New Slanted-Edge Image Quality Measurements: the Edge variance method for Camera Information Capacity

Key Performance Indicators for Machine Vision and Artificial Intelligence Systems

Original publication November 7, 2022

Major revision March 12, 2023

minor revision March 12, 2022

Norman Koren, Imatest LLC

www.imatest.com

https://www.imatest.com/wp-content/uploads/2022/11/Information_capacity_slanted_edge.pdf

We present several new image quality measurements that can be made from the familiar slanted-edge test pattern, and can be applied to a archives of existing test images.

This document (Part 1) describes the Edge variance method of measuring camera information capacity, which combines sharpness and noise.

Part 2 describes the Noise image method of measuring Power Spectral Density (PSD), Noise Equivalent Quanta (NEQ), Information Capacity, and Ideal Observer Signal-to-Noise Ratio (SNRI), which measures detectability of small objects.

The measurements take advantage of two newly-discovered capabilities of slanted-edges. They have great potential as a fundamental figure of merit for image quality, particularly for Machine Vision and Artificial Intelligence.

Abstract— As Machine Vision (MV) and Artificial Intelligence (AI) systems are incorporated in an ever-increasing range of imaging applications— from medical images to autonomous vehicles— there is a corresponding need for camera measurements that can accurately predict the performance of these systems. Until now, the standard practice was to separately measure the two major factors— sharpness and noise (or Signal-to-Noise Ratio) as well as several additional factors, including tonal and color response, optical distortion, sensor linearity, and dynamic range, then to estimate system performance based on a combination of these factors. This estimate is usually based on experience, and is often more of an art than a science.

Camera information capacity, based on Claude Shannon's groundbreaking work on information theory [1],[2] holds great promise as a figure of merit for a variety of imaging systems, especially for Machine Vision and Artificial Intelligence that operate on *information* rather than *pixels*, but it is unfamiliar and has traditionally been difficult to measure. Appendix 1 contains an introduction to information theory.

In 2019 we developed a method for measuring it using the Siemens star test chart [3] that gave good results, especially when the effects of image processing such as demosaicing, compression, and bilateral (nonuniform) filtering were of concern, but was slower and less convenient than optimum.

In this white paper, We describe the new Edge variance method for measuring camera information capacity, C, from the well-known slanted-edge test pattern, specified by the ISO standard, ISO 12233:2017 Photography — Electronic still picture imaging — Resolution and spatial frequency responses [4]. Measuring information capacity requires no extra effort: it essentially comes for free along with slanted-edge MTF measurements. C has units of bits per pixel or total bits per image for a specified ISO speed and chart contrast, making it easy to compare very different cameras. The new measurement can be used to solve some important problems, such as finding a camera that meets information capacity requirements with a minimum number of pixels— important because fewer pixels mean faster processing as well as lower cost.

A follow-up white paper will describe the newer Noise image method for calculating the Noise Power Spectrum (NPS), Noise Equivalent Quantum (NEQ), Ideal observer SNR (SNRI), in addition to information capacity. Until it is available, the method is described on the web page, New measurements from Slanted-edges: Information capacity, NPS, NEQ, & SNRI.

The Slanted-edge measurement

The slanted-edge information capacity measurement uses an overlooked capability of slanted-edge regions that was quite literally hidden in plain sight. To understand it, we present a brief summary of the standard ISO 12233 Edge SFR (e-SFR) algorithm. For a full (but dense) explanation, see Appendix D of the ISO standard.


*Figure **1.** Slanted edge*

1. **The image should be well-exposed**, avoiding the dark "toe" and light "shoulder" regions, where the image deviates from standard log-linear behavior, $\log(pixel\ level) = \gamma_{encoding} \times \log(exposure)$
2. **Linearize the image** by applying the inverse of the encoding gamma curve or using the edge itself to obtain an approximate linearization if the chart contrast is known.
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.** In ISO 12233:2017+, it can be up to 6th order.
5. Depending on the location of the curve on the scan line, add each appropriately shifted scan line to one of four bins.
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 2.
$$\mu_s(\mathrm{x}) = \frac{1}{L}\sum_{l=0}^{L-1} y_l(x-\delta)$$
7. Note that we are effectively performing **signal averaging**, where the signal-to-noise ratio improves by the square root of the number of samples
8. **Modulation Transfer Function $MTF(f)$** is calculated by differentiating the averaged edge, windowing it, then taking the magnitude of the Fourier transform, normalized to 1 (100%) at zero frequency. $MTF(f)$ is illustrated in the lower plots of Figure 2, below.

The overlooked capability of the ISO 12233 binning algorithm: the Edge variance method

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

In addition to the sum, $\sum y_l(x)$, **Calculate the <u>sum of the squares of each scan line</u>**, $\sum y_l^2(x)$.

This allows **variance** σ_s^2 of the edge be calculated from the two arrays,

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

$\sigma_s^2(x)$ and $\sigma_s(x)$ are the noise power and noise voltage ($N(x)$ and $\sqrt{N(x)}$), *not the noise itself*, at each position on the oversampled array — including the edge transition, where noise was traditionally difficult to measure. Noise voltage $\sigma_s^2(x)$ is shown in Figures 3, 5, and 11.

Slanted-edge results for the same image saved as raw (converted to TIFF) and JPEG

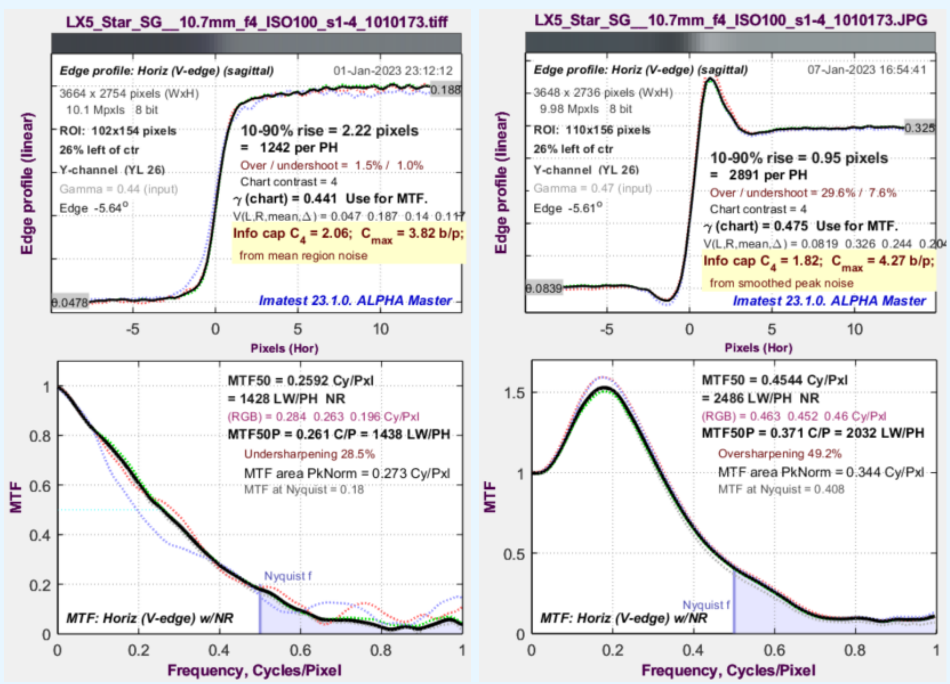


Figure 2. *Edge and MTF plots for compact digital camera, ISO 100, Y (Luminance) channel*
Upper plot: Mean edge μ_s Lower plot: MTF(f).
Left: Unsharpened TIFF from RAW Right: Strongly-sharpened in-camera JPEG

The 4:1 edge information capacity C_4 for the JPEG image (above-right) is sensitive to the region used to calculate noise, as discussed in **Noise Power (*N*)**. The minimally-processed TIFF on the left calculates *N* from mean of the edge. The bilateral-filtered JPEG on the right uses the peak of the smoothed noise near the edge. C_4 is slightly lower for the JPEG, but C_{max} is significantly higher, most likely because the bilateral filtering adversely affects noise measurements N_1 and N_2 on either side of the edge, and hence the N_{mean} calculation, described in the section, **calculating C_{max}**.

$\sigma_s^2(x)$ and $\sigma_s(x)$ must be corrected for binning noise — a recently-discovered artifact of the ISO 12233 binning algorithm, described below.

Noise N measured in the presence of a signal is more accurate than noise measured in flat areas, and can be entered into the Shannon-Hartley equation for channel capacity C.

