Image information metrics from slanted edges:

a toolkit of metrics to aid object recognition, machine vision, and artificial intelligence systems

Norman L. Koren, Imatest LLC, Boulder, Colorado, USA

Abstract

We present a comprehensive framework for conveniently measuring camera information capacity and related performance metrics from the widely used slanted-edge (e-SFR) test pattern.

The goal of this work is to develop a set of image quality metrics that can predict the performance of Machine Vision (MV) and Artificial Intelligence (AI) systems, assist with camera selection, and use for designing electronic filters to optimize system performance. The new methods go far beyond the standard approach of estimating system performance based on sharpness and noise (or Signal-to-Noise Ratio) — which often involves more art than science.

Metrics include Noise Power Spectrum (NPS), Noise Equivalent Quanta (NEQ), and two metrics that quantify the detectability of objects and edges: Independent Observer Signal-to-Noise Ratio, SNRi, and Edge SNRi. We show how to use these metrics to design electronic filters that optimize object and edge detection performance.

The new measurements can be used to solve several problems, including finding a camera that meets performance requirements with a minimum number of pixels— important because fewer pixels mean faster processing and lower energy consumption as well as lower cost.

Introduction

We introduce the concept of information capacity, which is calculated from signal power, noise power, and bandwidth, then we describe two methods for calculating noise power (and hence information capacity) in the presence of the signal from the familiar slanted-edge (e-SFR) test pattern, specified by the ISO 12233:2014/2017/2023 standard [1].

1. The edge variance method, which calculates spatially dependent noise power, $\sigma_s^2(x)$, and
2. The noise image method, which calculates frequency-dependent noise (noise power spectrum, $NPS(f)$), as well as several additional metrics, including Noise Equivalent Quanta, *NEQ*, and metrics for object and edge detection, *SNRi*, and *Edge SNRi*.

We describe how to use these metrics to design electronic filters, called *matched filters*, for optimizing object and edge detection performance.

Information capacity

Camera information capacity, based on Claude Shannon's ground-breaking work on information theory [2-3] has long held promise as a figure of merit for a variety of imaging systems. It has been discussed in several technical papers [4-7] and in two textbooks [8-9], but it has failed to gain widespread adoption in the imaging industry, primarily because it was difficult to measure

In electronic communications systems, channel (information) capacity, C, defines the maximum rate in bits per second that information can be transmitted through a channel without error. For additive white gaussian noise, it is given by the deceptively simple Shannon-Hartley equation.

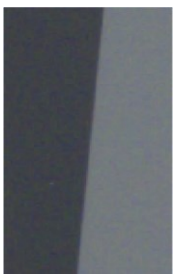
$$C = W \log_2\left(1 + \frac{S}{N}\right) = \int_0^W \log_2\left(1 + \frac{S(f)}{N(f)}\right) \qquad (1)$$

It is quite logical to apply this definition to imaging systems, where C has units of bits/pixel. Signal power, $S(f)$, and noise power, $N(f)$, must be measured with care, which was traditionally difficult and error-prone because $S(f)$ and $N(f)$ had to be measured at separate locations. To make matters worse, most JPEG images from consumer cameras usually have nonuniform image processing (bilateral filtering) [10] that sharpens images near contrasty features such as edges (boosting high frequencies) but reduces noise elsewhere (lowpass filtering). This increases the measured information capacity while removing information.

Because nonuniform image processing is so commonly applied, it is highly desirable to measure signal and noise at the same location in the image, i.e., to measure noise in the presence of signal. This is what the new slanted-edge methods accomplish.

The slanted-edge measurement

For context, we briefly review the slanted-edge algorithm.



1. **The image should be well-exposed**, avoiding the dark "toe" and light "shoulder" response regions.
2. **Linearize the image** by applying the inverse of the encoding gamma curve.
3. **Find the center of the transition** between the light and dark regions for each horizontal scan line, $y_l(x)$.
4. **Fit a polynomial curve** to the center locations.
5. **Add each appropriately shifted scan line to one of four bins**, depending on the location of the

curve relative to the scan line. [We also tried a method, described below, that interpolates each scan line.]

6. **Combine** the mean signal in each bin to obtain the 4× oversampled averaged edge for L scan lines, $\mu_s(x)$, illustrated in the upper plot of Figure 7.

$$\mu_s(\mathrm{x}) = \frac{1}{L}\sum_{l=0}^{L-1} y_l(x-\delta) \qquad (2)$$

7. **Calculate Spatial Frequency Response, *SFR(f)* (synonymous with *MTF(f)*),** by differentiating the averaged edge, windowing it, then taking the magnitude of the Fourier transform, normalized to 1 (100%) at zero frequency. Illustrated in the lower plot of Figure 7.

Nomenclature— Spatial Frequency Response (SFR) and Modulation Transfer Function (MTF) are used synonymously in the literature, but SFR is generally preferred in recent literature [1]. We use SFR here, although some plots are labeled MTF (which is more familiar) and we keep summary metrics such as MTF50, the spatial frequency where SFR drops to 50% of its zero-frequency value.

Modified ISO 12233 slanted-edge (e-SFR) calculation with interpolated binning

The slanted edge is one of several patterns for calculating SFR. The Siemens star produces smooth, consistent results, but requires far more space and computation time than slanted edges, which are small and fast, but often have somewhat rough response and noise-like artifacts at high frequencies (>0.3 C/P) when processed with the ISO 12233 binning algorithm. This can make it difficult to measure MTF10— the spatial frequency where SFR drops to 10% of its zero-frequency value, roughly equivalent to the Rayleigh diffraction limit. Figure 1 shows an example for a 12-megapixel camera with a 1-inch sensor at ISO 1600.

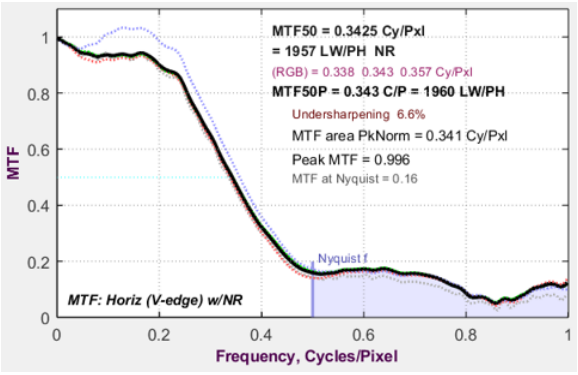


Figure 1. *SFR (MTF) for 1 inch sensor camera at ISO 1600: Current ISO 12233 binning algorithm (uninterpolated).*

The anomalous response at $f > 0.5$ C/P has little effect on common SFR summary metrics such as MTF50, which is why it has been mostly, though not entirely, ignored. But we were concerned about how it might affect the consistency of metrics such as NEQ, that depend on $K(f) = SFR(f)^2/NPS(f)$.

Figure 2. *Interpolation diagram: N to 2N-1 pixel count*

We explored a modification of the ISO 12233 algorithm that uses interpolated data. Its algorithm is simple. Before performing the binning, interpolate the N pixels each scan line to obtain $2N$-1 pixels. In MATLAB, this can be easily done with the interp2 function. 'cubic' interpolation gives good results, but not very different from 'linear,' which is faster. The polynomial fit equation and frequency scale are adjusted accordingly. The result is an SFR curve that has reduced artifacts and SFR above the Nyquist frequency, shown in Figure 3.

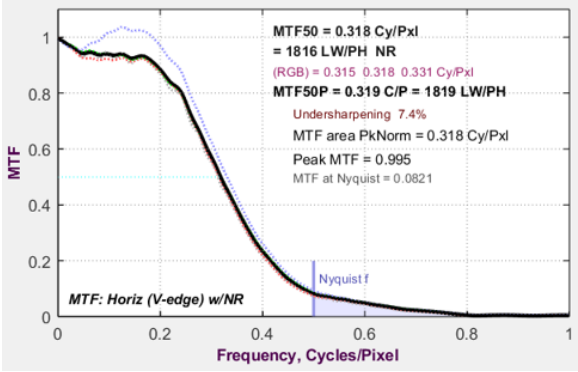


Figure 3. *SFR (MTF) for 1 inch sensor camera at ISO 1600: New interpolated binning algorithm.*

Interpolated binning reduces high frequency artifacts in SFR curves we've observed [11], ranging from sharp images (with significant energy above $f_{Nyq}/2 = 0.25\ C/P$), to motion-blurred images that suffered from sawtooth Line Spread Function artifacts. Slanted-edge results are much closer to Siemens star results for uniformly processed images, and it is now possible to reliably measure MTF10.

However, we have become concerned about the frequency response of the interpolation, which may drop off as $\mathrm{sinc}^2(fT)$ for sampling interval T. This would account for the cleaned-up high frequency response. We have also observed that the noise-like artifacts may originate with the demosaicing in some cameras, i.e., they may be real. As a result, we are giving the new method further study.

We have added the new interpolated binning technique to a variant of SFRMAT5— the free program available from Burns Digital Imaging [12] that is used as reference code for ISO 12233 standard. [11] contains more detail, and will be kept up-to-date.

The Edge Variance method for calculating noise and information capacity

We will concisely review the Edge Variance method, which was introduced in an earlier paper [13].

A simple addition to the ISO 12233 binning algorithm described above allows the ***variance*** of the signal, σ_s^2 (the noise power), to be calculated in addition to the mean, μ_s.

In addition to $\sum y_l(x)$, **calculate the sum of the squares of each scan line**, $\sum y_l^2(x)$. Then, the noise power is

$$N(x) = \sigma_s^2(\mathrm{x}) = \frac{1}{L}\sum_{l=0}^{L-1}(y_l(x) - \mu_s(x))^2 = \frac{1}{L}\sum_{l=0}^{L-1} y_l^2(\mathrm{x}) - \left(\frac{1}{L}\sum_{l=0}^{L-1} y_l(\mathrm{x})\right)^2 \quad (3)$$

$\sigma_s(x)$ is the noise amplitude, $\sqrt{N(x)}$. This equation holds for the entire oversampled array— including at the edge transition, where noise was traditionally difficult to measure.

