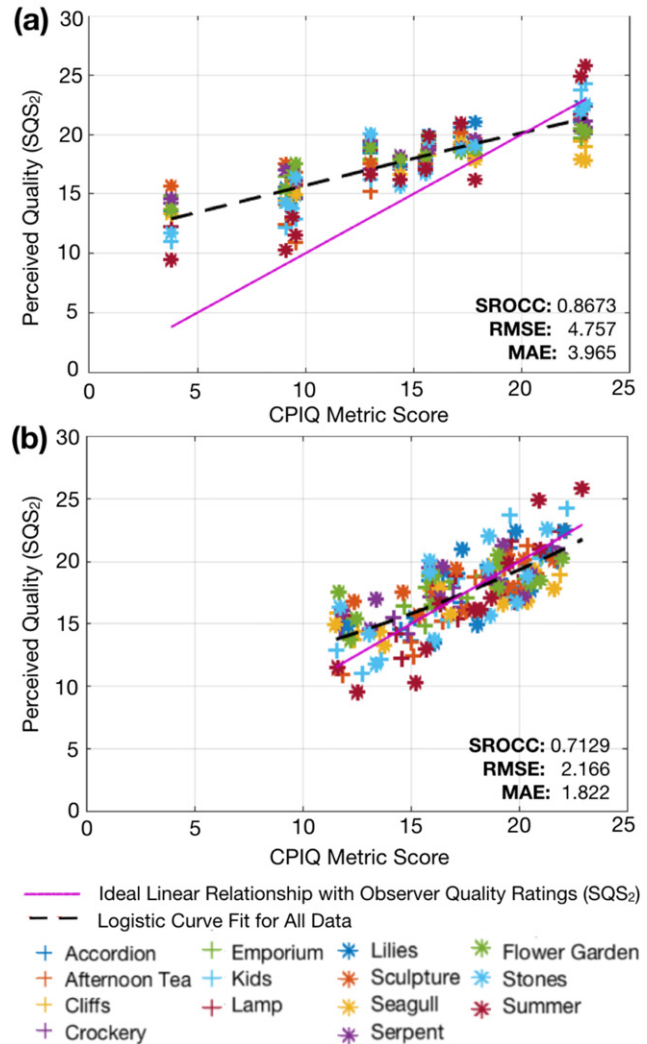
For three out of five metrics tested, the most accurate variants employed the dead leaves SPD-MTF. The other two metrics performed most accurately using the pictorial image SPD-MTF. Thus, there appear to be trade-offs between the negative effect of bias in the latter measure and the positive effect of it accounting more comprehensively for system signal transfer scene dependency.

Research efforts in deriving SPD-MTFs from extracted pictorial scene edges continue [51]. It remains to be seen whether the resultant MTFs are more suitable parameters for image quality modeling.

Utilizing contextual detection (cCSF) or discrimination models (cVPF) that account for visual masking did not improve IQM accuracy, regardless of whether or not they were normalized.

We recommend further investigations to establish whether (i) contextual visual models have a genuine role in quality modeling, (ii) validating the IQMs using real capture systems corroborates the results of this article, and (iii) the IQMs of this article describe accurately the quality of compressed images (JPEG and JPEG 2000).

The novel and revised IQMs of this article and their SPD-MTF, SPD-NPS, cCSF, and cVPF input parameters represent a new paradigm of image quality models, imaging systems characterization measures and human visual models that account for relevant scene dependencies in spatial imaging system and visual performance. The benefits of these more complex measures/metrics are demonstrated in this article and in our previous work [52].

