Introduction to Image Information Metrics

Norman Koren, Imatest, LLC

December 5, 2023

The market for cameras that produce images for Machine vision (MV) and Artificial Intelligence (AI), in contrast to pictorial images for human vision, is steadily growing. Applications include automotive (driver assistance and autonomous vehicles), robotics, security, and medical imaging systems.

Two questions arise when designing camera systems for such applications.

1. How best to select (or qualify) cameras for MV/AI applications?
2. What image processing (ISP or filtering) is optimal?

To answer these questions, we must go beyond standard measurements of sharpness (MTF) and noise and apply metrics derived from **information theory**, including information capacity and related metrics for object and edge detection.

These metrics are important because Object Recognition (OR), MV, and AI algorithms operate on *information*, not *pixels*, making them far better predictors of system performance than MTF or noise.

Imatest has developed a highly convenient method for measuring information capacity and related metrics from the most widely used ISO standard resolution test pattern — the slanted edge. We describe how the new metrics can be used to select (or qualify) cameras and determine the optimum Image Signal Processing (ISP) for Object Recognition, which is likely to improve the performance of MV and AI algorithms.

This white paper is a simplified and much more readable version of "Image Information Metrics and Applications: Reference," which has the full technical detail, and is linked from www.imatest.com/solutions/image-information-metrics.

This document describes features of ***Imatest 24.1***, which will be available in the ***Imatest 24.1 Pilot program*** until the spring 2024 release.

Table of Contents

Introduction

Traditional image quality measurements, including sharpness, noise, dynamic range, optical distortion, tonal and color response, and spatial uniformity, have proven useful for human vision, where tradeoffs are evaluated visually. For example, sharpening makes fine features more visible, but increases noise. Choices are often based on experience; they come down to what looks best, i.e., what has the most pleasing appearance for the application.

Object Recognition (OR), Machine Vision (MV), and Artificial Intelligence (AI) systems are different. System performance is not dependent on image appearance. A more objective metric is required.

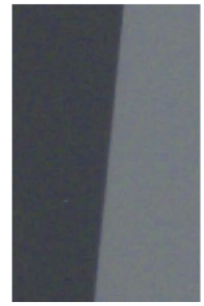
Information

Information is a metric that quantifies how much is learned from a measurement. For example, an individual pixel in a blurred image is highly correlated with its neighbors, so little is learned from its contents. But if the image is sharp, it is weakly correlated, and much more can be learned from its contents, i.e., it contains more information.

The concept of information dates from 1948 and 49 in two celebrated papers by Claude Shannon [1],[2]. Appendix I contains a brief introduction to information theory.

In electronic communications, information capacity is the maximum rate that information can be transmitted through a channel without error. In images, it is the maximum amount of information that a pixel or image can hold.

The slanted edge



Imatest calculates information capacity from the ***slanted edge***, which is a key part of the ISO 12233 standard, "Photography — Electronic still picture imaging — Resolution and spatial frequency responses" [3], is the most convenient and widely used resolution test pattern. It is highly efficient in its use of space (with multiple edges, sharpness can be mapped over the image surface), and calculations are very fast.

Imatest offers several charts with multiple edges that can be automatically detected and rapidly analyzed. Some of the charts offer additional color, tone, noise, and distortion analysis.

The ISO 12233 algorithm linearizes the image, finds the center of each scan line, fits a curve to the centers, then uses the curve to add each appropriately shifted scan line to one of four bins. The bins are combined to form a 4× oversampled averaged edge, which is used to calculate MTF.

This white paper contains a brief, abbreviated description of the calculations. The full description, with all the equations, is in "Image Information Metrics and Applications: Reference," linked from www.imatest.com/solutions/image-information-metrics

The Edge Variance method

The Edge Variance method uses an overlooked capability of the ISO 12233 slanted-edge binning algorithm to calculate information capacity.

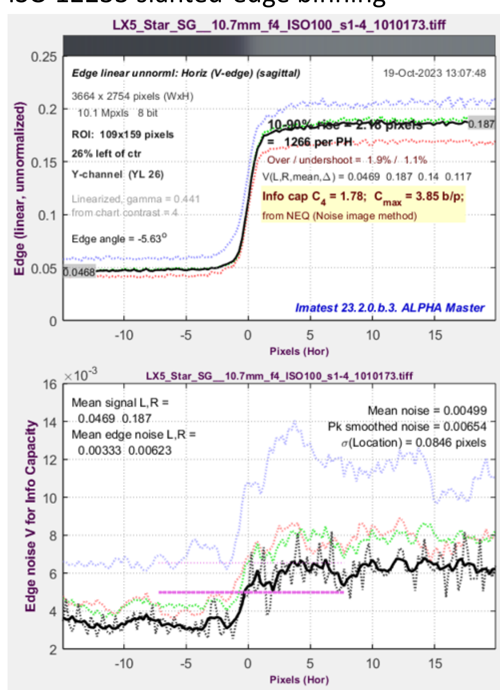
By summing of the squares of each scan line, $\rho_s(x) = \frac{1}{L}\sum_{l=0}^{L-1} y_l^2(x)$, we calculate the edge variance (the spatially dependent noise power) $\sigma_s{}^2(x) = N(x)$ and noise amplitude $\sigma_s(x)$ in addition to the mean, $\mu_s(x) = \frac{1}{L}\sum_{l=0}^{L-1} y_l(x)$.