[13] describes a form of quantization noise called *binning noise* that is largest near the image transition—where the Line Spread Function, $LSF(x) = d\mu_s(x)/dx$ is maximum. It is subtracted from $\sigma_s^2(x)$ to improve calculation accuracy. The interpolated e-SFR calculation has only half the binning noise of the original calculation. Binning noise has only a minor effect on computation accuracy.

Signal power, *S*

The peak-to-peak signal amplitude, $V_{P\text{-}P}$, (Figure 4) at low spatial frequencies is the measured difference between the means of the light and dark regions of the linearized slanted edge, $\mu_s(x)$.

$$V_{p-p} = \Delta\mu_s = \mu_{sLight} - \mu_{sDark} = V_{max} - V_{min} \quad (4)$$

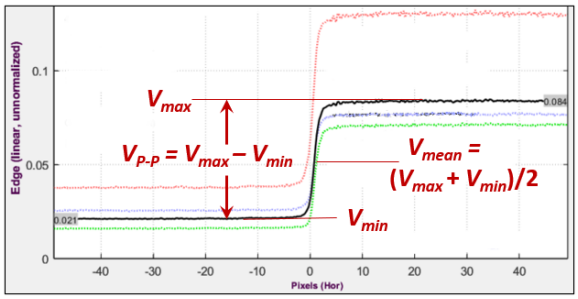


Figure 4. *Slanted-edge amplitudes (voltages, V)*

Since our intent is to calculate the information (or channel) *capacity*, which is the *maximum* information for the $V_{P\text{-}P}$ signal, we assume a signal distribution that maximizes information: the uniform distribution. The variance of the uniformly-distributed signal, which is the average signal power at low spatial frequencies, is

$$\sigma_V^2 = S_{avg}(0) = (\mu_{sLight} - \mu_{sDark})^2/12 = V_{p-p}^2/12 \quad (5)$$

The Shannon-Hartley equation uses the *average* frequency-dependent signal power, $S_{avg}(f)$.

$$S_{avg}(f) = \left(V_{p-p}\, MTF(f)\right)^2/12 \quad (6)$$

Signal power, S, is proportional to the square of the chart's Michelson contrast, $C_{Mich} = (\mu_{sLight} - \mu_{sDark})/(\mu_{sLight} + \mu_{sDark})$, for a properly linearized image, which is easy to obtain if the camera does not approach saturation at low or high pixel levels. Note that $S_{max} \leq 1$ for linearized images normalized to 1.

Noise power, *N*

Noise power, N, has the same units as signal power, S; hence S/N is dimensionless.

Depending on how the image has been processed, system performance may be dominated by noise inside or outside the edge transition region, which is defined by the Line Spread Function, $LSF(x) = d\mu_s(x)/dx$, shown in Figure 5.

For an LSF peak located at X_{LSF}, the edge transition region is defined by $X_{LSF} - PW20 < x < X_{LSF} + PW20$, where $PW20$ is the length of the region where $LSF(x) = d\mu_s(x)/dx \geq 0.20\, LSF_{max}$.

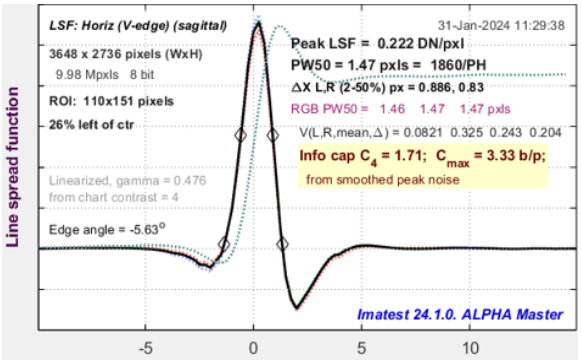


Figure 5. *Line Spread Function, LSF = dμs(x)/dx for the strongly sharpened bilateral-filtered image in the example below.*

Two image processing types cover most cases of interest. A third type is extremely rare.

1. Minimally and uniformly -processed (but not strongly sharpened) images, often TIFFs converted from raw files (raw→TIFF). Most cameras to be evaluated for Machine Vision/Artificial Intelligence are in this category. Minimally processed images can be identified by the lack of a strong noise peak near the transition.

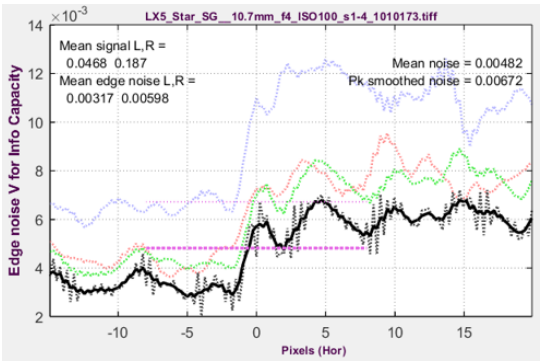


Figure 6. *Noise amplitude $\sigma_s(x)$ for uniformly processed image (TIFF from raw; no sharpening or noise reduction). ISO 100. The* ***bold black*** *curve is the Y-channel, smoothed with a 1.25-pixel kernel before 4× oversampling.*

Since noise amplitude, $\sqrt{N(x)}$, can be quite rough (Figure 6), the entire acquired edge is recommended for calculating N to enter into the Shannon-Hartley equation.

$$N_{uniform} = \text{mean}\left(\sigma_s^2(x)\right) \tag{7}$$

2. Bilateral-filtered images [10] include nearly all JPEG images from consumer cameras. Bilateral filters sharpen images near contrasty features such as edges, but blur (lowpass-filter) them to reduce noise elsewhere. This causes a distinct noise peak, shown in Figure 7, close to the edge transition, which can dominate camera performance because SFR is also measured at the transition. The noise-reduced (lowpass-filtered) region should not be used for information-related calculations. We have long known about the noise peak, but we previously had no convenient way to observe it.

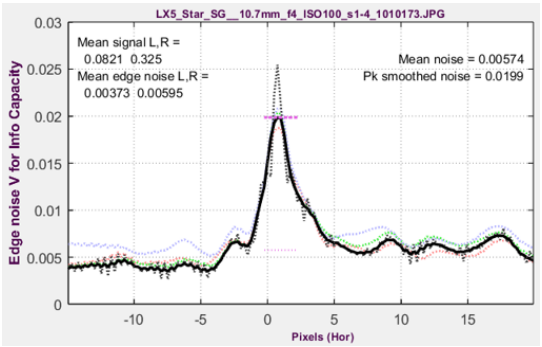


Figure 7. *Noise amplitude $\sigma_s(x)$ for bilateral-filtered image (sharpened near edges; noise-reduced elsewhere) from a camera JPEG. ISO 100..*

For bilateral-filtered images, use Noise power, N_{peak}, the square of the peak noise amplitude, smoothed with a rectangular kernel of length $PW20/2$. This is an arbitrary choice, but it removes most of the jaggedness and produces reasonably consistent results.

3. Despite our best efforts, a very few **strongly but uniformly-sharpened images** may have strong noise peaks caused by defects such as dust specks. These defects have little effect on standard SFR measurements. In the rare event that this occurs, noise $N_{uniform\text{-}sh}$ can be calculated from mean$\left(\sigma_s^2(x)\right)$ *excluding* the edge transition region, X_{LSF} – PW20 < x < X_{LSF} + PW20.

The noise ($N_{uniform}$, N_{peak}, or $N_{uniform\text{-}sh}$) to enter into the Shannon-Hartley equation is selected based the presence of a detected peak near the transition. Some additional considerations:

- Noise is close enough to white to yield good results. This assumption is supported by experimental results in [22]. The Noise Image method, below, calculates the noise spectrum.
- Noise power is larger on the lighter side of the edge due to photon shot noise, which increases with the number of photons reaching the sensor pixels. The mean, $N_{uniform}$, includes both sides.
- For linear sensors, noise power increases with exposure, following the function $N(V) = k_0 + k_1 V$, where k_1 is the coefficient for photon shot noise, derived in [13].
- N_{peak} provides a reasonable approximation to the information capacity for bilateral-filtered images, which includes almost all JPEGs from cameras. It is less accurate than measurements from minimally or uniformly processed images, but it can be useful for estimating the performance of "black box" cameras, which have unknown image processing.

Bandwidth, W

Bandwidth, W, is always 0.5 cycles/pixel (the Nyquist frequency, f_{Nyq}). Signals above Nyquist do not contribute to the information content; they can reduce it by causing aliasing— spurious low frequency signals like Moiré that can interfere with the actual image. Frequency dependence comes from $SFR(f)$, which is a component of $S_{avg}(f)$.

Combining S, N, and W to obtain information capacity, C

Once signal power, S, and noise power, N ($N_{uniform}$, $N_{uniform\text{-}sh}$, or N_{peak}, as appropriate), have been obtained, information capacity, C, can be calculated.

$$C = \int_0^{0.5} \log_2\left(1 + \frac{S_{avg}(f)}{N}\right) df \cong \sum_{i=0}^{0.5/\Delta f} \log_2\left(1 + \frac{S_{avg}(i\Delta f)}{N}\right) \Delta f \tag{8}$$

Edge variance results: Edge, *SFR*, *C*

Figure 8 shows the Edge and SFR response as well as calculated information capacity values (C_4 and C_{max}, to be

introduced below). Similar plots in [13], made with the uninterpolated e-SFR calculation, are rougher, as expected.