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

where S and N are the average signal and noise power. C is strongly dependent on test chart contrast ratio and exposure, both of which affect signal S. Hence it is not a reliable indicator of a camera's full information capacity, especially for low contrast charts (≤10:1), which are recommended for the most accurate MTF measurements. After measurement details are presented, we will derive a more stable metric for the camera's full capacity, C_{max}.

An advantage of this method is that it allows noise to be calculated at *every* point where the signal is present — especially near the edge transition, where it has the greatest impact on system performance — not just in the flat, uniform areas.

Binning noise

Binning noise, which has identical statistics to quantization noise, is a recently-discovered artifact of the ISO 12233 binning algorithm. It is largest near the image transition — where the Line Spread Function $LSF(x) = d\mu_s(x)/dx$ (Figure 4) is maximum, and it can affect information capacity measurements. It appears because the individual scan lines are added to one of four bins, based on a polynomial fit to the center locations of the scan lines, which is a continuous function.

Assume that n identical signals $\mu_s(x)$ are binned over an interval $\{-\Delta/2, \Delta/2\}$, where $\Delta = 1$ in the 4× oversampled output of the binning algorithm (noting that Δ = (original pixel spacing)/4). If there were no binning noise, we would expect the binning noise power σ_{Bnoise}^2 to be zero. However, the values of $\mu_s(x_k)$ are summed at uniformly-distributed locations x_k over the interval Δ, so they take on values $\mu_k = \mu_s(x_k) = \mu_s(x_0 + \delta) = \mu_s(x_0) + \delta \frac{d\mu(x)}{dx} = \mu_s(x_0) + \delta\, LSF(x)$ for Line Spread Function LSF (Figure 4). Noting that δ is uniformly distributed over $\{-1/2, 1/2\}$ we apply the equation for the variance of a uniform distribution (similar to quantization noise) to get

$$\sigma_{Bnoise}^2(x) = LSF^2(x)\sigma_{Uniform}^2 = LSF^2(x)/12 \quad \text{or} \quad \sigma_{Bnoise} = LSF(x)/\sqrt{12}.$$

Although this equation involves some approximations, we have had good success calculating the corrected noise, $\sigma_s^2(\text{corrected}) = \sigma_s^2 - \sigma_{Bnoise}^2$. Binning noise has no effect on conventional MTF calculations.

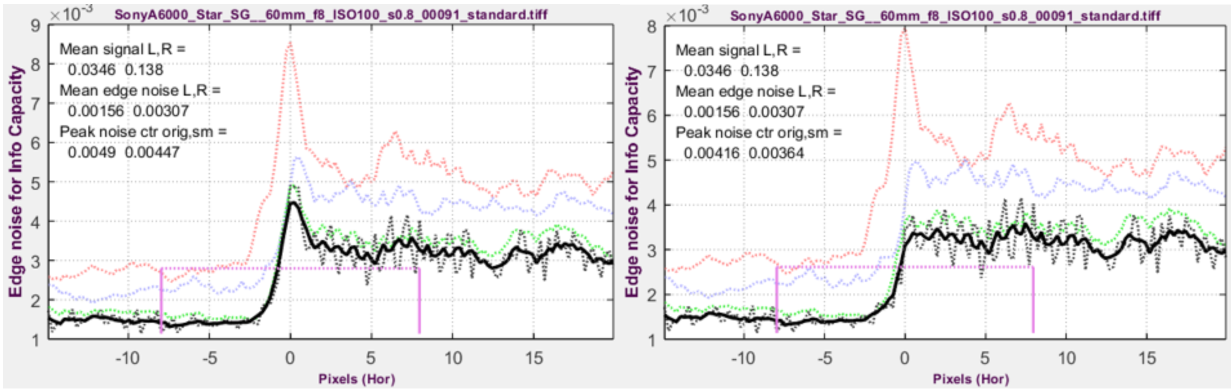


Figure 3. *Edge noise for a Micro Four-Thirds digital camera, ISO 100, Y (Luminance) channel from raw image converted to TIFF with minimal processing.*
Left: with binning noise (σ_s^2(uncorrected)); Right: with binning noise removed (σ_s^2(corrected))

Binning noise also affects JPEG files with bilateral filtering (nonuniform sharpening). Removing it is important for robust calculations.

Signal power S

The peak-to-peak signal voltage at low spatial frequencies is the measured difference between the means of the light and dark regions of the linearized slanted edge $\mu_s(x)$.

$$V_{p-p} = \Delta\mu_s = \mu_{sLight} - \mu_{sDark}$$

The signal power is the *variance* of this signal. If we assume a uniform distribution between the limits μ_{sLight} and μ_{sDark}, which maximizes information capacity, we note that the variance of the uniform distribution , which is the average signal power at low spatial frequencies, is

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

The Shannon-Hartley equation uses the *average* frequency-dependent signal power, $S(f)$.

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

Signal power S is proportional to the square of the chart contrast if the image has been properly linearized. $S_{max} \leq 1$ 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.

Noise N near the edge transition dominates system performance — not noise measured in flat patches. The calculation of N depends on the detected image type. Two distinct image types cover most cases of interest.

(1) **Uniformly or minimally-processed images**, often TIFFs converted from raw files (raw→TIFF), that don't have bilateral filtering, i.e., they either have no or uniform sharpening or noise reduction. Most cameras intended for Machine Vision/Artificial Intelligence fall into this category.

Since noise can be a very rough function of *x* (Figures 5 and 6), a large region size is required for a stable value of N. We average over all values of *x* in the ROI. (Formerly, we averaged noise over a region defined as the edge center ± 1.5×*PW20*, where *PW20* is the width of the region where the Line Spread function $LSF \geq 0.20\ LSF_{max}$.)

$$N_{method\_1} = \text{mean}(\sigma_s^2(x)) \quad \text{for all values of } x \text{ in the ROI.}$$

(2) **Bilateral-filtered images** include most JPEG images from consumer cameras.*

*Bilateral filtering and JPEG data compression are completely independent, though they're often found together in consumer cameras.

Bilateral filters sharpen images near contrasty features such as edges, but blur them (to reduce noise) elsewhere. This causes a noise peak near the edge. The blurring improves Signal-to-Noise Ratio (SNR), but it removes information. Because of this, noise near the edge can dominate camera performance, and should be strongly weighted in calculating N.

We have long known about the noise peak, but until the present method was developed, there was no good way to measure it (or detect bilateral filtering).

For calculating information capacity C, we use the noise at the peak, smoothed slightly (with a rectangular kernel of length *PW*20/2) to remove jaggedness. This is a somewhat arbitrary choice that produces reasonably consistent results. This method also works with minimally-processed images, but results are less consistent than method (1).

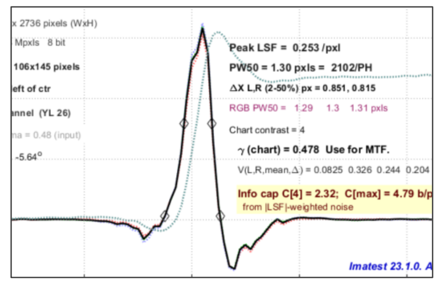


***Figure 4**. Line Spread Function (LSF) for a sharpened JPEG.*

The Imatest program has options for manually or automatically selecting the noise calculation method based on whether or not a peak is detected near the transition. The automatic detection is not 100% reliable (though if it fails, information capacity is only slightly changed). Some additional considerations:

- Noise is not exactly white, but these approximations are close enough to yield good results. This assumption is strongly supported by experimental results in [13].
- Noise power is larger on the lighter side of the edge because of photon shot noise, which increases with the number of photons reaching the sensor pixels. The mean includes both sides.
- More generally, noise power N increases with exposure. For linear sensors, $N(V) = k_0 + k_1 V$, where k_1 is the coefficient for photon shot noise.

Figure 5 illustrates the noise voltage $\sigma_s(x) = \sqrt{N(x)}$ for minimally-processed TIFF files (on the right) and for bilateral-filtered JPEG files (on the left) for a compact camera. The JPEGs have a large, distinct peak that is absent for the TIFFs. The solid dark curves are for the luminance (Y) channel smoothed with a 5 pixel-wide rectangular function (1.25 pixels before 4× oversampling) to improve plot appearance.