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

Signal and noise results

The figure on the right shows the results of the ISO 12233 calculation, including the Edge Variance method of measuring spatially dependent noise.

- Upper plot: the average 4× oversampled edge, $\mu_s(x)$. The thick black line is the luminance channel. Information capacities are shown with a yellow background.
- Lower plot: the noise amplitude (voltage), $\sigma_s(x)$. The thick black line is the smoothed luminance channel. $\sigma_s(x)$ plot is a new measurement: spatially dependent noise was previously difficult to observe.



Edge amplitude and spatially dependent noise, calculated by the Edge Variance method

Calculating information capacity from $\mu_s(x)$ and $\sigma_s(x)$

The next step is to calculate the information capacity, C, typically in units of bits per pixel, by entering appropriate values of the signal and noise power, $S(f)$ and $N(f)$, into the **Shannon-Hartley equation**.

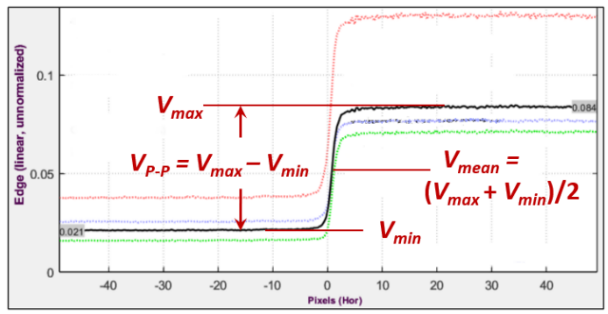
$$C = \int_0^W \log_2\left(1 + \frac{S(f)}{N(f)}\right) df$$

$S(f)$ and $N(f)$ are frequency-dependent signal and noise power, and W is the bandwidth, which is always equal to 0.5 cycles/pixel (the Nyquist frequency). Frequency-dependence is entered into $S(f)$ using $MTF(f)$ (described below).

Signal power S

Assuming that the signal is uniformly distributed over the $V_{p\text{-}p}$ range, the *average* frequency-dependent signal power, $S_{avg}(f)$, is

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



Signal amplitude from slanted edge.

Noise power N

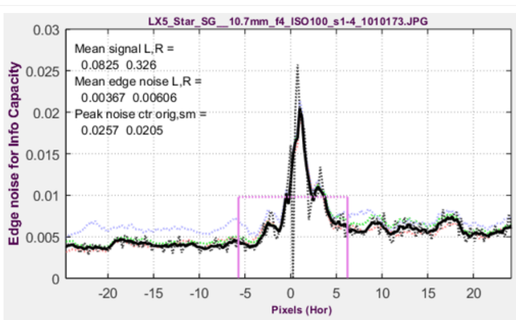
Noise power N has the same units as signal power S; hence S/N is dimensionless.

In examining a great many images, we observe two broad classes of images with very different noise properties, visible in $\sigma_s(x)$. We call them (1) uniformly/minimally processed and (2) bilateral filtered images. The value of noise power, N, used to calculate C, is different for the two image types.

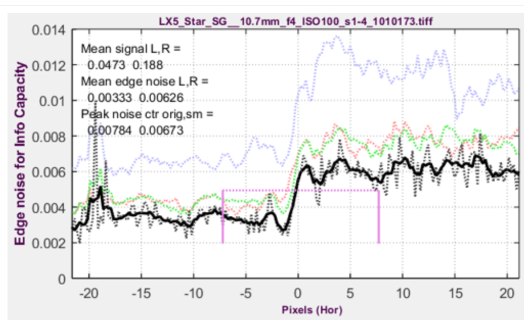
1. Uniformly/minimally processed images use the mean value of $\sigma_s^2(x)$ for calculating C. They are preferred when available, and should *always* be used for evaluating cameras for MV/AI systems.

2. Bilateral-filtered images, which include most in-camera JPEGs, are sharpened near edges and often noise-reduced (lowpass filtered) away from the peaks. They can be identified by a peak in the spatially dependent noise. The smoothed peak noise is used for calculating C.