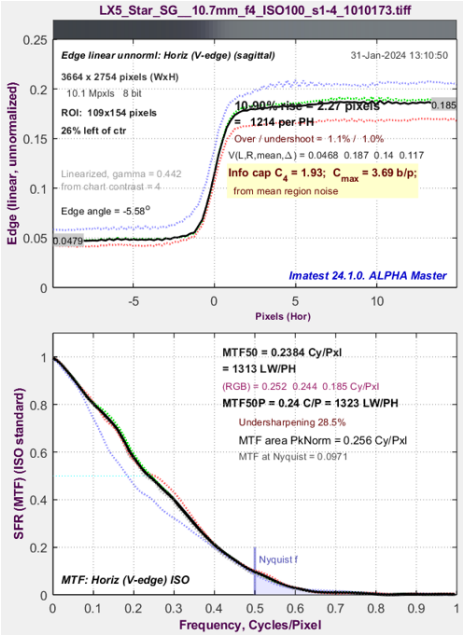


***Figure 8**. Edge and SFR (MTF) plot for compact digital camera for an unsharpened TIFF from raw. Upper: Mean edge $\mu_s(x)$. Lower: SFR(f). **C_4** is the Shannon information capacity for a 4:1 contrast ratio edge.*

SFR(*f*), which is sometimes confused with bandwidth, can take a large bite out of *C*, especially since it is squared in the above equation. [13] contains an explanation of how increasing SFR can lead to significant aliasing-related artifacts, such as Moiré, that degrade performance.

Measurement technique

Test chart edge contrast should be between 2:1 and 10:1, with 4:1 (specified in the ISO 12233 e-SFR standard) recommended. Edge contrast greater than 10:1 increases the likelihood of nonlinear operation (saturation or clipping), that will compromise the results.

Images should be well-exposed because saturation or clipping can cause misleading results. It may be of interest, however, to measure *C* as a function of exposure.

The camera should be well-focused. Sturdy camera support should be employed.

Although results are relatively insensitive to ROI selection, some care must be taken to obtain good consistency. ROIs should be reasonably large; at least 30x60 pixels is recommended. If possible, the edge should be centered in the selected region, and there should a reasonable amount of "breathing room" on the sides.

Additional assumptions

A key assumption is that the camera's dynamic range (the range of tones that can be reproduced with good contrast and Signal-to-Noise Ratio (SNR)) is sufficient for the intended task. Most modern image sensors have dynamic ranges greater than 60dB (1000:1); high dynamic range (HDR) sensors have 120 dB or more. The majority of scenes in pictorial, medical, or robotic (but *not* automotive) imaging have tonal ranges under 60 dB. Lens flare (stray light) typically limits practical camera dynamic range to under 100 dB, which can impact automotive night driving by fogging important dark to middle tones. If there are concerns about dynamic range, we strongly recommend measuring it with a transmissive chart.

Other assumptions: sensor nonuniformities (fixed-pattern noise, also called PRNU (Photo Response Nonuniformity) are included in noise measurements. Tonal response is well-behaved (typically following a gamma curve, except for the extreme highlights and shadows). Stray (flare) light is not too severe where *SFR* is measured.

Because the measured value of C is closely tied to the n:1 chart contrast ratio, where $n \leq 10$ to minimize saturation or clipping, n should be specified when C is reported, e.g., C_4 for charts with a 4:1 contrast ratio.

Sensitivity to exposure

Because both noise power, *N*, and amplitude range, ΔV, increase with exposure, **C_4 is a strong function of exposure,** as illustrated in Figure 9**.**

Consistent exposure can be difficult to achieve with autoexposure consumer cameras because their JPEG output files often have "shoulders" in their tonal response (regions of reduced highlight contrast intended to improve pictorial quality by minimizing saturated ("burnt out") highlights).

Implementing a shoulder requires extra headroom, i.e., a degree of underexposure, which can vary for different camera models. Since autoexposure is optimized for JPEG output, minimally processed files, typically TIFFs converted from raw with simple gamma curves (raw→TIFF), often appear to be underexposed.

Maximum information capacity C_{max} — a more stable metric than C_4

Because the strong exposure-dependence of C_4 (Figure 9) affects its value as a performance metric, we have developed a stable metric for maximum information capacity, ***C_{max}***, that is nearly independent of exposure. It is obtained in two steps.

Step1: Replace the measured peak-to-peak amplitude range, $V_{P\text{-}P}$, with the maximum allowable value, $V_{P-P\_max} = 1$ (for systems normalized to a maximum

amplitude of 1). This may seem like a simplification, but it works well for most cameras. Referring to the section on Noise Power, N,

$$S_{avg}(f) = \left(V_{p-p\_max}\, MTF(f)\right)^2/12 = MTF(f)^2/12 \quad (9)$$

Step 2: Replace the measured noise power, N, with N_{mean}, the mean of N over the range $0 \le V \le 1$ (where 1 is the maximum allowable normalized signal amplitude V). The general equation for N for linear image sensors is

$$N(V) = k_0 + k_1 V \quad (10)$$

Equations for k_0 and k_1 and an adjustment to C_{max} for bilateral-filtered images (which are less accurate than for minimally processed images) are derived in [13].

C_{max} (Figure 9) is nearly independent of exposure for minimally or uniformly-processed images with linear sensors, where noise power, N, is a known function of signal amplitude, V, but it is only approximate for imaging systems with bilateral filtering or HDR (nonlinear) sensors, where noise power N is not a simple function of V.

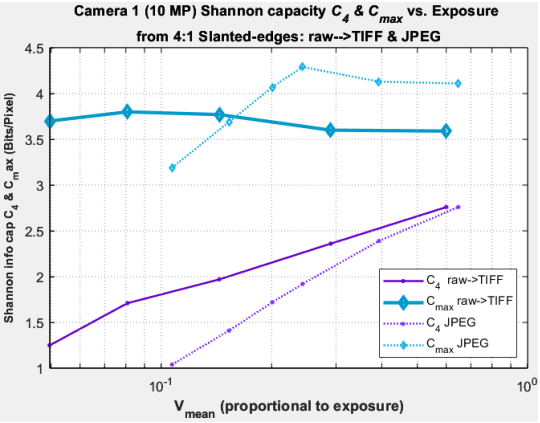


Figure 9. *C_4 and C_{max} for minimally processed raw→TIFF and JPEG images for a 10 MP compact camera. C_{max} is consistent, especially for the raw→TIFF image.*

High Dynamic Range (HDR) images

Special care must be taken when calculating C_{max} for HDR sensors, which have several cycles of SNR and noise as exposure increases [14].

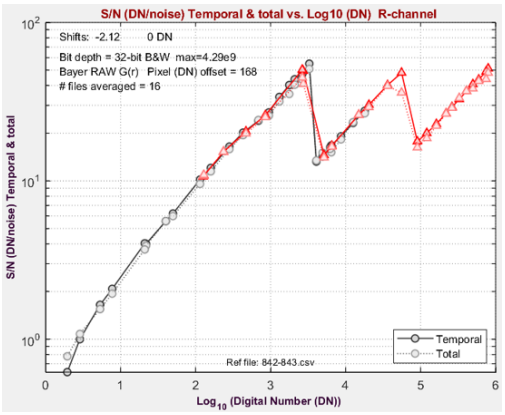


Figure 10. *Cyclic response of Signal-to-Noise Ratio for HDR sensor*

Noise $N(V)$ increases monotonically, but jumps at the discontinuities of the SNR plot (Figure 10), around $\text{Log}_{10}(DN) = 3.6$ and 4.8. The noise measurement depends on the location on the sawtooth curve where the measurement is made. Because of this and because each HDR sensor is different, there is no simple equation, comparable to (10) for calculating C_{max}. It will require a separate measurement and an assumption about the maximum SNR, perhaps limiting it to the mean value in the sawtooth region (above $\text{Log}_{10}(DN) = 3$ in Figure 10). For now, we recommend caution when calculating C_{max} for HDR sensors.

Information capacity results

Table 1 shows three cameras with both raw and JPEG output that we tested for information capacity as a function of Exposure Index (ISO speed setting).

Table 1. Cameras used in the tests

1.	**Panasonic Lumix LX5**	2.14 µm pixel pitch. Compact 10.1-megapixel camera with a Leica f/2 zoom lens set to f/4.
2.	**Sony A6000**	3.88 µm pixel pitch. 24-megapixel micro four-thirds camera
3.	**Sony A7Rii**	4.5 µm pixel pitch. A 42-megapixel full-frame camera with a Backside-Illuminated (BSI) sensor

The image in Figure 11, which was analyzed in [15], contains a 50:1 contrast Siemens star and four 4:1 contrast slanted edges. We used the upper-left slanted edge for most tests. The average background of the chart is close to neutral gray (18% reflectance) to ensure a good exposure.

Figure 11. *Typical image (cropped) including Siemens star and slanted-edges to the left and right of the star.*

We captured both JPEG images and raw images, converted by LibRaw to 24-bit sRGB TIFF (designated as raw→TIFF) with minimal processing (no sharpening, no noise reduction, and simple gamma-encoding). The luminance channel (Y = 0.2125×R + 0.7154×G + 0.0721×B) was analyzed. Results with 48-bit Adobe RGB conversion were similar.

Figure 12 shows C_4 as a function of ISO speed (Exposure Index, which is proportional to analog gain) for raw→TIFF images (solid lines) and JPEG images (dotted lines). For the raw→TIFF images, the relationship between ISO speed and C is similar for all three cameras.

$N_{uniform}$ was used for the raw→TIFF images; N_{peak} was used for the bilateral-filtered JPEGs.

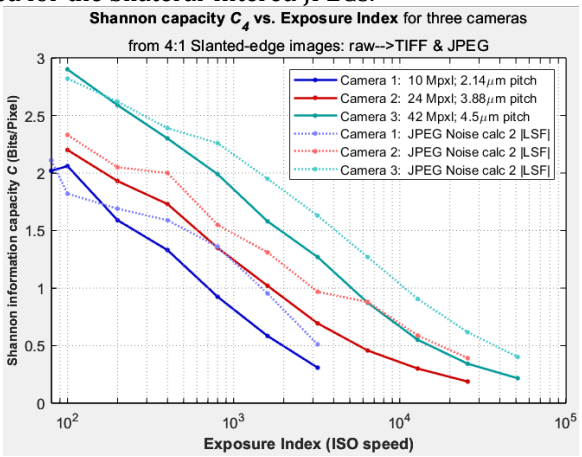


Figure 12. *Information capacity, C_4, from 4:1 slanted-edge images. Solid lines for raw→TIFF images; Dotted lines for JPEGs.*

C_{max} has a similar trend to C_4, but is higher about 1.6. [13] contains more detail on the behavior of C_{max}.