APPENDIX A
Abbreviations
cCSF... Contextual Contrast Sensitivity Function
cVPF... Contextual Visual Perception Function
CPIQ Metric... IEEE P1858 CPIQ Standard Metric
Log NEQ... Log Noise Equivalent Quanta Metric
MF-IQM... Multivariate Formalism Image Quality Metric
PIC... Perceived Information Capacity Metric
SPD-MTF... Scene-and-Process-Dependent MTF
SPD-NEQ... Scene-and-Process-Dependent NEQ
SPD-NPS... Scene-and-Process-Dependent NPS
SQRIn... Square Root Integral with Noise Metric
STV-IQM... Signal Transfer Visual Image Quality Metric
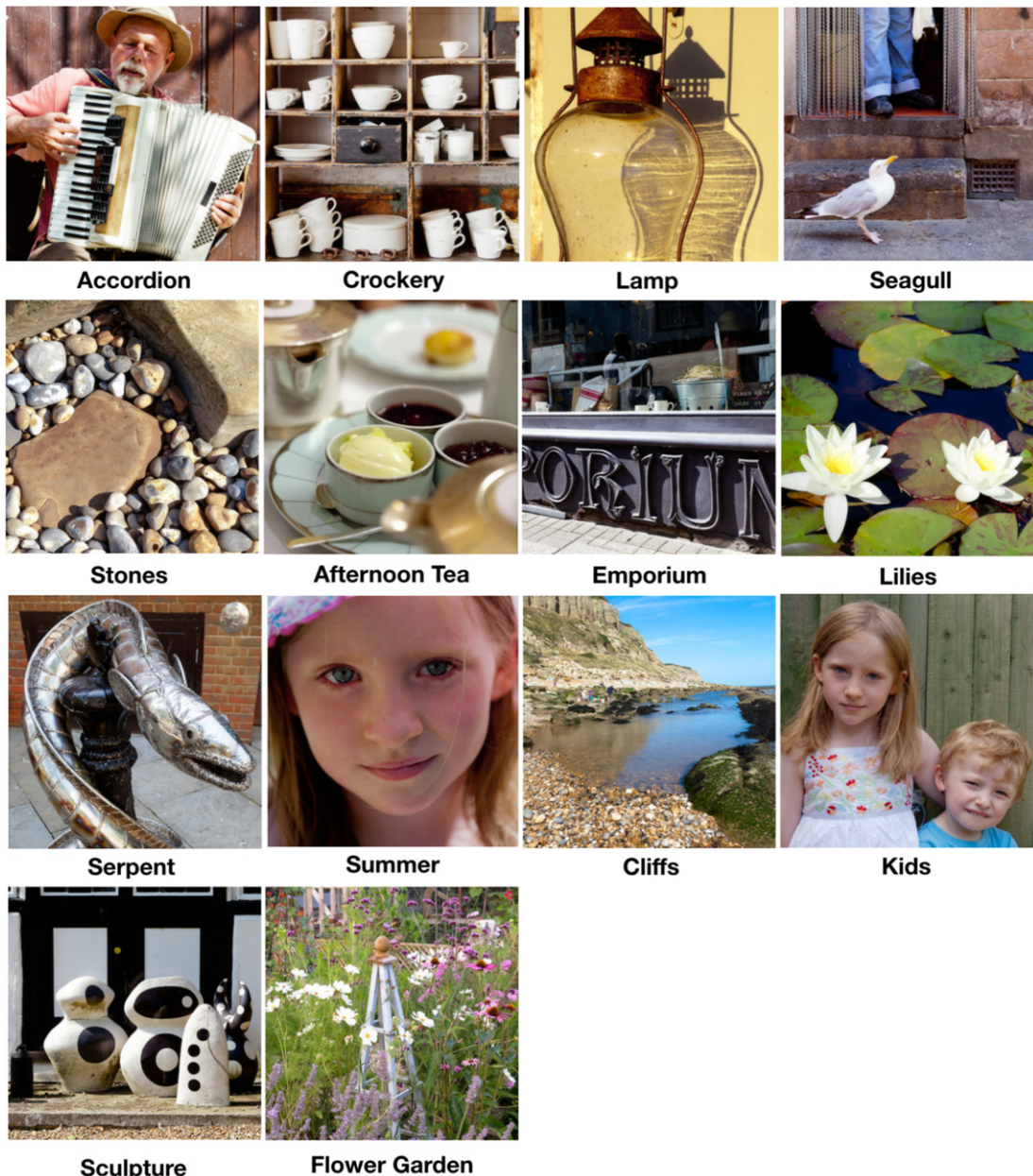Visual log NEQ... Visual log Noise Equivalent Quanta Metric

Figure A1. Input scenes to the image capture simulations. All scenes were captured and processed by Allen [41].

REFERENCES

1 P. G. Engeldrum, *Psychometric Scaling: A Toolkit for Imaging Systems Development* (Imcotek Press, Winchester MA, 2000).

2 Iso 20462-2 - 2005: Photography — Psychophysical experimental methods for estimating image quality — Part 2: Triplet Comparison Method, vol. 3 (ISO, Geneva), www.iso.org.

3 Iso 20462-3 - 2012: Photography – Psychophysical experimental methods for estimating image quality – Part 3: Quality ruler method (ISO, Geneva), www.iso.org.

4 Z. Wang and A. Bovik, "Reduced and no-reference image quality assessment," IEEE Signal Process. Mag. **28**, 29–40 (2011).

5 E. W. S. Fry, S. Triantaphillidou, R. E. Jacobson, J. R. Jarvis, and R. B. Jenkin, "Bridging the gap between imaging performance and image quality measures," *Proc. IS&T CIC26: Twenty-Sixth Color and Imaging Conf.* (IS&T, Springfield, VA, 2018).