Bilateral-filtered (typical of in-camera JPEGs) **Minimally or uniformly-processed**

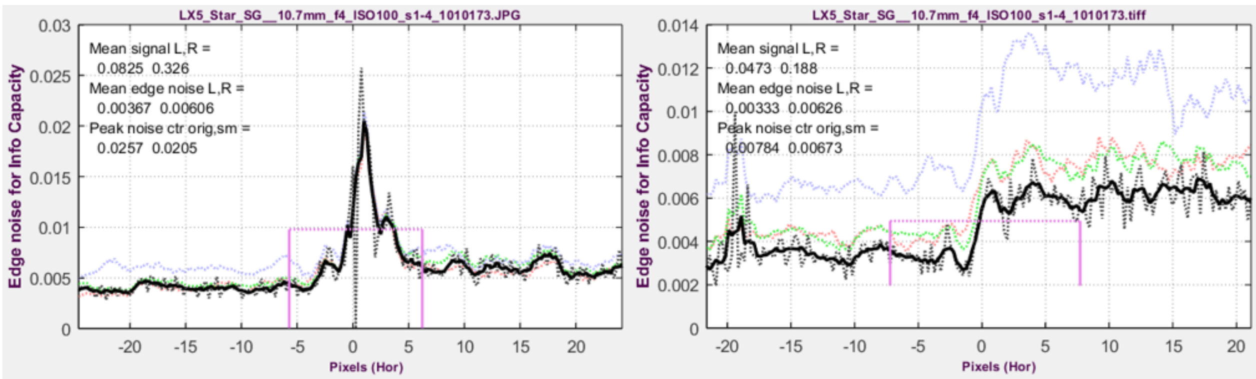


Figure 5. *Edge noise voltage @ ISO 100. Left: Bilateral-filtered in-camera JPEG; Right Unsharpened TIFF from raw.*
The x-axis is the original pixel location of the 4× oversampled signal.
Note that the spike around x = -19 of the plot on the right is a noise outlier likely caused by a speck of dust on the chart.
*It looks worse in the unsmoothed plot (dotted curve). The smoothing (**bold** curve) controlled it well.*

Bandwidth W

Bandwidth W is always 0.5 cycles/pixel (the Nyquist frequency). Signals above Nyquist do not contribute to the information content; they can actually reduce it by causing aliasing — spurious low frequency signals like Moiré that can interfere with the true image. Frequency-dependence comes from $MTF(f)$. Using the uniformly-distributed assumption,

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

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

$S_{avg}(f)$, N, and W are entered into the Shannon-Hartley equation.

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

$\mathrm{MTF}(f)$ can take a large bite out of C, especially since it is squared in the above equation. Because of its frequency-dependence, it is sometimes confused with bandwidth. For the raw-converted image in Figure 2, lower-left, it drops to zero around 0.6 cycles/pixel — typical of a well-focused high quality camera with no sharpening that makes good use of the sensor pixels. Since it is a nearly straight line, $MTF(f) \cong 1 - f/0.6$ for $f \leq 0.6$. The integral of $MTF^2(f)$ for $0 \leq f \leq 0.5$ is approximately 0.204: a significant loss from the value of 0.5 for a perfect (no rolloff) response.

Would increasing MTF help? The relationship between MTF and signal spread (or extent) is explored for diffraction-limited systems in [5] and summarized in the ***Imatest*** web page, Diffraction, Optimum Aperture, and Defocus. If all the energy of a point of light were inside one pixel, there would be no MTF loss. This corresponds to the case of Q = 0.5, where MTF is 0.69 at the Nyquist frequency (0.5 C/P), dropping to 0.4 at twice Nyquist (1 C/P). Such a system would have extreme aliasing (low frequency) artifacts that would degrade its performance. The camera in Figure 2, lower-left has $Q \approx 1.8$, with only a little energy above Nyquist, so aliasing is reasonably well-controlled. But the pulse is spread over two pixels, with the 10-90% rise distance = 2.22 pixels in Figure 2, upper-left, leading to a significant loss. This appears to be an unavoidable tradeoff.

Technique

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

Images should be well-exposed because saturation or clipping (both deviations from log-linearity) can give misleading results indicating better than real performance.

The camera should be well-focused (unless you're testing misfocus), and 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; we recommend at least 30x60 pixels. The edge should be centered in the selected region, and there should a reasonable amount of "breathing room" on the sides. A good initial "rule of thumb": The ratio of light to dark space at the ends of the edges (the top and bottom of Figure 1) should be no larger than 35/65 (or 65/35, depending on the orientation).

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 likely to occur in pictorial, medical, or robotic imaging have tonal ranges under 60 dB. Lens flare (stray light) typically limits practical camera dynamic range to 100dB or less. If there are questions or doubt about the camera's dynamic range, we strongly recommend measuring it.

The one class of images where scene contrast is likely to exceed 60 dB is night driving, where the scenes can include both light sources and shadows that contain objects of importance. A "worst case" scene is a forward view from a car emerging from a dark tunnel into daylight, where it's important to see the road ahead inside and outside the tunnel. The key concern is that flare light (or veiling glare) from the light sources can fog the important middle to dark tones.

Sensor nonuniformities (fixed-pattern noise, also called PRNU (Photo Response Nonuniformity) are included in noise measurements. Tonal response should be well-behaved (typically following a gamma curve, except possibly in the extreme highlights and shadows).

Because C is a strong function of the test chart contrast ratio, we include the contrast when reporting C, i.e., C_n for *n*:1 contrast ratio. C_4, for the ISO 12233-standard 4:1 contrast ratio, is used throughout this document.

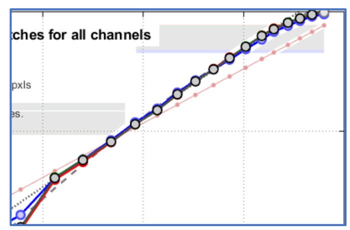
Sensitivity to exposure

Because the old ISO 12233:2000 standard (replaced in 2014) used high contrast edges (≥50:1), 4:1 edges may appear to have relatively low contrast. Nevertheless, they can occupy a substantial portion of the available linearized/normalized signal voltage *V*, where $0 \le V \le 1$. The portion is strongly dependent on exposure. For a standard "good" exposure, where $V_{mean} \approx 0.20$, the voltage from the 4:1 edge occupies 24% of the total range. But it can occupy as much as 75% for a 1.64 f-stop overexposed image.

V_{mean}	V_{min} (0.4 V_{mean})	V_{max} (1.6 V_{mean})	Range = ΔV = V_{max}-V_{min} = $V_{p\text{-}p}$
0.12	0.048	0.192	0.144
0.20	0.08	0.32	0.24
0.40	0.16	0.64	0.48
0.60	0.24	0.96	0.72

***Table 1**: V_{mean}, V_{min}, V_{max}, and range of normalized signal voltage $\Delta V = V_{max}$-$V_{min} = V_{p\text{-}p}$ for 4:1 contrast ratio edges*

Because both noise power *N* and voltage range ΔV increase with exposure, **C_4 is a strong function of exposure.**



Consistent exposure can be difficult to maintain 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, often appear to be underexposed.

Because C_4 is a strong function of exposure and consistent exposure can be difficult to obtain, we have developed a new metric, ***C_{max}***, described below, that is relatively insensitive to exposure.

Visualizing 4:1 contrast — Because our eyes have a logarithmic response to light, 4:1 contrast edges appear to have a constant contrast regardless of their brightness (as long as they're not saturated), even though ΔV in the above table varies widely.

4:1 contrast ratio slanted edges are preferred for measurements because

- They are specified by the ISO 12233 standard since 2014. There is a huge archive of such images.
- The moderate contrast level is characteristic of features (edges, etc.) in typical images.
- Higher contrast edges can cause sensors to operate in nonlinear regions, often causing clipping. This reduces accuracy.

Maximum information capacity C_{max} — a more consistent metric

The strong dependence of C_4 on exposure reduces its value as a performance metric. The reasons for this dependence are (1) voltage range $\Delta V = V_{p\text{-}p}$ is a strong function of exposure, and (2) noise power N is also a function of exposure.

We have developed a new 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 voltage range $V_{p\text{-}p}$ with the maximum allowable value, $V_{p-p\_max} = 1$. This may seem like a simplification, but it works well for most cameras. Referring to the section on Signal Power S,

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

If the camera has a pixel offset (a minimum allowable pixel level) or a maximum pixel level, $V_{p\text{-}p\_max}$ may need to be set lower than 1.