   Special care is required when measuring C with bilateral-filtered images because the noise reduction can increase the measured value of C while *reducing* information. That is why the noise at the peak (where noise reduction is least likely to be applied) is used. The strong peak (below, left) is a signature of bilateral filtering.



Noise $\sigma_s(x)$ for bilateral-filtered image



Noise $\sigma_s(x)$ for uniformly processed image

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 reduce it by causing aliasing — spurious low-frequency signals like Moiré that can interfere with the true image. Frequency-dependence comes from $MTF(f)$.

Combining $S_{avg}(f)$, N, and W to obtain C

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

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

C is measured with relatively low contrast test charts to minimize errors from saturation to ensure that the camera is operating in its linear region. For most of our work, we use charts with a 4:1 contrast ratio, following the ISO 12233 standard [3]. Since $V_{P\text{-}P}$ is directly proportional to chart contrast, we label C according to the contrast ratio: C_n for n:1 contrast ratio. We use C_4 throughout this document.

C_4 is highly dependent on the exposure level, and does ***not*** represent the maximum information capacity of the camera.

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

C_4 is strongly dependent on exposure because (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 (derived from image sensor properties).

We have developed a metric for maximum information capacity: C_{max}, that is nearly independent of exposure. It is obtained in two steps.

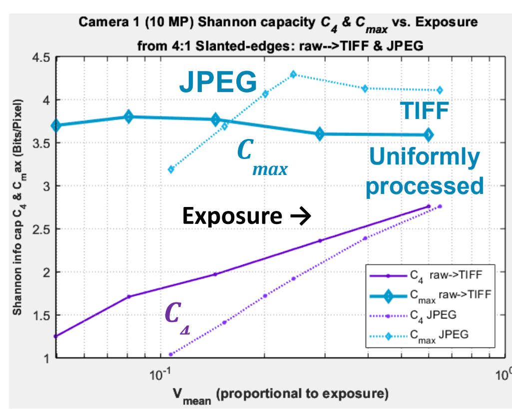
Step1: Replace the measured peak-to-peak voltage range $V_{p\text{-}p}$ with the maximum allowable value, $V_{p-p\_max} = 1$.

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). This is not difficult for linear sensors where the relationship between V and N is known, but it can be complex for HDR (High Dynamic Range) sensors. Calculation details can be found in "**Image Information Metrics and Applications: Reference**," linked from www.imatest.com/solutions/image-information-metrics.

C_{max} may need to be adjusted if the image is incapable of spanning the entire range of Digital Numbers (DNs), for example, 0-255 for images with bit depth = 8. Information capacity measurements fail if local tone mapping has been applied.

Consistency of C_{max}

We performed a set of analyses with a range of exposures (indicated by V_{mean}). The results showed that C_{max} is highly consistent with exposure for the raw→TIFF images (which were not bilateral-filtered), but less consistent with the bilateral-filtered (JPEG) images. C_4 varied as expected.



C_4 and C_{max} for raw→TIFF and JPEG image

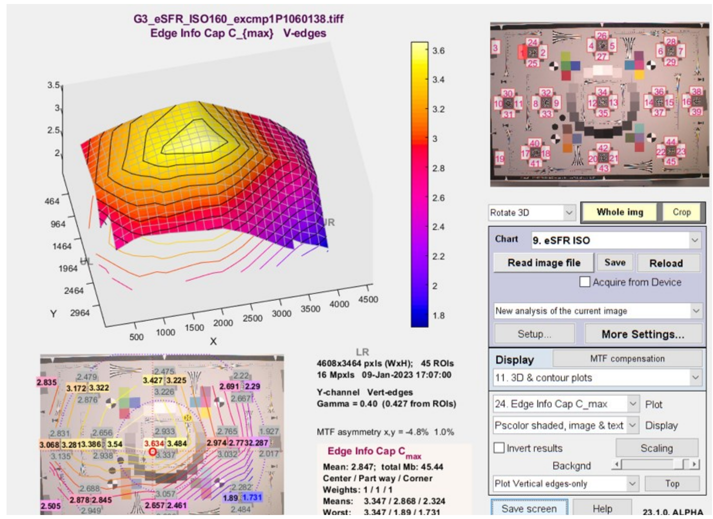
Total information capacity

Thus far, we have presented information capacity C in bits per pixel. The total information capacity, C_{total}, for the entire image accounts for variations in C over the image.

C_{total} and the mean value of C_{max} for auto-detected slanted-edge modules, SFRplus, eSFR ISO, or Checkerboard, are shown on the right.

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

The mean information capacity C_{max} is 2.847 bits/pixel. The total capacity C_{total} for the Luminance (Y) channel of the 16 MP camera = 45.44 MB.