6 P. D. Burns, "Signal-to-noise ratio analysis of charge-coupled device imagers," Proc. SPIE **1242**, 187–194 (1990).

7 B. W. Keelan, "Objective and subjective measurement and modeling of image quality: a case study," Proc. SPIE **7798**, 779815-1–779815-9 (2010).

8 B. W. Keelan, "Imaging applications of noise equivalent quanta," *Proc. IS&T CIC24: Twenty-Fourth Color and Imaging Conf.* (IS&T, Springfield, VA, 2016), pp. 1–7.

9 E. W. S. Fry, S. Triantaphillidou, R. B. Jenkin, J. R. Jarvis, and R. E. Jacobson, "Validation of modulation transfer functions and noise power spectra from natural scenes," J. Imaging Sci. Technol. **63** (2019) (in print).

10 S. Triantaphillidou, J. R. Jarvis, A. Psarrou, and G. Gupta, "Contrast detection in images of natural scenes," Signal Process., Image Commun. **75**, 64–75 (2019).

11 S. Triantaphillidou, J. R. Jarvis, and G. Gupta, "Spatial contrast sensitivity and discrimination in pictorial images," Proc. SPIE **9016**, 901604-1–901604-15 (2014).

12 IEEE Std 1858TM-2016: IEEE Standard for Camera Phone Image Quality. IEEE Standards Association Board of Governors.

13 C. E. Shannon, "A mathematical theory of communication," Bell Syst. Tech. J. **27**, 623–656 (1948).

14 O. H. Schade, "Image quality: A comparison of photographic and television systems," *RCA Lab.* (RCA Laboratories, Princeton, NJ, 1975).

15 C. N. Nelson, "Image sharpness criteria," J. Opt. Soc. Am. **63**, 1289 (1973).

16 P. G. J. Barten, "Evaluation of the effect of noise on subjective image quality," Proc. SPIE **1453**, 2–15 (1991).

17 K. Topfer and R. E. Jacobson, "The relationship between objective and subjective image quality criteria," J. Inf. Rec. Mater. **21**, 5–27 (1993).

18 J. C. Dainty and R. Shaw, *Image Science: Principles, Analysis and Evaluation of Photographic-type Imaging Processes* (Academic Press Ltd., London, 1974).

19 B. W. Keelan, *Handbook of Image Quality, Characterization and Prediction* (Marcel Dekker, New York, 2002).

20 R. N. Shepherd, "Metric structures in ordinal data," J. Math. Psychol. **3**, 287–315 (1966).

21 E. W. Jin, J. B. Phillips, S. Farnand, M. Belska, V. Tran, E. Chang, Y. Wang, and B. Tseng, "Towards the development of the IEEE P1858 CPIQ standard – a validation study," *Proc. IS&T CIC25: Twenty-Fifth Color and Imaging Conf.* (IS&T, Springfield, VA, 2017), pp. 88–94.

22 P. D. Burns, "Refined measurement of digital image texture loss," Proc. SPIE **8653**, 86530H (2013).

23 U. Artmann, "Measurement of noise using the dead leaves pattern," *Proc. IS&T CIC26: Twenty-Sixth Color and Imaging Conf.* (IS&T, Springfield, VA, 2018), pp. 341-1–341-6.

24 F. Cao, F. Guichard, and H. Hornung, "Measuring texture sharpness of a digital camera," Proc. SPIE **7250**, 72500H (2009).

25 ISO 12233 - 2017, Photography - Electronic still picture imaging-Resolution and spatial frequency responses, (ISO, Geneva), www.iso.org.

26 R. B. Jenkin, R. E. Jacobson, and M. A. Richardson, "Use of the first order wiener kernel transform in the evaluation of SQRI n and PIC quality metrics for JPEG compression," Proc. SPIE **5294**, 60–70 (2004).

27 A. M. Ford, "*Relationships Between Image Quality and Image Compression*," Ph.D. thesis (University of Westminster, UK, 1997).

28 L. Kirk, P. Herzer, U. Artmann, and D. Kunz, "Description of texture loss using the dead leaves target - current issues and a new intrinsic approach," Proc. SPIE **9023** (2014).

29 P. D. Burns and J. M. Bauza, "Intrinsic camera resolution measurement," Proc. SPIE **9396**, 939609 (2015).

30 J. McElvain, S. P. Campbell, J. Miller, and E. W. Jin, "Texture-based measurement of spatial frequency response using the dead leaves target: extensions, and application to real camera systems," Proc. SPIE **7537**, 75370D-1–75370D-11 (2010).