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 voltage V). The general equation for noise power N as a function of V for linear image sensors is

$$N(V) = k_0 + k_1 V$$

k_0 is the coefficient for constant noise (dark current noise, Johnson (electronic) noise, etc.). k_1 is the coefficient for photon shot noise. Noise powers $N_1 = \sigma_1^2$ and $N_2 = \sigma_2^2$ are measured along with signal voltages V_1 and V_2 on either side of the edge transition.

Assuming $N_1 = k_0 + k_1 V_1$ and $N_2 = k_0 + k_1 V_2$ we can use two equations in two unknowns to solve for k_0 and k_1.

$$k_0 = \frac{N_1 V_2 - N_2 V_1}{V_2 - V_1}\ ; \quad k_1 = \frac{N_2 - N_1}{V_2 - V_1}$$

N closely approximates the noise used in noise calculation method (1) (used for minimally-processed images that don't have bilateral filtering). But if method (2) (the smoothed peak noise) is used (recommended for in-camera JPEGs with bilateral filtering), N is generally larger, and must be modified.

$$N \rightarrow k_N N, \text{ where } k_N = N_{method\_2}/N_{method\_1}$$

In bilateral-filtered images (most JPEGs from consumer cameras), lowpass filtering (for noise reduction) may be affect N_1 and N_2 strongly enough so the equation $N(V) = k_0 + k_1 V$ does not reliably hold. This can adversely affect the accuracy of C_{max}.

The mean noise power N_{mean} over the range $0 \le V \le 1$ for calculating C_{max} is

$$N_{mean} = \int_0^1 N(V)\, dv \Big/ \int_0^1 dv = \int_0^1 (k_0 + k_1 V) dv = k_0 + k_1/2$$

Using $N = N_{mean}$, $V_{p-p\_max} = 1$ and $S_{avg}(f) = MTF(f)^2/12$,

$$C_{max} = \int_0^{0.5} \log_2\left(1 + \frac{MTF(f)^2}{12\ N_{mean}}\right) \mathrm{df} \cong \sum_{i=0}^{0.5/\Delta f} \log_2\left(1 + \frac{MTF(i\Delta f)^2}{12\ N_{mean}}\right)\Delta f$$

C_{max} (Figure 6) is nearly independent of exposure for minimally or uniformly-processed images with linear sensors, where noise power *N* is a known function of signal voltage *V*.

It is approximate for other imaging systems (with bilateral-filtered in-camera JPEGs, HDR sensors, etc.) where noise power N is not a simple function of *V*.

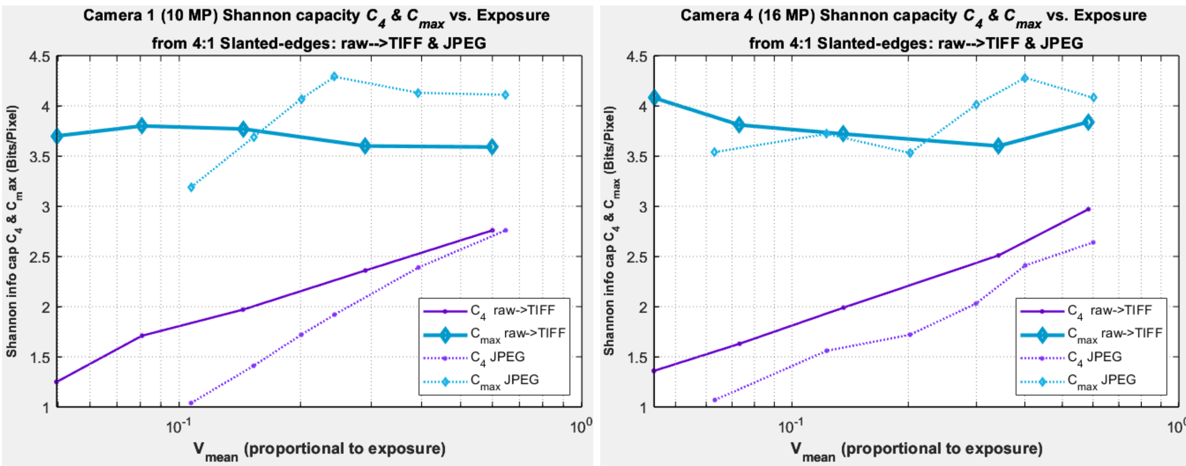
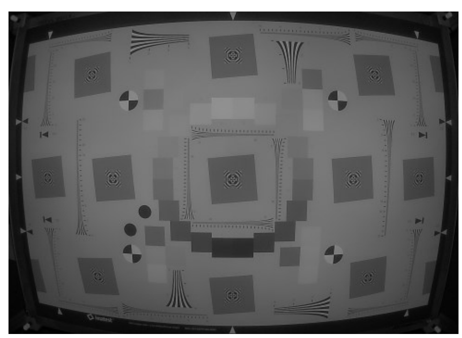


Figure 6. *C_4 and C_{max} for minimally processed raw→TIFF and JPEG images for two cameras:*
1. Panasonic Lumix LX5 (10 MP compact), 2. Panasonic Lumix G3 (16 MP micro 4/3).
The left edge of the central square was used for the analysis.
C_{max} results are consistent for raw→TIFF images, except for the darkest (severely underexposed).

Potential problems with C_{max}



The C_{max} calculation assumes that the image is capable of spanning the entire range of Digital Numbers (DNs), for example, 0-255 for images with bit depth = 8. This isn't always the case, especially with cameras under development where image processing isn't functioning well. We don't know the history of the image on the right — it may be a prototype near infrared (NIR) camera.

The pixel level in the brightest OECF patch is only 145 — far below the maximum DN for 8-bit files of 255, and gamma = 0.24 is exceptionally low (0.45 is nominal for sRGB color space). The gamma curve is well-behaved, showing few signs of veiling glare or saturation.

We may need to enter information about the maximum and perhaps minimum available DN to get a reasonable value of C_{max}. The dynamic range should be measured, possibly using a [newly-developed chart that works in the visible and NIR regions].

Tone-mapped images, which can be pictorially pleasing with High Dynamic Range (HDR) scenes, but ruin measurements, **cannot be used for C_{max}**.

Information capacity results

We tested three cameras that produced both raw and JPEG output for information capacity C 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.** An older (2010) compact 10.1-megapixel camera with a Leica f/2 zoom lens set to f/4.
2.	Sony A6000	**3.88 µm pixel pitch.** A 24-megapixel micro four-thirds camera with a 60mm Canon macro lens set to f/8
3.	Sony A7Rii	**4.5 µm pixel pitch.** A 42-megapixel full-frame camera with a Backside-Illuminated (BSI) sensor and a 90mm f/2.8 Sony macro lens set to f/8

We captured both JPEG and raw images, which were converted to 24-bit sRGB (encoding gamma ≅ 1/2.2) TIFF images (designated as raw→TIFF) with LibRaw, with minimal processing (defined as no sharpening, no noise reduction, and a simple gamma-encoding function). Results with 48-bit Adobe sRGB conversion were nearly identical.

The image in Figure 7 below, which was analyzed in "Camera Information Capacity: A Key Performance Indicator for Machine Vision and Artificial Intelligence Systems", contains a 50:1 contrast Siemens star and four 4:1 contrast slanted edges on the sides. We used the upper-left slanted edge for most tests. The average background of the chart should be close to neutral gray (18% reflectance) to ensure a good exposure (exposure compensation may be applied if needed and available).

Figure 7. Typical image (cropped) including Siemens star and slanted-edges to the left and right of the star.

Results for JPEG and minimally-processed raw→TIFF images

Figures 8-10 show results for the luminance (Y) channel as a function of ISO speed (Exposure Index) for the 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. Noise calculation (1) is used for the raw→TIFF images; noise calculation (2) is used for the bilateral-filtered JPEGs.

Results are for the luminance (Y) channel, where Y = 0.2125·R + 0.7154·G +0.0721·B.

C_4 4:1 slanted edge

The information capacity for 4:1 contrast edges, C_4, shows similar trends to C_{max} and C_{star}, but since the relatively low 4:1 contrast uses only a fraction of the available signal level, C_4 is lower than either measurement.