Color channels

The separate R, G, and B channels tend to have slightly lower C_4 than the Y-channel because the noise in the separate channels is uncorrelated. Color is discussed in more detail in [13].

Although this paper has focused on demosaiced images, the slanted-edge method can also be applied to raw (undemosaiced) images.

Effects of sharpening

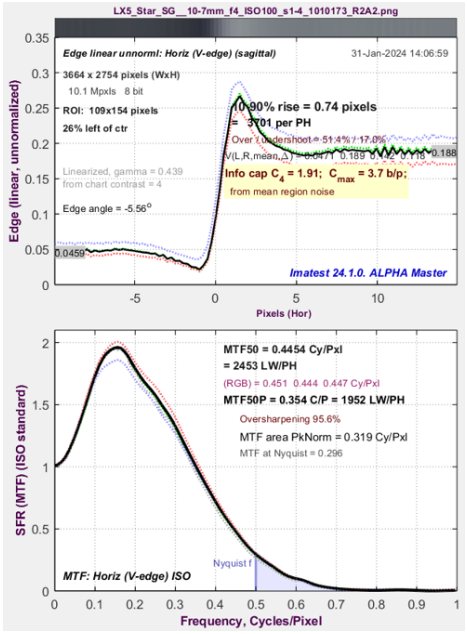


Figure 13*. Edge/SFR (MTF) plots derived from the same image as Figure 7, where C_4 = 1.93 b/p and C_{max} = 3.69 b/p, raw→TIFF, ISO 100 Sharpening Radius = 2; Amount = 2. C_4 = 1.91 b/p; C_{max} = 3.7 b/p.*

The examples in Figures 8 and 13 (and many others we ran) show that sharpening has little effect on slanted-edge information capacity, as expected. The image used for Figure 13 (initially a minimally-processed TIFF) has been strongly Unsharp Mask (USM) sharpened with Radius = 2 and Amount = 2 (R2A3). It can be compared to Figure 8, where $C_4 = 1.93$ and $C_{max} = 3.69$ b/p. We observed a similar insensitivity of C to sharpening with Siemens stars [15].

Total information capacity

The total information capacity, C_{total}, for the entire image is calculated from

$$C_{total} = \text{mean}(C) \times \text{megapixels} \qquad (11)$$

From Figure 14, the mean value of C_{max} is 2.847 bits/pixel. Since this camera has 16 Megapixels, the total capacity, $C_{maxTotal}$, for the Luminance (Y) channel = 45.55 MB.

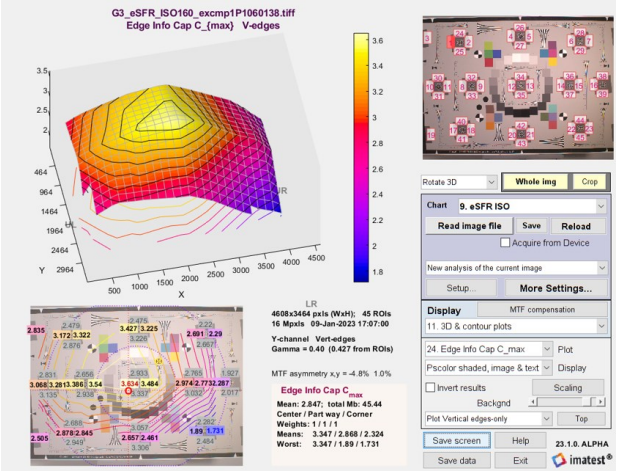


Figure 14. *3D contour eSFR ISO plot of Cmax for the Luminance (Y) channel, ISO 100*

Summary of the Edge variance method

The edge variance method calculates the spatially dependent noise in addition to the signal from a slanted edge. A mean (or peak value) of the noise is substituted into the Shannon-Hartley equation to calculate information capacity, C_n (measured directly from an n:1 contrast chart) or C_{max} (the maximum information capacity, which is independent of exposure and chart contrast).

The Noise Image method, below, calculates the noise spectrum, which is used to derive several useful metrics related to edge and object detection. The noise spectrum is normalized using the edge variance noise.

The Noise Image Method

The noise image method is the second of two methods for calculating noise and image information metrics. It calculates the frequency-dependent noise (the Noise Power Spectrum, $NPS(f)$) instead of the spatially dependent noise power, $\sigma_s^2(x)$. This enables the calculation of a particularly rich set of metrics.

The method involves inverting the ISO 12233 binning procedure. Noting that the 4× oversampled edge was created by interleaving the contents of 4 bins, each of which contains an averaged (noise-reduced) signal derived from the original image, we apply an inverse of the binning algorithm to set the contents of each scan line to its corresponding interleave (Inverse-binned, below). Since the inverse-binned image is a nearly noiseless replica of the original image, we can create a noise image by subtracting the inverse-binned image from the original image (which must be corrected for illumination nonuniformity in the direction of the edge).

The three images are shown in Figure 15. The other images are displayed with gamma-correction.

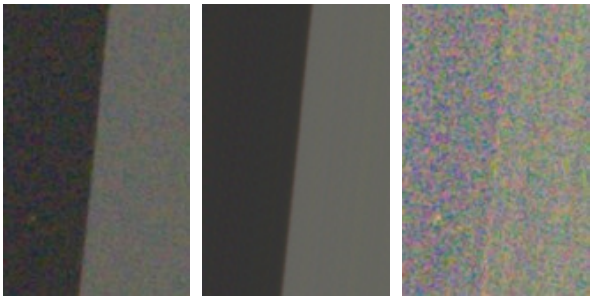


Figure 15. *Noise image method, for a noisy (high ISO speed) image*

$$\text{Noise image} = \text{Original image} - \text{Inverse-binned image} \qquad (12)$$

The noise image, which has a mean of zero, is displayed with an offset, lightened, and boosted in contrast for visibility.

These images allow several additional image quality parameters to be calculated, including Noise Power (Wiener) Spectrum ($NPS(f)$) and Noise Equivalent Quanta ($NEQ(f)$), well-known in medical imaging systems, and described in an excellent review paper by Ian Cunningham and Rodney Shaw [16], and also in the obscure but valuable ICRU Report 54 [17]. (ICRU is the International Commission on Radiation Units & Measurements.) These measurements are little-known outside of medical imaging, in part because they have been difficult to measure.

One caution is in order: the Noise Image method is invalid for bilateral-filtered images and has limited value for images that have been noise-reduced.

Noise Power Spectrum (*NPS*)

$NPS(f)$, also called the *Wiener spectrum* (Figure 16), is used in the calculation of the key information metrics. The Noise Amplitude (Voltage) Spectrum, $N_V(f) = \sqrt{NPS(f)}$ is also of interest.

The 1D Noise Power or Voltage spectrum is derived from a 2D Fourier transform (FFT) of the noise image.

- Noting that $f = 0$ at the center of the 2D FFT image (after applying the MATLAB fftshift function), divide it into several annular regions, and find the average noise power for each region. This procedure has been used for the Imatest Spilled Coins/Dead Leaves calculations since 2013, and has been tested thoroughly.
- Because this procedure does not maintain the invariance in energy between the spatial and frequency domains implied by Parseval's theorem, $NPS(f)$ is normalized so that the one-dimensional integrals in frequency and spatial domain are equal.

$$\int NPS(f)\, df = \int \sigma_s^2(x)\, dx = \int N(x)\, dx \qquad (13)$$

Even though this paper focuses on metrics derived from the noise image method, the spatially dependent noise from the Edge variance method, $\sigma_s^2(x)$, is useful for two purposes.

- as an indicator of the type of image processing.
- to normalize *NPS*(*f*), ensuring consistent scaling, even for small ROIs, where the de-binned image is not entirely noiseless.

Demosaicing typically causes the Noise Power Spectrum to drop to about half its low frequency value at the Nyquist frequency ($f_{Nyq} = 0.5$ C/P).

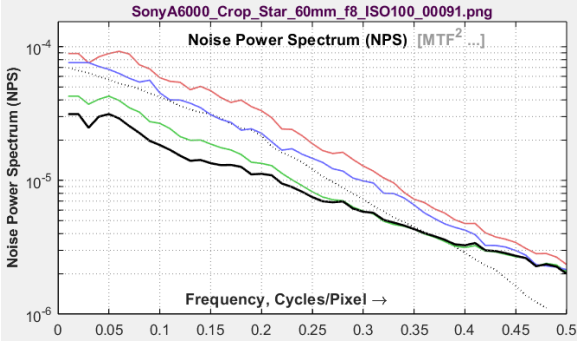


Figure 16. *Noise Power Spectrum (NPS(f))*

NPS(*f*) is used to calculate most of the image information metrics introduced in this paper— *NEQ*, C_{NEQ}, *SNRi*, *Edge SNRi*, and matched filter transfer functions.

It is a part of the ***kernel k*** that appears in the equations for most of the metrics.

$$K(f) = SFR^2(f)/NPS(f) \qquad (14)$$

Equations will be written in standard form, then with *K*(*f*). Because uniform filtering affects *SFR²*(*f*) and *NPS*(*f*) identically, *K*(*f*) is not affected by uniform filtering, such as sharpening or lowpass filtering.

Noise autocorrelation, which is inverse Fourier transform (IFT) of *NPS*(*f*), is potentially useful for evaluating the crosstalk between image sensor pixels, but the Bayer Color Filter Array (CFA) makes such measurements challenging.

Noise Equivalent Quanta, *NEQ*

NEQ(*f*) (Figure 17) is a frequency-dependent Signal-to-Noise (power) Ratio, related to the number of quanta that would result in the measured SNR when photon shot noise is dominant. It was described in 1999 by Cunningham and Shaw [16] and in 2016 by Keelan [18], and it is used in medical imaging [16, 17, 19].

$$NEQ(f) = \frac{\mu^2\, SFR^2(f)}{NPS(f)} = \mu^2\, K(f) \qquad (15)$$

where the mean linear signal, μ, can be defined in either of two ways, depending on how *NEQ* is to be applied.