31 G. M. Johnson and M. D. Fairchild, "A top down description of S-CIELAB and CIEDE2000," Color Res. Appl. **28**, 425–435 (2003).

32 P. G. J. Barten, "The square root integral (SQRI): a new metric to describe the effect of various display parameters on image quality," Proc. SPIE **1077**, 73–82 (1989).

33 J. A. Movshon and L. Kiorpes, "Analysis of the development of spatial sensitivity in monkey and human infants," J. Opt. Soc. Am. **5**, 2166–2172 (1988).

34 E. W. S. Fry, S. Triantaphillidou, J. Jarvis, and G. Gupta, "Image quality optimization, via application of contextual contrast sensitivity and discrimination functions," Proc. SPIE **9396** (2015).

35 P. G. J. Barten, *Contrast Sensitivity of the Human Eye and its Effects on Image Quality* (SPIE Press, Bellingham, WA, 1999).

36 R. Branca, S. Triantaphillidou, and P. D. Burns, "Texture MTF from images of natural scenes," *Proc. IS&T CIC25: Twenty-Fifth Color and Imaging Conf.* (IS&T, Springfield, VA, 2017), pp. 113–120.

37 G. E. Legge, D. Kersten, and A. E. Burgess, "Contrast discrimination in noise," J. Opt. Soc. Am. A **4**, 391–404 (1987).

38 P. G. J. Barten, "Physical model for the contrast sensitivity of the human eye," *Proc. IS&T Eighth Int'l. Congress on Advances in Non-Impact Printing Technologies* (IS&T, Springfield, VA, 1992), pp. 57–72.

39 D. Laming, *The Measurement of Sensation* (Oxford University Press, Oxford, 1997).

40 Handbook of Photographic Science and Engineering, edited by C. N. Proudfoot, 2nd ed., (IS&T, Springfield, VA, 1997).

41 E. Allen, "*Image Quality Evaluation in Lossy Compressed Images*," Ph.D. thesis (Faculty of Media Arts and Design, University of Westminster, UK, 2017).

42 S. Triantaphillidou, E. W. S. Fry, V. Sanchis-Jurado, and A. Pons, "Image quality loss and compensation for visually impaired observers," *Proc. IS&T CIC25: Twenty-Fifth Color and Imaging Conf.* (IS&T, Springfield, VA, 2017), pp. 365-1–365-6.

43 H. S. Malvar, L. He, and R. Cutler, "High-quality linear interpolation for demosaicing of Bayer-patterned color images," *IEEE Int'l. Conf. Acoust. Speech Signal Process.* (IEEE, Piscataway, NJ, 2004), pp. 5–8.

44 Y. M. Lu, M. Karzand, and M. Vetterli, "Demosaicking by alternating projections: Theory and fast one-step implementation," IEEE Trans. Image Process. **19**, 2085–2098 (2010).

45 K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising with block-matching and 3D filtering," Proc. SPIE **6064**, 606414-1–606414-12 (2006).

46 K. He, J. Sun, and X. Tang, "Guided image filtering," IEEE Trans. Pattern Anal. Mach. Intell. **35**, 1397–1409 (2010).

47 IEC/4WD 61966-2-1: Colour Measurement and Management in Multimedia Systems and Equipment - Part 2-1: Default RGB Colour Space - sRGB. International Electrotechnical Commission (IEC) (1998).

48 J. Schneider, "Snellen Near Vision Test Card," Available: http://plantphys.info/sciencematters/snellenchart.pdf, accessed March 2018, (2002).

49 D. J. Baxter and A. Murray, "Calibration and adaptation of ISO visual noise for I3A's Camera Phone Image Quality initiative," Proc. SPIE **8293**, 829303-1–829303-14 (2012).

50 IEEE Standards Association, "IEEE Standards Downloads and Executable Files: IEEE 1858TM," Available: https://standards.ieee.org/content/dam/ieee-standards/standards/web/download/1858-2016_downloads.zip, accessed October 2018, (2016).

51 O. van Zwanenberg, S. Triantaphillidou, R. Jenkin, and A. Psarrou, "Edge detection techniques for quantifying spatial imaging system performance and image quality," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition Workshops* (IEEE, Piscataway, NJ, 2019).

52 E. W. S. Fry, S. Triantaphillidou, R. B. Jenkin, R. E. Jacobson, and J. R. Jarvis, "Noise power spectrum scene dependency in simulated image capture systems," *Proc. IS&T Electronic Imaging 2020: Image Qual. Syst. Perform. XVII* (IS&T, Springfield, VA, 2020), (in print).