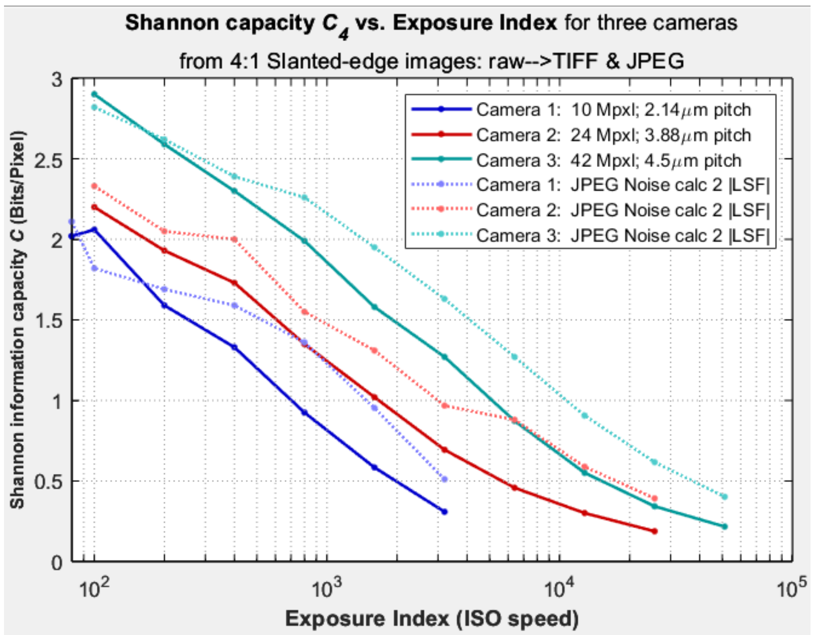


Figure 8. Information capacity C_4 from 4:1 slanted-edge images. *Solid lines for raw→TIFF images; Dotted lines for JPEGs.*

C_{max} maximum information capacity

C_{max} is the maximum information capacity of the camera, derived from measurements of 4:1 edges. It's relatively accurate for minimally or uniformly-processed (often raw→TIFF) images, and it's much less sensitive to exposure than C_4, making it a very robust measurement, well-suited for comparing the performance of different cameras.

Comparisons of information capacity between different cameras are similar regardless of the measurement method.

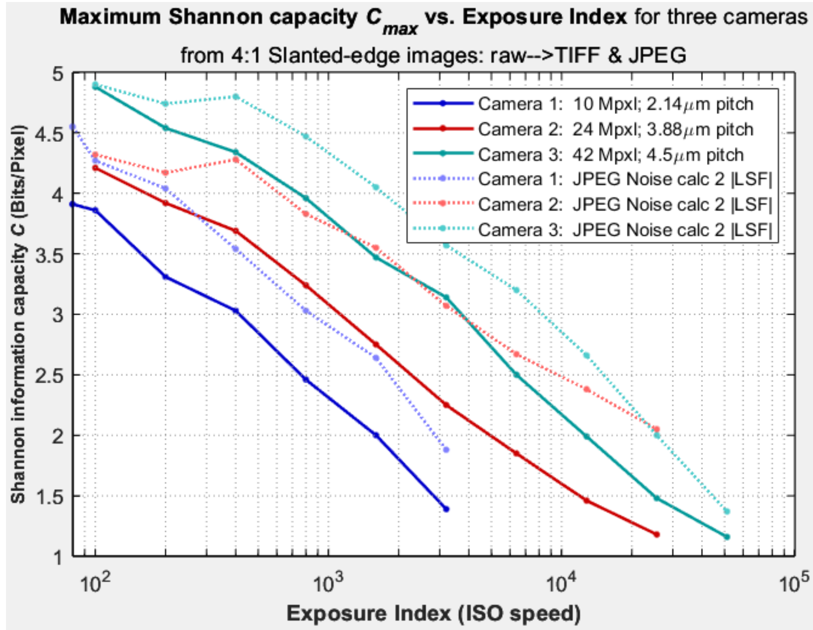


Figure 9. Information capacity C_{max} from slanted-edge results.
Solid lines for TIFFs derived from raw images; dotted lines for JPEGs.

C_{star} Siemens star (50:1 contrast) from the 2020 Information Capacity white paper

Information capacities C_{star} for the star are generally higher than C_4 but lower than C_{max} for slanted edges because the range of tones in the star is between the two, i.e., the star images don't use the entire available tonal range.

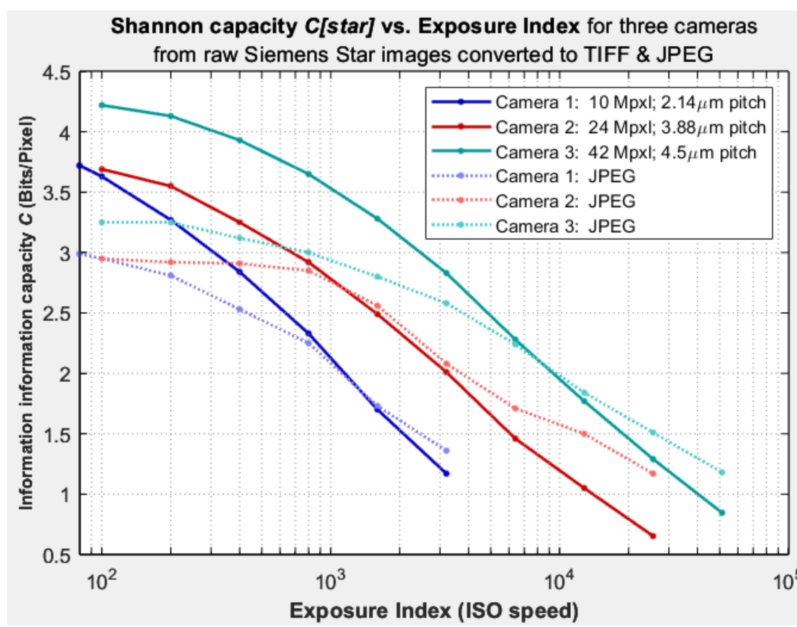


Figure 10. Information capacity from Siemens star images for the three cameras as a function of Exposure Index: Solid lines for TIFFs derived from raw images; dotted lines for JPEGs.

Color channels

The separate R, G, and B channels tend to have slightly lower C_4 than the Y-channel because the uncorrelated noise from the separate channels are combined in the Y-channel. Example: for Camera 2 (24 Megapixels, Micro Four-Thirds) at ISO 400, C_{4Y} = 1.96, C_{4R} = 1.16, C_{4G} = 1.81. and C_{4B} = 1.36 bits/pixel. The noise for each channel is shown on the Right: from best (lowest noise, highest C) to worst: Y, G, B, and R. The green channel has the best SNR because the image sensor is most sensitive to green, and hence the green channel had the least boost in the white balance process.

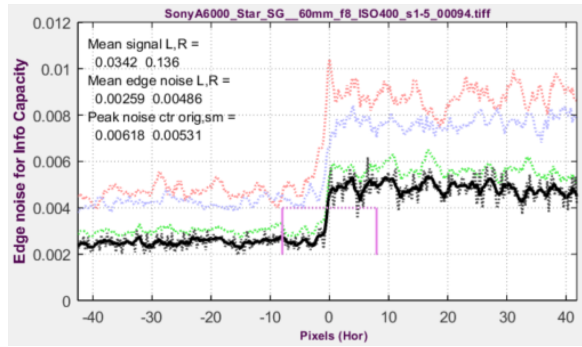


Figure 11. *Noise for camera 2 @ ISO 400, showing the different color channels*

$C_R + C_G + C_B$ is nearly triple C_Y. But this is to be expected because the three color channels are in 24-bit (RGB) pixels instead of 8-bit (single channel) pixels.

Even though we've focused on demosaiced images, the slanted-edge method can be applied to raw (undemosaiced) images. For this camera, C_{4Ru} = 2.09, C_{4GRu} = 2.43, C_{4Bu} = 1.6. and C_{4GBu} = 2.47 bits/pixel, where the undemosaiced (u) channels have a quarter as many pixels as the demosaiced channels. We haven't worked on interpreting these results.

Effects of sharpening

The two examples below show that sharpening has no significant effect on slanted-edge information capacity, as expected for a valid measurement. The two images (originally a minimally-processed TIFF) have been strongly USM sharpened in the Imatest Image Processing module with Radii = 1 and 2 and Amount = 2. They can be compared to Figure 2 (left), where $\boldsymbol{C_4}$ = 2.06 and $\boldsymbol{C_{max}}$ = 3.82 b/p. We observed a similar insensitivity of C to sharpening with Siemens stars.