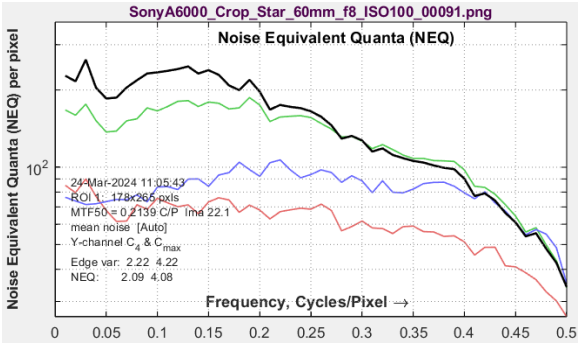


Figure 17. *Noise Equivalent Quanta (NEQ(f))*

In the standard definition of *NEQ*, $\mu^2 = V_{mean}^2$ (Figure 4). If *NPS* is dominated by photon shot noise, we can use the well-known property of the Poisson distribution, $SNR^2 = \bar{q}$ (where $\bar{q}$ is the mean count of the detected quanta.) to show that $NEQ(0) = \bar{q}$. For example, *NEQ*(0) = 200 corresponds to a mean of $\bar{q} = 200$ detected quanta per pixel. But because nonuniform illumination can increase NPS at the lowest spatial frequencies, causing NEQ to decrease, the maximum value of *NEQ* appears to be better for estimating $\bar{q}$.

The above equation can be used for calculating Detective Quantum Efficiency), $DQE(f) = NEQ(f)/\bar{q}_i$, where $\bar{q}_i$ is the mean number of quanta *incident* on each pixel. This requires a separate (and exacting) measurement of $\bar{q}_i$. The two levels of the slanted edge make *DQE* measurements particularly challenging.

NEQ(*f*) is not affected by electronic filtering.

Information capacity from NEQ, C_{NEQ}

$NEQ_{info}(f)$ (a variant of *NEQ*), calculated from (15) using $\mu = V_{P-P}/\sqrt{12}$ (to be consistent with the Edge Variance calculation for uniformly distributed levels), is signal power divided by noise power. It can therefore be substituted into the Shannon-Hartley equation to calculate information capacity, C_{NEQ}.

$$C_{NEQ} = \int_0^W \log_2\left(1 + NEQ_{info}(f)\right) df = \int_0^{0.5} \log_2\left(1 + \mu^2\, K(f)\right) df \qquad (16)$$

where $W = f_{Nyq} = 0.5$ Cycles/Pixel. [Author's note: I thought I discovered this connection, but I found it in papers on PET scanners and Digital Mammography by Christos Michail et. al. [20-21]— almost certainly unknown outside medical imaging.]

The key results, $C_4(NEQ)$ and $C_{max}(NEQ)$, are included in the Results summary (Figure 18). They are slightly different from the Edge Variance results, most likely because $NPS(f)$, is used. (The Edge Variance calculation assumes constant NPS, i.e., white noise.)

Channel	R	G	B	Y
Info capacity C_{Max}(EdgeVar) =	3.54	4.11	3.76	4.23
Info capacity C_4(EdgeVar) =	1.63	2.12	1.71	2.22
Info capacity C_{Max}(NEQ) =	3.87	4.57	4.02	4.66
Info capacity C_4(NEQ) =	1.61	2.26	1.72	2.36

Figure 18. *Information capacity C from the two methods: Edge variance and Noise image.*

Ideal Observer SNR (*SNRi*)

SNRi (Figure 20) is a measure of the detectability of objects. It was introduced and rigorously correlated with Bayesian detection statistics in the 1996 ICRU Report 54 [17], then reintroduced to the imaging community Paul Kane [23] and Orit Skorka and Paul Kane [22]. The two-dimensional equation in [22] gives the best results.

$$SNRi^2 = \iint \left(\frac{\left|G(\nu_x,\nu_y)\right|^2 SFR^2(\nu_x,\nu_y)}{NPS(\nu_x,\nu_y)} \right) d\nu_x\, d\nu_y = \iint \left|G(\nu_x,\nu_y)\right|^2 K(\nu_x,\nu_y)\, d\nu_x\, d\nu_y \qquad (17)$$

where $G(\nu_x,\nu_y)$ is the two-dimensional Fourier transform of the rectangular object to be detected, $g(x,y)$, and $SFR(\nu)$ and $NPS(\nu)$ are defined in one dimension for spatial frequency $\nu = \sqrt{\nu_x^2 + \nu_y^2}$, which has units of Cycles/Pixel.

Objects to be analyzed are typically rectangles of dimensions $w \times kw$, where $k = 1$ for a square or 4 for a 1:4 aspect ratio rectangle. Amplitude, V_{P-P}, is typically taken from the chart (typically with a 4:1 contrast ratio). The equation for the rectangular object (Figure 19) is

$$g(x,y) = V_{P-P} \cdot \text{rect}\left(\frac{x}{w}\right) \cdot \text{rect}\left(\frac{y}{kw}\right) \qquad (18)$$

where rect(x/w) = 1 for $-w/2 < x < w/2$; 0 otherwise

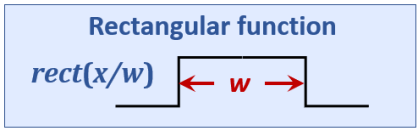


Figure 19. *rect function*

$$G(\nu_x,\nu_y) = kw^2\, V_{P-P} \frac{\sin(\pi w \nu_x)}{\pi w \nu_x} \frac{\sin(\pi k w \nu_y)}{\pi k w \nu_y} = V_{P-P}\, G_{rect}(\nu_x,\nu_y) \qquad (19)$$

where $G_{rect} = w\ \text{sinc}(\omega w/2) = w\ \text{sinc}(\pi w \nu)$ is the Fourier transform of rect(x/w) for frequency ν. Note that G_{rect} has units of $1/\nu^2$, and since ν has units of cycles/pixel, $G(\nu_x, \nu_y)$ has units of pixels^2.

$SNRi^2$ is calculated numerically by creating a two-dimensional array of frequencies (0 to 0.5 c/p in 51 steps) with ν_x on the x-axis and ν_y on the y-axis, filled with frequency $\nu = \sqrt{\nu_x^2 + \nu_y^2}$. These frequencies are used to create a 2D array that can be numerically summed [23].

$$SNRi^2 = \Delta\nu_x\, \Delta\nu_y \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \frac{SFR^2(i,j)\, G^2(i,j)}{NPS(i,j)} \qquad (20)$$

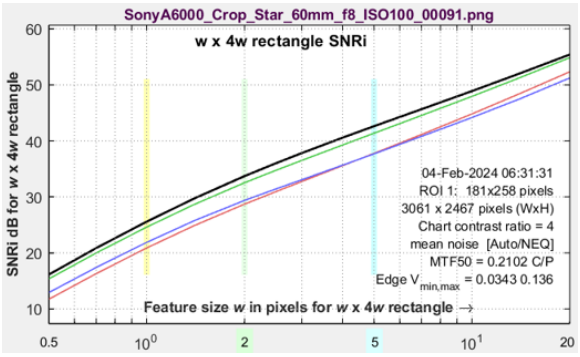


Figure 20. *SNRi for a sharp, low-noise (ISO 100) image*

Note that like C_4, $SNRi$ is strongly affected by exposure and chart contrast. But unlike C_4, $SNRi$ is also affected by Image Signal Processing (ISP; electronic filtering such as sharpening or lowpass filtering).

Although $SNRi$ is a powerful measurement, we also give weight to a closely related measurement, *Edge SNRi*, for determining the performance of ISP applied before sending the image to the Object Recognition/Machine Vision/AI block).

As a result of Parseval's theorem, which states that the integrals of a Fourier transform pair, $r(x)$ and $R(\omega)$, must be equal,

$$\int_{-\infty}^{\infty} |r(x)|^2 dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} |R(\omega)|^2 d\omega = \int_{-\infty}^{\infty} |R(2\pi f)|^2 df \qquad (21)$$

$SNRI^2$ is equivalent to the total (integrated) noise-whitened Signal/Noise energy of the object in the spatial domain.

SNRi displayed in dB per pixel squared

Because standard *SNRi* plots can be difficult to read (in part because *SNRi* has units of pixels^2), *SNRi* can also be plotted in dB per pixel^2 (Figure 21). It is somewhat easier to read than the standard *SNRi* image, but it is more of a *relative* measurement— useful for evaluating changes from image processing.

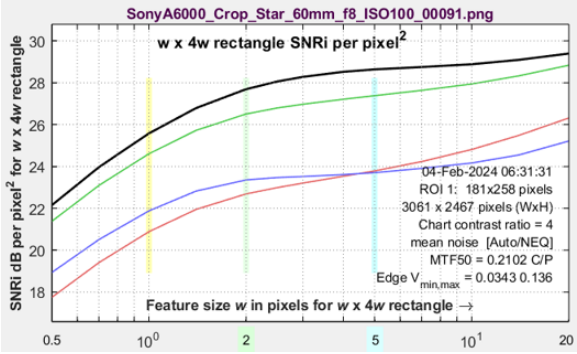


Figure 21. *SNRi per pixel^2 for a sharp, low-noise (ISO 100) image*

Object visibility and *SNRi*

One of the uses of *SNRi* measurements is to predict object visibility for small, low contrast squares or rectangles. The *SNRi* prediction begs for visual confirmation.

We have developed a display that does this with real slanted-edge image data. Despite the trickery, the data is directly from the acquired image.

We show two images, below: Figure 22 for a relatively low-noise image and Figure 23 for a noisy image (both from a camera with Micro Four-Thirds sensors, at ISO 100 and 12800). The sides of the squares are $w = 1$, 2, 3, 4, 7, 10, 14, and 20 pixels. The original chart has a 4:1 contrast ratio (light/dark = 4), equivalent to a Michelson contrast C_{Mich} = (light-dark)/(light+dark) = 0.6. The outer squares have $C_{Mich} = 0.6$. The middle and inner squares have $C_{Mich} = 0.3$ and 0.15, respectively.