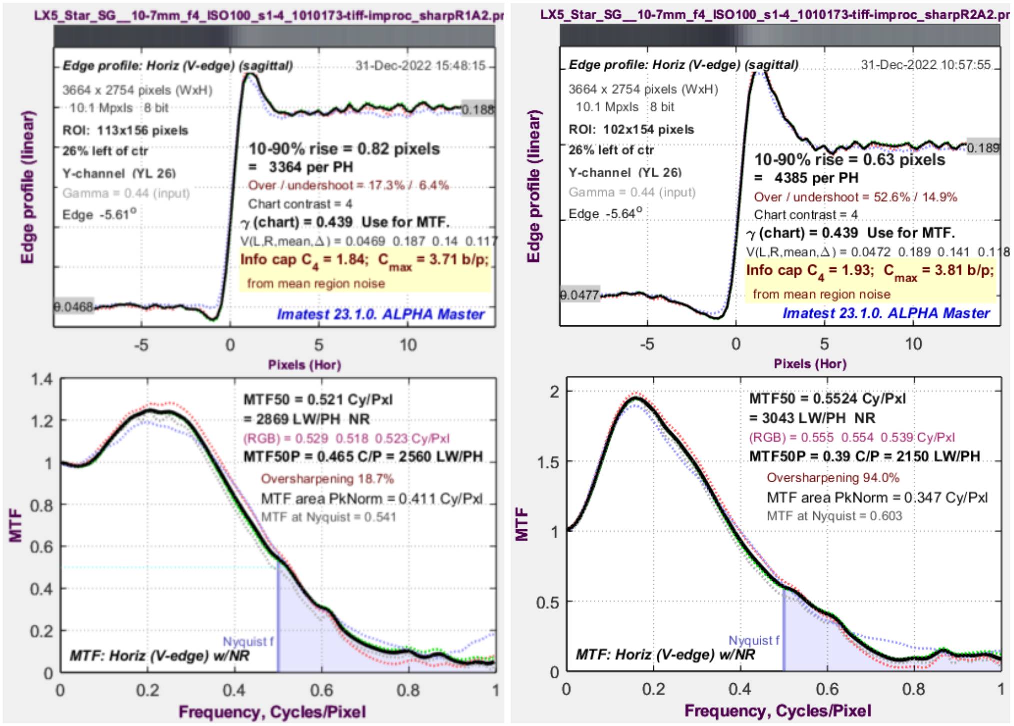


Figure 12. *Edge/MTF plots of same image as Figure 2 (left where C_4 = 2.35 b/p): TIFF from RAW, ISO 100*
Left: sharpening Radius = 1, Amount = 2. C_4 = 2.15 b/p. Right: sharpening Radius = 2; Amount = 2. C4 = 1.93 b/p.

This highlights another benefit of the information capacity measurement. It does not reward excessive sharpening, which creates "halos" near edges. These halos improve the perception of sharpness (by the human eye) when applied in moderation, but create artifacts that degrade image appearance when applied in excess [6]. They also have a bad reputation for machine vision applications.

Total information capacity

The results we have presented thus far are information capacity C in bits per pixel. The total information capacity C_{total} for the entire image must take variations in C over the image into account. In ***Imatest***, the mean value of C for the image is displayed in the 3D plots for auto-detected slanted-edge modules, SFRplus, eSFR ISO, and Checkerboard, which can map a variety of results, including C, over the image. For the information capacity plots ($\boldsymbol{C_4}$ and $\boldsymbol{C_{max}}$), the zone weights are always [1, 1, 1], and the total capacity is displayed.

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

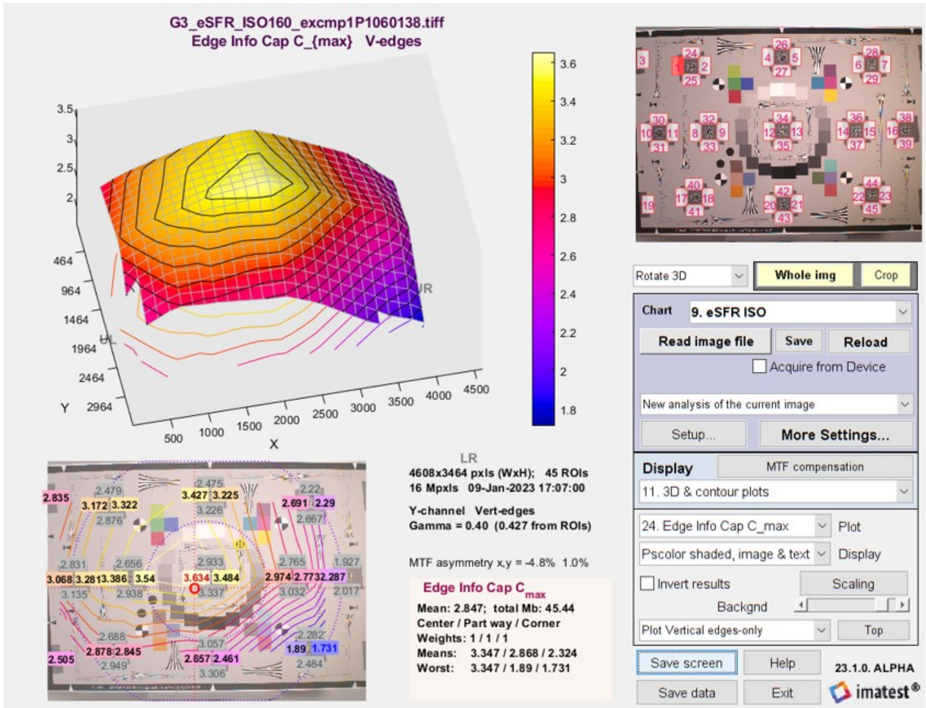


Figure 13. *3D contour **eSFR ISO pl**ot of C_{max} for the Luminance (Y) channel, ISO 100*

The mean information capacity C_{max} (unweighted for information capacity calculations) is 2.847 bits/pixel. Since the camera has 16 Megapixels, total capacity $C_{maxTotal}$ for the Luminance (Y) channel = 45.44 MB.

Signal Averaging

Extremely noisy images, typically acquired in dim light or at high ISO speeds (exposure indices), may have inaccurate MTF calculations, and hence information capacity measurements. A classic technique for reducing the effect of noise and obtaining reliable measurements from noisy images is *signal averaging*, where n identical captures of the same image are averaged. When this is done, the sum of the signal voltage and the sum of the noise *power* (noise voltage^2), which is uncorrelated, are both proportional to n. This causes noise voltage to be proportional to $\sqrt{n}$ and SNR to increase by $\sqrt{n}$ for n averages: by 3dB whenever n is doubled. To obtain correct information capacity measurements when the signal is averaged, the noise power is multiplied by n in the information capacity calculation.

This effect is illustrated below for a camera with a one-inch sensor, which wasn't perfectly focused, at ISO 12800. A single image is shown on the left. Note that MTF is rough and has significant high frequency bumps. For the average of 8 images is shown on the right, information capacity C is lower because MTF is better behaved.

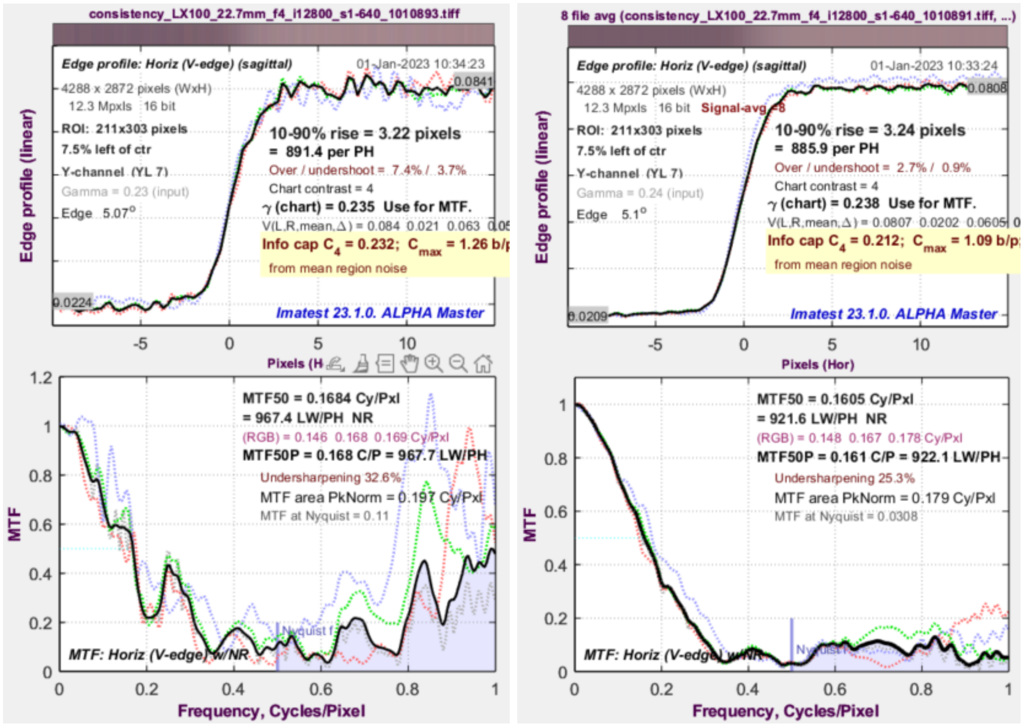


Figure 14. *Results without (left) and with (right) signal averaging (8×). 1 inch sensor @ ISO 12800. Note that Information capacity C_4 and C_{max} for the averaged images are lower because there is less spurious MTF caused by noise at high frequencies. (Unaveraged C is very inconsistent.)*