How to use these images

The yellow numbers are the square widths in pixels. The outer (left and right) patches correspond to the *SNRi* curves for the ISO 12233-standard 4:1 contrast ratio, where, according to the Rose model [16], *SNRi* of 5 (14 dB) should correspond to the threshold of visibility.

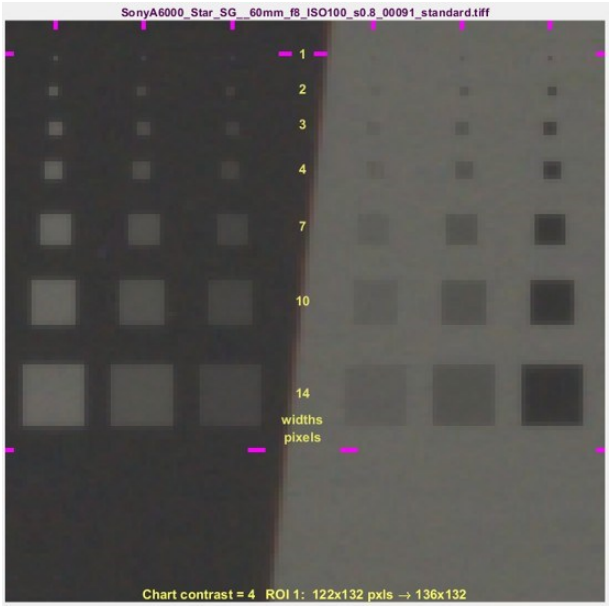


Figure 22. *Low noise image, ISO 100.*

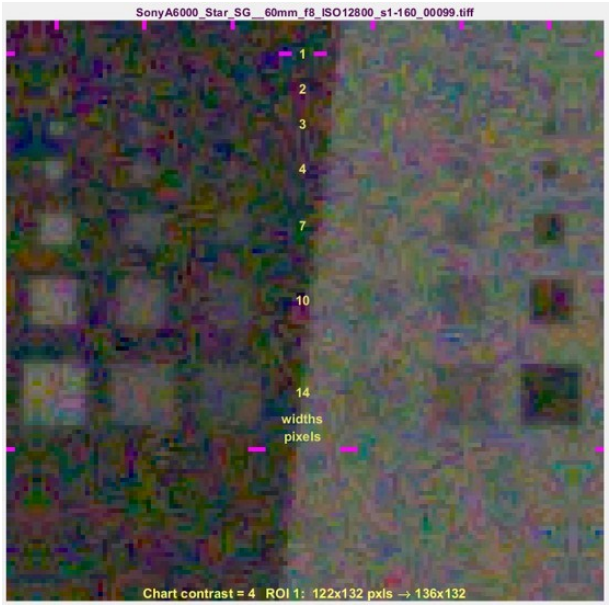


Figure 23. *Noisy image, ISO 12800.*

The *SNRi* curve in Figure 24 is for the noisy ISO 12800 image in Figure 23, above. The $w = 1$ squares are invisible; the $w = 2$ and 3 squares are only marginally visible, and $w = 4$ squares are clearly visible. In Figure 23, the Y (luminance) channel *SNRi* at $w = 2$ is 15 dB; it reaches 19 dB for $w = 3$; a little above the expectation that the threshold of visibility is around 14 dB.

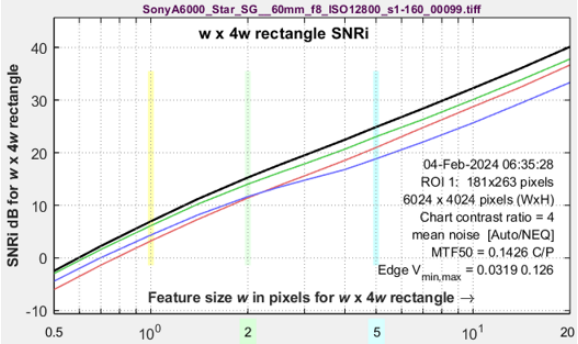


Figure 24. *SNRi for noisy ISO 12800 image (above)*

Only original pixels were used in these images of squares, but we used some "smoke and mirrors" tricks to make squares that have the same blur as the original image

How the squares were made

1. Expand the image if needed to make room for all the squares by adding mirrored versions of image to the sides, top, and bottom of the image.
2. Create a (horizontal) mirror of the full image. This is the "mirror" part.
3. Create a mask consisting of ideal $w \times w$ squares, with 0 in the background, 1 in the squares, and sharp sides.
4. Blur the squares with the MATLAB filter2 function. This is the "smoke" part. Determining the blur kernel was challenging. We found that we couldn't get good results by just using the 1D Line Spread function (LSF) in 2D. A more complex transformation was required.
5. Linearize the two images (remove the gamma encoding).
6. Combine them using the mask, using the original image where the mask = 0, the mirrored image where the mask = 1, and blending them elsewhere.
7. Reapply the gamma encoding.

Edge Signal-to-Noise Ratio (*Edge SNRi*)

Edge SNRi is a measure of the detectability of the *edges* of objects. It is similar to *SNRi*, described above and in [17, 22, 23].

$$Edge\ SNRi^2 = \iint \left(\frac{\left|H(\nu_x,\nu_y)\right|^2 SFR^2(\nu_x,\nu_y)}{NPS(\nu_x,\nu_y)} \right) d\nu_x\, d\nu_y = \iint \left|H(\nu_x,\nu_y)\right|^2 K(\nu_x,\nu_y)\, d\nu_x\, d\nu_y \qquad (22)$$

$H(\nu_x,\nu_y)$ is the Fourier transform of the *edges* (the gradient) of the object to be detected.

For a rectangle of dimensions $w \times kw$, the function is the derivative, $h(x, y)$, of the rectangle, $g(x, y)$, that describes the object.

V_{P-P} is typically obtained from a chart with a 4:1 contrast ratio. *SNRi* and *Edge SNRi* are both proportional to the Michelson contrast of the chart ($(n\text{-}1)/(n\text{+}1)$), and can be scaled for different contrast levels.

$$h(x,y) = V_{P-P} \cdot d\left[\text{rect}\left(\frac{x}{w}\right)\right]/dx \cdot d\left[\text{rect}\left(\frac{y}{kw}\right)\right]/dy = V_{P-P} \cdot I_I\left(\frac{x}{w}\right) \cdot I_I\left(\frac{y}{kw}\right) \qquad (23)$$

where $I_I(x/w) = d(\text{rect}(x/w)/dx$ (Figure 25) is called the "odd impulse pair," consisting of a pair of Dirac delta functions of opposite polarity separated by the object width w.

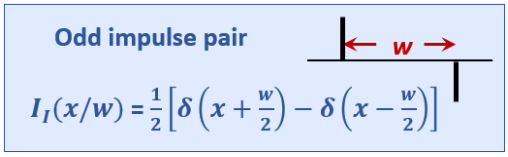


Figure 25*. Odd impulse pair*

$H(\nu_x,\nu_y)$ is the Fourier transform of the edges of the object to be detected, equivalent to $2\pi\nu\, G(\nu_x,\nu_y)$ for frequency ν. Expressed in two dimensions,

$$H(\nu_x,\nu_y) = 2\, V_{P-P} \sin(\pi w \nu_x) \sin(\pi k w \nu_y) \qquad (24)$$

*Edge SNRi*² (Figure 26) is numerically calculated using a similar equation to *SNRi*².

$$Edge\ SNRi^2 = \Delta\nu_x\, \Delta\nu_y \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \frac{SFR^2(i,j)\, H^2(i,j)}{NPS(i,j)} \qquad (25)$$

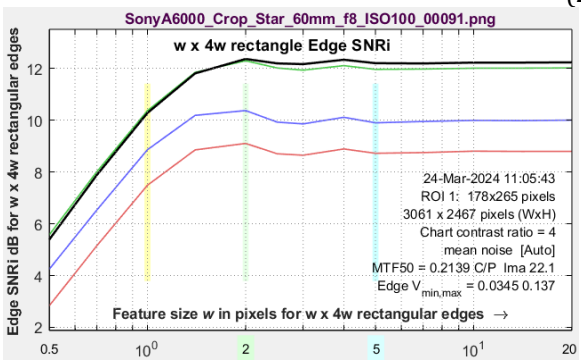


Figure 26. *Edge SNRi for a sharp, low-noise (ISO 100) image*

SNRi and *Edge SNRi* are affected by signal processing (sharpening, etc.), making them useful for evaluating filtering (ISP filtering applied prior to the object recognition/machine learning/AI blocks).

Line Spread Function (LSF) doublet results

Edge SNRi is based on pairs of Line Spread Functions of opposite polarity called *LSF doublets*, $r(x)$ (Figure 27), which are used in several key calculations.

$$r(x) = (LSF(x) - LSF(x-w))/\sigma \quad \text{and}$$

$$R(v) = \frac{H(v)\, MTF(v)}{\sqrt{NPS(v)}} \qquad (26)$$

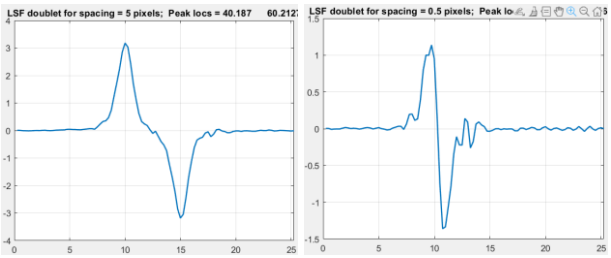


Figure 27. *LSF doublets. w = 5.0 pixels (left), w = 0.5 pixels (right) Amplitude for w = 0.5 is 1/3 as large as for w = 5.0 pixels.*

As spacing *w* decreases, the peaks are shifted more from their original locations and amplitude decreases.

As a result of Parseval's theorem, *Edge SNRI²*, which is defined in frequency domain, is equivalent to the total (integrated) Line Spread Function doublet energy divided by Noise energy in the spatial domain.

Effects of Image Signal Processing (ISP)

Several recent papers [24-26] state that appropriate image processing prior to Object Recognition, Machine Vision or Artificial Intelligence algorithms may improve the performance (accuracy, speed, and power consumption) of AI systems. Because information capacity is independent of Image Signal Processing— at least with ISP that does not remove information, such as sharpening— it provides little guidance about filter design for optimal image processing.

SNRi has some drawbacks for predicting the quality object detection. It indicates how well the presence of an object can be detected, but it says nothing about its *shape*. Shape detection is dependent on the edge detection, which is quantified by *Edge SNRi*. And there is the problem of object color. What If the object has the same color as the background? (Think of gray cars in front of gray concrete.) In such cases it is the *edge* that matters. For this reason, *Edge SNRi* should be given comparable weight to *SNRi* when designing filters.

Image signal processing algorithms can be designed to optimize a specific task, for example, the detection of an object of a specific size, often a small rectangle, or the detection of its edges. In practice, ISP needs to perform a multitude of tasks: detecting objects and edges greater than a minimum size and limiting interference from neighboring objects.

The results in the table below were obtained starting with an unsharpened image, applying various combinations of sharpening (Radius R, Amount A) and gaussian lowpass filtering (blurring) to the image, which was linearized prior to filtering, then restored to its original gamma encoding afterwards. SNRi and Edge SNRi are for $w \times 4w$ rectangles (w in pixels). $N_{uniform}$ was used for these images, which had no noise peak. (A peak was only present in the JPEG, which was not analyzed.) Standard (non-interpolated) binning was used.

Table 2. Results for various filters for a 24 MP Micro Four-Thirds camera, 1 inch sensor, ISO 100, *w* × 4*w* rectangle.

Filter (ISO 100)	MTF50 C/P	Edge SNRi *w* = 1	Edge SNRi lg *w*	SNRi dB/Pxl² w = 1	SNRi dB/Pxl² w = 5	C_{max} (NEQ)
None	0.214	10.3	12.2	25.5	28.6	4.08
σ=07 LPF	0.160	10.0	12.9	27.0	30.5	3.91
Sharp R2A3	0.541	8.843	10.4	23.5	26.2	3.8
Sharp R2A3, σ=0.7	0.335	9.47	11.7	25.1	27.8	3.86
Sharp R1A2 σ=0.7	0.265	10.4	12.4	25.8	28.7	4.09

Table 3. Results for various filters for a 24 MP Micro Four-Thirds camera, ISO 800, *w* × 4*w* rectangle.

Filter (ISO 800)	MTF50 C/P	Edge SNRi w = 1	Edge SNRi lg w	SNRi dB/Pxl² w = 1	SNRi dB/Pxl² w = 5	C_{max} (NEQ)
None	0.216	2.30	4.54	18.2	22.2	2.96
σ=0.7 LPF	0.154	3.69	6.44	20.4	24.5	2.87
Sharp R2A3	0.527	1.86	3.54	16.9	20.3	2.82
Sharp R2A3, σ=0.7	0.342	3.06	5.00	18.4	21.7	2.83
Sharp R1A2 σ=0.7	0.275	3.04	5.10	18.7	22.4	2.96

We can make several observations from these results.

- A pure lowpass filter (LPF $\sigma = 0.8$) improves *Edge SNRi* and *SNRi*.
- Pure sharpening (R2A3) degrades both metrics. This illustrates how strong oversharpening, which is often used boost summary metrics like MTF50, can seriously degrade performance. It can be identified by strong edge "halos" and SFR peaks (Figure 13).
- Sharpening + LPF slightly improves Edge SNRi for small *w*, and improves other metrics somewhat less than a pure LPF. However, both the sharpened + LPF images will be less susceptible to interference from neighboring objects.

The bottom line is that appropriate lowpass filtering as well as lowpass filtering with some sharpening can

potentially improve performance. Of course, these are just a few of many filter combinations of potential interest. And they are only for one specific camera at two ISO speeds (Exposure Indices 100 and 800).

And as we indicated, filters should be designed for more than a single task. The pure LPF performs well in the absence of interfering objects. Some sharpening may reduce the effects of interference, and as we have shown, sharpening, when combined with lowpass filtering, causes little performance degradation.

Matched filters

A matched filter [27] (sometimes called a "noise-whitened matched filter") is a custom filter that maximizes the SNR, i.e., the detection probability, for

- a system with a specific response, and
- a specific object (or edge).

Matched filters were originally developed for impulse detection in radar (a single airplane at a large distance). They are discussed in ICRU Report 54 [17], but were mostly ignored outside medical imaging because they were not directly relevant to human vision. That has changed with the advent of machine vision and artificial intelligence.

For an impulse (a δ-function, i.e., the airplane), the matched filter transfer function is identical to the noise-whitened system response (where "noise whitening" is division by $N_V(f)$, or equivalently, $\sqrt{NPS(f)}$).

$$\mathcal{F}_{matched}(f) = SFR(f)/(\sqrt{NPS(f)}) = \sqrt{K(f)} \quad (26)$$

A *matched filter*, $\mathcal{F}_{matched}(f)$, which optimizes *SNRi* or *Edge SNRi*, has the same frequency spectrum as the system, including the edge or object. For edge or object detection,

$$\mathcal{F}_{matched}(f) = \frac{|P(f)|\,SFR(f)}{\sqrt{NPS(f)}} = |P(f)|\sqrt{S(f)} \quad (27)$$

$P(f)$ is equal to $G(f)$ (Equation (19); the Fourier transform of the object) for *SNRi* or $H(f)$ (Equation (24); the Fourier transform of the edge) for *Edge SNRi*.

Figures 28 and 29 show the transfer functions for matched filters for *SNRi* or *Edge SNRi*. The Sharpening + LPF filter designed to approximate it (closest to *Edge SNRi*; Figure 30) has a response peak near 0.20 Cycles/Pixel.

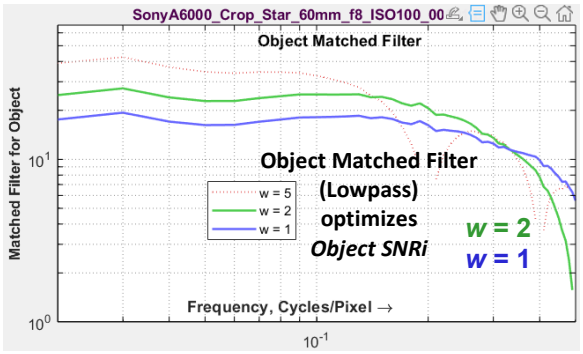


Figure 28. *Matched filter for optimum object: Lowpass (LPF).*

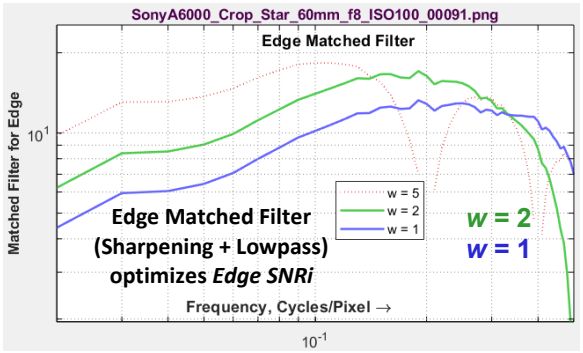


Figure 29. *Matched filter for optimum edge detection. Can be approximated with sharpening + LPF.*

Matched filters are optimized for a single task: detecting an object or edge of a certain size (by maximizing *SNRi* or *Edge SNRi*). But real-world filters must perform a multitude of tasks: they must detect objects and edges of varying sizes, contrasts, and colors. This calls for tradeoffs, which are quite mild. Since large objects are usually detected well, filters should be designed to perform well with small objects or edges.

Figure 30 shows the transfer function of a filter with a $\sigma = 0.7$ gaussian LPF + sharpening with Radius = 1 and Amount = 2 (R1A2G07).

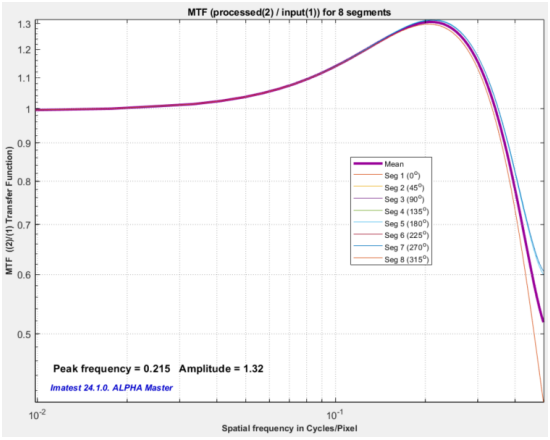


Figure 30. *Filter transfer function for σ = 0.7, R1A2. An approximate tradeoff between optimizing SNRi and Edge SNRi.*

This transfer function approximates the edge matched filter in Figure 29. The matched filter can be realized by combining a standard lowpass filter (Bessel, Butterworth, etc.) with standard sharpening or unsharp masking (USM). Details for realizing this filter are beyond the scope of this paper.

Exposure and pixel level

Most of the performance metrics discussed in this paper, including $NEQ(f)$, C_{NEQ}, *SNRi*, and *Edge SNRi*, are sensitive to exposure, increasing as exposure increases. For this reason, we need to maintain consistent exposure when acquiring images, especially when comparing different cameras. Linearized Digital Numbers (DNs or pixel levels) should be similar. A reasonable level will need to be established as we develop best practices for measurement.

Low light measurements are, of course, important, especially for video systems, where the maximum exposure time is limited, making it necessary to increase the analog gain (at the expense of SNR) at low light levels. It may be fruitful to measure C_4 (rather than C_{max}) as a function of exposure, especially at low light. We are still developing best practices for the new metrics.