Comparisons of the slanted-edge and Siemens star methods

Slanted-edge methods

- A key advantage is that any image of a slanted-edge test chart with printed contrast ≤ 10:1 can be used to obtain C. The ISO 12233 2014+ standard contrast ratio of 4:1 is strongly recommended. Imatest and many of its customers have large archives of suitable images that can be run without modification.
- Works well with multi-region images: C can be easily mapped over the entire image (along with a great many other results).
- For bilateral-filtered images (most in-camera JPEGs): Noise is reduced in smooth areas, but the calculation for C uses the peak noise near the transition, which is not reduced. Results from the Edge variance method are useful, but results are more consistent and reliable for minimally or uniformly processed images.
- Does not measure the effects of artifacts very well. Clipped images may show improved performance due to sharp corners on the edge transition and reduced noise in the clipped areas. The Siemens star gives more accurate results for clipped images, but clipping should be avoided whenever accurate, repeatable measurements are required.

Siemens star method

- Generally excellent results, but sensitive to optical distortion. Best if the star is in the center of the image; may not work as well with multiple stars if distortion is present. Not good for mapping C over the image or finding the total information capacity.
- Slower than slanted edges.
- For bilateral-filtered images (most in-camera JPEGs): fine high-frequency detail is reduced (smoothed), but since the star calculation defines noise as the difference between the inferred original and acquired signal, noise is increased, appropriately reducing C.
- Appropriate response to image processing artifacts: enables comparison of demosaicing techniques, image compression, aliasing, etc. C is reduced appropriately for clipped images.

Both methods assume that the camera dynamic range is sufficient for the intended task. This is likely to hold for many MV/AI applications, but it may not hold for automotive night driving, where dynamic range may be compromised by stray light and veiling glare. If there are any questions about dynamic range, we recommend testing it using a transmissive chart. Imatest's new VIS/NIR charts, which are spectrally neutral and work in the visible and near infrared (NIR) ranges, are especially suitable.

When comparing cameras, the same measurement method (chart type, contrast, etc.) should be used.

Future work

- Work with partners in industry and academia to correlate information capacity C with performance of Machine Vision and Artificial Intelligence systems.
- Work to include camera information capacity in several standards, especially ISO TC42.
- Do more to correlate information capacity C with the subjective visual appearance of a variety of images, without and with additional image processing. We may extend the model of C to include viewing conditions and the human visual system to obtain a "visual information capacity", analogous to visual noise or acutance.
- Determine if we can predict the effects of image processing (sharpening, noise reduction) on MV/AI performance. Our initial approach will be to calculate SNRI, based on the work of Paul Kane and collaborators [5], [8], [9].

Summary

We have developed two methods for measuring camera information capacity, C, from slanted edges. This white paper has described the Edge variance method.

Information capacity has great potential value as a figure of merit for evaluating camera image quality because it combines the effects of sharpness, noise, and (for the Siemens star calculation) several types of artifact. The new slanted edge method are fast and convenient — requiring no special effort.

The key concepts we have presented are

1. Information capacity is a fundamental figure merit for imaging systems — more so than sharpness or any other metric.
2. The spatially-varying noise power $N(x)$ can be extracted from slanted-edge regions.

3. The noise peak in bilateral-filtered images allows them to be distinguished from uniformly-processed images, so that noise calculations can be selected for optimum results with each.
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 extended to calculate a maximum information capacity C_{max}, that is insensitive to these factors.
5. Reliable C_{max} measurements can be obtained from minimally or uniformly-processed (often raw→TIFF) images, but C_{max} is less accurate (but still useful) with bilaterally-filtered images because of the effects of nonuniform processing on noise measurements and possibly the effects of the sharpening peak in strongly-sharpened images.
6. The Noise image method can measure several image quality factors in addition to information capacity: Noise Power Spectrum (NPS), Noise Equivalent Quantum (NEQ), and Ideal observer SNR (SNRI), but it is not intended for use with bilateral-filtered images. It will eventually have its own white paper, but for now it's described in New measurements from Slanted-edges: Information capacity, NPS, NEQ, & SNRI.

The earlier Siemens star method [3] is effective for quantifying artifacts caused by demosaicing, data compression (as in JPEG images), and clipping, but it is somewhat slow, sensitive to optical distortion, and not well-suited for calculating total information capacity.

Both slanted-edge methods are fast, work with a vast archive of existing slanted-edge test chart images, require no changes in test or analysis procedure (apart from making a selection). They are very convenient for measuring total information capacity C_{total}.

The Siemens star and slanted-edge methods have similar *relative* trends (for comparing cameras).

Camera information capacity is still a novel measurement. Significant effort will be required to make it better known and give it traction in the industry. But despite its unfamiliarity, the units of camera information capacity— information bits per pixel (or total image) for a specified ISO speed— are intuitive and easy to understand.

We would like to see information capacity — either expressed as bits per pixel or megabits total at specified ISO speeds (exposure indices) or light (lux) levels — become an integral part of a standard camera specifications, particularly for machine vision applications. Using information capacity in this way will help in selecting cameras that have maximum performance with minimum pixel count, which will improve speed and energy consumption, which is not insignificant. See **Computers that power self-driving cars could be a huge driver of global carbon emissions.**

Appendix I. Information theory background

Because concepts of information theory are unfamiliar to most imaging engineers, we present a very brief introduction. To learn more, we recommend a text such as "Information Theory— A Tutorial Introduction" by James V Stone, available on Amazon. Shannon's 1948 and 1949 papers [1],[2] are highly readable.

What is information?

Information is a measure of surprise or the resolution of uncertainty. The classic example is a coin flip. For a "fair" coin, which has a probability of 0.5 for either a head or tail outcome (which we can designate

1 or 0), the result of such a flip contains one bit of information. Note that two coin flips have four possible outcomes (00, 01, 10,11); three coin flips have eight possible outcomes, etc. The number of information bits is $\log_2$(the number of outcomes), which is the number of flips.

Now, suppose you have a weirdly warped coin that has a probability of 0.99 for a head (1) and 0.01 for a tail (0). Little information is gained from the results of a flip. The equation for the information in a trial with m outcomes, where $p(x_i)$ is the probability of outcome i and $\sum_{i=1}^{m} p(x_i) = 1$, is

$$H = \sum_{i-1}^{m} p(x_i) \log_2 \frac{1}{p(x_i)}$$

H is called "entropy", and is often used interchangeably with "information". It has units of bits (binary digits). Note that this definition is subtly different from the physical memory element called a "bit".

For the fair coin, where $p(x_1) = p(x_2) = 0.5$, $H = 1$ bit. For the warped coin, where $p(x_1) = 0.99$ and $p(x_2) = 0.01$, $H = 0.0808$ bits. If the results of the warped coin toss were transmitted without coding, each channel bit would contain 0.0808 information bits. That would be extremely inefficient.

Claude Shannon was one of the genuine geniuses of the twentieth century— renowned among electronics engineers, but little known to the general public. The medium.com article, 10,000 Hours With Claude Shannon: How A Genius Thinks, Works, and Lives, is a great read. There are also nice articles in The New Yorker and Scientific American. The 29-minute video "Claude Shannon – Father of the Information Age" is of particular interest to the author of this white paper because it was produced by the UCSD Center for Memory and Recording Research, which he visited frequently in his previous career.