Summary

The basic premise of this work is that traditional sharpness and noise metrics are insufficient to directly predict object and edge detection performance, and hence are poor predictors of Machine Vision/Artificial Intelligence system performance. And they provide no insight into how to design image processing for optimum performance. The new metrics and techniques described here are intended to accomplish this. They include

1. Two methods for measuring camera noise from slanted edges— spatially dependent noise, $\sigma_s(x)$, from the edge variance method, and the noise power spectrum, $NPS(f)$, from the noise image method.
2. Techniques for calculating camera information capacity from the average signal and noise power from both methods.
3. A set of metrics for quantifying object and edge detection performance, primarily derived from the noise image method, including $NPS(f)$, $NEQ(f)$, C_{NEQ}, *SNRi*, and *Edge SNRi*.
4. A technique for designing matched filters for optimizing detection performance, based on a tradeoff between maximizing *SNRi* and *Edge SNRi* and minimizing interference from nearby objects.
5. Note that the equations for the information metrics and matched filters all contain the equation kernel, $K(f) = SFR^2(f)/NPS(f)$. This unifies the results from the noise image method.

We need to verify that these calculations and design techniques work as intended— that they correlate well with MV/AI system performance.

The key concepts presented in this paper are

1. Information capacity, which combines sharpness, noise, contrast loss, is a fundamental figure merit for imaging systems that is appropriate for selecting cameras.
2. Both spatial and frequency-dependent noise can be measured from slanted-edge regions *at the same location, in the presence of the signal*. Co-locating signal and noise measurements makes the measurements convenient, robust, and reduces the likelihood of error.
3. A noise peak in $\sigma_s(x)$ allows bilateral-filtered images to be distinguished from uniformly-processed images for "black box" cameras with unknown image processing, so that the optimum noise calculation can be selected.
4. Information capacity, C_n, measured from n:1 contrast slanted edges (typically 4:1), is sensitive to chart contrast and exposure, but it can be extrapolated to calculate a stable maximum information capacity, C_{max}.

Camera information capacity and related information metrics are still novel in the imaging industry. Significant effort will be required to make them better known. But the units for C— information bits per pixel (or total image) for a specified ISO speed or exposure— are intuitive and easy to understand.

We would like to see information capacity become a standard specification for cameras intended for machine vision. And we would like to see better use made of edge and object detection metrics, *SNRi* and *Edge SNRi*. Cameras should be characterized with measurements made over a range of ISO speeds (exposure indices) and/or light (lux) levels. We are optimistic that this will lead to improved performance and reduced energy use [28].

Future work

- Collaborate with partners in academia and industry to correlate camera information capacity, C, and object and edge detection metrics, *SNRi* and *Edge SNRi*, with the performance of Machine Vision and Artificial Intelligence systems.
- Verify the validity of the new *Edge SNRi* metric, which is similar to *SNRi*, but has not has not been subjected to the same rigorous verification [17, 23]— correlating it with Bayesian statistics.
- Become familiar with the measured numbers for the metrics (e.g., what are "good" values of *SNRi* or

Edge SNRi?) Transform units from native sensor native units of cycles/pixel to practical units like cycles/angle or cycles per object distance, as needed.

- Determine best practices for measuring the information capacity of High Dynamic Range (HDR) sensors.
- Determine best practices for designing the output (matched) filter: How much weight should *SNRi* and *Edge SNRi* be given? What about sharpening to limit external interference?
- Study the effects of demosaicing, which may include nonlinear processing that enhances edges, on information capacity.
- Work on the ISO 23654 standard, Photography - Digital cameras - image Information Metrics, overseen by ISO TC42.
- Explore the correlation between *C* with the subjective visual appearance of images. This is challenging because visual perception is strongly affected by image processing (sharpening, color balance and saturation, tonal response, etc.).

References

[1] ISO/DIS 12233 Digital Cameras Resolution and spatial frequency responses, https://www.iso.org/standard/88626.html.

[2] C. E. Shannon, "A mathematical theory of communication," Bell Syst. Tech. J., vol. 27, pp. 379–423, July 1948; vol. 27, pp. 623–656, Oct. 1948.

[3] C. Shannon, "Communication in the Presence of Noise," Proceedings of the I.R.E., January 1949, pp. 10-21.

[4] R. Shaw, "The Application of Fourier Techniques and Information Theory to the Assessment of Photographic Image Quality," Photographic Science and Engineering, Vol. 6, No. 5, Sept.-Oct. 1962, pp. 281-286. Available for download from https://www.imatest.com/wp-content/uploads/2022/12/Rodney-Shaw_Fourier_Information_full.pdf.

[5] Robin Jenkin, Paul Kane, "Fundamental Imaging System Analysis for Autonomous Vehicles," Proc. IS&T Int'l. Symp. on Electronic Imaging: Autonomous Vehicles and Machines, 2018, pp. 105-1 – 105-10, https://doi.org/10.2352/ISSN.2470-1173.2018.17.AVM-105

[6] Frédéric Cao, Frédéric Guichard, Hervé Hornung, "Information capacity: a measure of potential image quality of a digital camera," DxO Labs, 2010.

[7] F.-X. Thomas, T. Corbier, Y. Li, E. Baudin, L. Chanas, F. Guichard, "RAW Image Quality Evaluation Using Information Capacity," Proc. IS&T Int'l. Symp. on Electronic Imaging: Image Quality and System Performance XVIII, 2021, pp. 218-1 - 218-7, https://doi.org/10.2352/ISSN.2470-1173.2021.9.IQSP-218

[8] J. C. Dainty and R. Shaw, "Image Science," Academic Press, 1974.

[9] Francis T. S. Yu, "Optics and Information Theory," Wiley, 1976.

[10] Tomasi, C., and R. Manduchi. "Bilateral Filtering for Gray and Color Images". Proceedings of the 1998 IEEE International Conference on Computer Vision. Bombay, India. Jan 1998, pp. 836–846.

[11] N. Koren, "Interpolated ISO 12233 slanted-edge SFR (MTF) calculation," www.imatest.com/2024/01/improved-slanted-edge-mtf-calculation/

[12] Peter Burns, Burns Digital Imaging, http://burnsdigitalimaging.com

[13] N. L. Koren, "Measuring camera information capacity with slanted-edges (Invited)" Electronic Imaging, 2023, pp 454--1 - 454-9, https://doi.org/10.2352/EI.2023.35.8.IQSP-454

[14] Orit Skorka et. al., "Evaluation of Signal and Noise Metrics of High Dynamic Range Image Sensors by IEEE P2020 Methodology," presented at Electronic Imaging 2024.

[15] Norman L. Koren, "Measuring camera Shannon Information Capacity with a Siemens Star Image", Proc. IS&T Int'l. Symp. on Electronic Imaging: Image Quality and System Performance XVII, 2020, pp 347-1 - 347-10, https://doi.org/10.2352/ISSN.2470-1173.2020.9.IQSP-347

[16] I.A. Cunningham and R. Shaw, "Signal-to-noise optimization of medical imaging systems", Vol. 16, No. 3/March 1999/pp 621-632/J. Opt. Soc. Am. A-optics Image Science and Vision

[17] ICRU Report 54, Medical Imaging – The Assessment of Image Quality, Bethesda, MD: International Commission on Radiation Units and Measurements, 1966.

[18] Brian W. Keelan, "Imaging Applications of Noise Equivalent Quanta," Proc. IS&T Int'l. Symp. on Electronic Imaging: Image Quality and System Performance XIII, 2016, https://doi.org/10.2352/ISSN.2470-1173.2016.13.IQSP-213.

[19] X. Tang, Y. Yang, S. Tang, "Characterization of imaging performance in differential phase contrast CT compared with the conventional CT: Spectrum of noise equivalent quanta NEQ(k)," Med Phys. 2012 Jul; 39(7): 4467–4482. https://doi.org/10.1118%2F1.4730287

[20] Michail C, et. al., Information Capacity of Positron Emission Tomography Scanners. *Crystals*. 2018; 8(12):459. https://doi.org/10.3390/cryst8120459.

[21] Christos M. Michail, et. al., "Figure of Image Quality and Information Capacity in Digital Mammography," *BioMed Research International*, vol. 2014, Article ID 634856, 11 pages, 2014. https://doi.org/10.1155/2014/634856.

[22] Orit Skorka, Paul J. Kane, "Object Detection Using an Ideal Observer Model", Proc. IS&T Int'l. Symp. on Electronic Imaging: Autonomous Vehicles and Machines, 2020, pp 41-1 - 41-7, https://doi.org/10.2352/ISSN.2470-1173.2020.16.AVM-041

[23] Paul J. Kane, "Signal detection theory and automotive imaging", Proc. IS&T Int'l. Symp. on Electronic Imaging: Autonomous Vehicles and Machines Conference, 2019, pp 27-1 - 27-8, https://doi.org/10.2352/ISSN.2470-1173.2019.15.AVM-027

[24] Chyuan-Tyng Wu et. al. (Intel Corp. Santa Clara), "VISIONISP: Repurposing the Image Signal Processor for Computer Vision Applications," IEEE International Conference on Image Processing (ICIP), 2019, pp. 4624-4628, https://doi.org/10.1109/ICIP.2019.8803607

[25] P. Hansen, et. al., ISP4ML: "Understanding the Role of Image Signal Processing in Efficient Deep Learning Vision Systems," 25th

International Conference on Pattern Recognition (ICPR), 2021, pp. 2438-2445, https://doi.org/10.48550/arXiv.1911.07954

[26] K. Park, M. Chae, and J.H. Cho, “Image Pre-Processing Method of Machine Learning for Edge Detection with Image Signal Processor Enhancement,” Micromachines 2021, 12, 73. https://doi.org/10.3390/mi12010073

[27] Nicholas R. Rypkema, “A Straightforward Derivation of the Matched Filter,” nrr.mit.edu/sites/default/files/documents/Matched_Filter.pdf

[28] Soumya Sudhakar, “Data centers on wheels: the carbon footprint of self-driving cars,” TEDxBoston, November 2022.

Author Biography

Norman Koren became interested in photography while growing up near the George Eastman House photographic museum in Rochester, NY. He received his BA in physics from Brown University (1965) and his Masters in physics from Wayne State University (1969), then worked in the computer storage industry simulating digital magnetic recording systems and channels for disk and tape drives from 1967-2001. He founded Imatest LLC in 2003 to develop software and test charts to measure the quality of digital imaging systems.