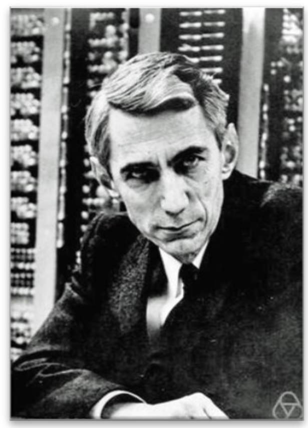


Figure 15. *Claude Shannon*

Channel capacity

Shannon and his colleagues developed two theorems that form the basis of information theory.

The first, Shannon's source coding theorem, states that for any message there exists an encoding of symbols such that each channel input of D binary digits can convey, on average, close to D bits of information. For the above example, it implies that a code can be devised that can convey close to 1 information bit for each channel bit—a huge improvement over the uncoded value of 0.0808.

The second, known as the Shannon-Hartley theorem, states that the channel capacity C, i.e., the theoretical upper bound on the information rate of data that can be communicated at an arbitrarily low error rate through an analog communication channel with bandwidth W, average received signal power S, and additive white Gaussian noise (AWGN) of power N is

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

This equation is challenging twaterffo use because bandwidth W is not well-defined, noise is not white, and it applies to one-dimensional systems, whereas imaging systems have *two* dimensions, at least for Siemens stars. Slanted-edge analysis is one-dimensional.

At this point we can hazard a guess as to why camera information capacity has been ignored for cameras. For most of its history the hot topic in information theory was the development of efficient codes, which didn't approach the Shannon limit until the 1990s—nearly fifty years after Shannon's original publication. But channel coding is not a part of image capture (though it's used downstream for image and video compression). Also, camera information capacity was not critically important when the primary consumers of digital images were humans (though it is related to perceived image quality), but that is changing rapidly with the development of new AI and machine vision systems. And finally, convenient methods of measuring it didn't exist.

Appendix 2. History

A key early paper that deals with Shannon information capacity is R. Shaw, "The Application of Fourier Techniques and Information Theory to the Assessment of Photographic Image Quality," Now available for download, and highly recommended. [7] This article reviews work on information capacity between Shannon's original 1948 publication and 1962, then uses the Shannon-Hartley equation to calculate information capacity in bits/cm^2 for three types of film. This was brilliant work, but difficult to perform, requiring exacting technique for writing patterns on the film, then analyzing them with a microdensito-meter. For comparison with Fig. 6, camera 2 (3.88µm = 0.000388cm pixel pitch) has 2.9 bits/pixel @ ISO 400 (the speed of HP3 in 1960) = 19.26 Megabits/cm^2 — about three orders of magnitude better than HP3.

R. Jenkin and P. Kane, Fundamental Imaging System Analysis for Autonomous Vehicles, Electronic Imaging, 2018. https://www.researchgate.net/publication/325622173 [8] discusses information capacity, focusing on modeling.

DXOMARK has published several papers on information capacity.

"Sensor information capacity and spectral sensitivities" (2009) [9] introduces CS_{RGB}, the number of different colors a sensor can distinguish, based on the eigenvalues of the noise covariance matrix. It contains little about spatial resolution.

Information capacity: a measure of potential image quality of a digital camera (2010) [10] separates information capacity into several factors: CS_{RGB}, blur, chromatic aberrations (lateral and longitudinal), light shading, and distortions. As outlined in section 6, these factors are measured separately, then combined to obtain information capacity, which is much higher than slanted-edge measurements for comparable cameras. Several of the factors (light shading, lateral chromatic aberration, and distortion) can be corrected in the image processing pipeline, and the resulting information loss is reflected in slanted-edge measurements.

RAW Image Quality Evaluation Using Information Capacity (2021) [11] continues the use of raw images for calculating information capacity, adding geometrical distortion, loss of sharpness in the field, vignetting and color lens shading. MTF is assumed to be a function of distance *r* from the image center, i.e., the lens is assumed to be perfectly centered, which is something we've rarely observed in mass-produced lenses. 2½ of the 6 pages are devoted to a model of optical distortion correction. The slanted-edge method lets you make fast and accurate measurements of information capacity directly from images without and with correction for distortion (or other aberrations).

Imatest has published two papers that led to the development of the slanted-edge method.

Camera Information Capacity: A Key Performance Indicator for Machine Vision and Artificial Intelligence Systems (2020) [3] uses the Siemens Star pattern. It's particularly good for analyzing the effects of artifacts (demosaicing, data compression, saturation, etc.), but it's slower and less convenient than the slanted-edge technique.

Using images of noise to estimate image processing behavior for image quality evaluation (2021) [12] presented a method of illustrating noise anywhere in an image. Difficult to implement because it required signal averaging fora large number of images, but it led to the present technique, which averages signals in multiple lines of a slanted-edge ROI in a single image.

Appendix 3. Approximations and assumptions

There are a number of approximations behind the calculations that will need to be examined in the process of making the measurement a standard. We list them here.

(1) Binning noise correction needs to be verified for a variety of images.
(2) Noise spectrum. We assume it's white. In general, we find it's close to white (from flat areas near edges), with some rolloff near Nyquist. When an image is sharpened, high frequency noise is increased, which increases the overall noise measurement. We seem to obtain good results, even though the noise spectrum $N(f)$ is not adjusted for the sharpening.
(3) The average noise power measurement, N_{avg} , is now different for uniformly-processed and bilateral-filtered images. Are the current choices optimum?
(4) What to do about overshoots in strongly-sharpened images? Currently we use the settling levels (away from the edge) for $V_{p\text{-}p}$. We know that strong sharpening has a bad reputation in the machine vision industry, and it can also cause saturation for large signals.
(5) How can we handle pixel offsets (minimum values) in the image?
(6) How should we handle complex tonal response curves, often with response "shoulders" (highlights where local contrast is reduced to improve pictorial quality).
(7) Going from C_4 to C_{max} is a big step that invites scrutiny. It seems to be straightforward for uniformly-processed files with simple log-linear response curves. But for JPEG files, it can be affected by bilateral-filtering, the tonal response curve, and overshoots (if any).
(8) Should gamma-encoding play a role in information capacity? We suspect that it improves the capacity of files, especially files with bit depth of 8, by more uniformly distributing the pixel levels. But it would not affect the camera's intrinsic information capacity.

References

[1] 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. people.math.harvard.edu/~ctm/home/text/others/shannon/entropy/entropy.pdf

[2] C. Shannon, "Communication in the Presence of Noise", Proceedings of the I.R.E., January 1949, pp. 10-21. fab.cba.mit.edu/classes/S62.12/docs/Shannon_noise.pdf

[3] N. L. Koren, "Measuring camera Shannon Information Capacity with a Siemens Star Image", Electronic Imaging conference, 2020. The Camera Information Capacity (Siemens Star) white paper is much more readable.

[4] ISO 12233:2017: https://www.iso.org/standard/71696.html

[5] Robert D. Feite, "Modeling the Imaging Chain of Digital Cameras", SPIE Digital Library, 2010, Chapter 8, "The Story of Q".

[6] N. L. Koren, Correcting Misleading Image Quality Measurements, Electronic Imaging Conference, 2020

[7] 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. Reprinted in "Selected Readings in Image Evaluation," edited by Rodney Shaw, SPSE (now SPIE), 1976. Available for download, and strongly recommended.

[8] R. Jenkin and P. Kane, Fundamental Imaging System Analysis for Autonomous Vehicles, Electronic Imaging, 2018

[9] F. Cao, F. Guichard, H. Hornung, L. Masson, "Sensor information capacity and spectral sensitivities", DxO Labs, 3 rue Nationale, 92100 Boulogne Billancourt, FRANCE, 2009

[10] Frédéric Cao, Frédéric Guichard, Hervé Hornung, "Information capacity: a measure of potential image quality of a digital camera", DxO Labs, 3 rue Nationale, 92100 Boulogne Billancourt, FRANCE, 2010

[11] F.-X. Thomas, T. Corbier, Y. Li, E. Baudin, L. Chanas, and F. Guichard. "RAW Image Quality Evaluation Using Information Capacity", DXOMARK, Boulogne-Billancourt, France, 2021, Also available here.

[12] N. L. Koren, "Using images of noise to estimate image processing behavior for image quality evaluation", Electronic Imaging conference , 2021

[13] Orit Skorka and Paul J. Kane, “Object Detection Using an Ideal Observer Model”, Electronic Imaging conference, 2020.