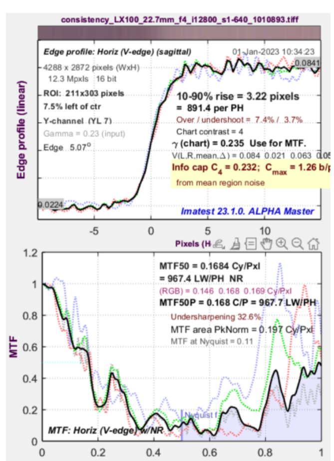
3D plot of information Capacity C_{max} over the image surface

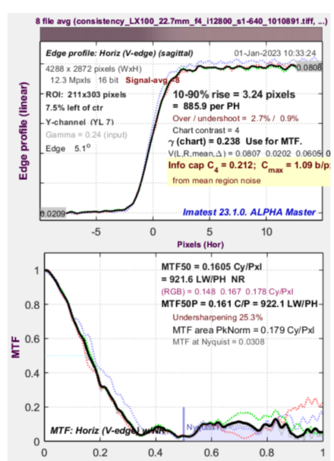
Signal averaging

Signal averaging is a well-known technique that can improve the accuracy and consistency for measurements of noisy images, typically acquired in dim light.

Signal averaging, where n identical captures of the same image are averaged, is a classic technique for obtaining better measurement consistency by reducing the effect of uncorrelated noise. When n images are averaged, SNR increases by $\sqrt{n}$: by 3dB whenever n is doubled. To obtain correct information capacity measurements when the signal is averaged, the noise power is multiplied by n.

This effect is illustrated below for a camera with set to ISO 12800. A single image is shown on the left. Note that MTF is rough and has significant high frequency noise bumps. The average of 8 images is shown on the right.



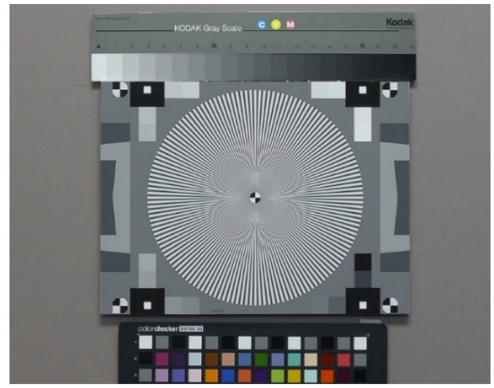


Single image ***n = 8 averaged***

Some key results of the Edge Variance method

We tested three cameras with different pixel sizes that produced both raw and JPEG output for information capacity C as a function of Exposure Index (ISO speed setting).

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).

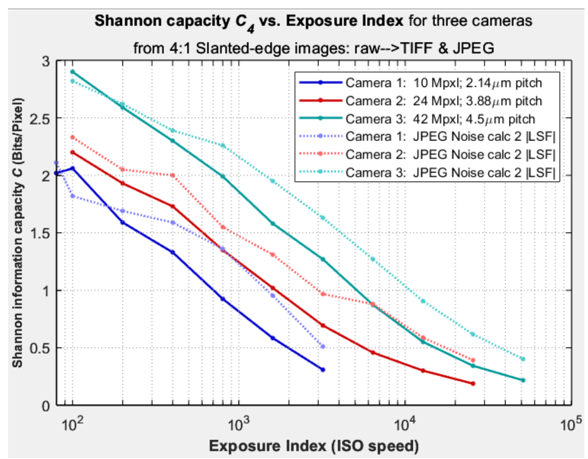


The figures below show results for the luminance (Y = 0.2125·R + 0.7154·G +0.0721·B) channel as a function of ISO speed (Exposure Index) for the raw→TIFF images (solid lines) and JPEG images (dotted lines). For the raw→TIFF images the relationship between ISO speed and C is similar for all three cameras.

C_4 4:1 slanted edge

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

C_{max} showed similar trends, but results were about 2 bits/pixel higher.



C_4 as a function of Exposure Index (EI) for TIFF and JPEG images

Sharpening

Simple sharpening, which has the same effect on the signal and noise response, and therefore does not change $S(f)/N(f)$, would not be expected to have a strong effect on C. We confirmed this by comparing an unsharpened image with two strongly sharpened images (all derived from a minimally processed TIFF from a consumer camera at ISO 100) characterized by USM (Unsharp Mask) Radius *R* and Amount *A*. Differences in C are numerical errors.

	MTF50 C/P	Peak MTF	C_4 B/P	C_{max} B/P
Unsharpened	0.265	1	2.06	3.82
USM R1 A2	0.521	1.22	1.84	3.71
USM R2 A2	0.552	1.9	1.93	3.81

This highlights another benefit of information capacity measurements. Unlike MTF50, they do not reward excessive sharpening, which creates "halos" near edges, making the image look sharp in small displays, but creating artifacts that degrade image appearance on large displays [5]. They also have a bad reputation for machine vision applications.

Summary of the Edge Variance method

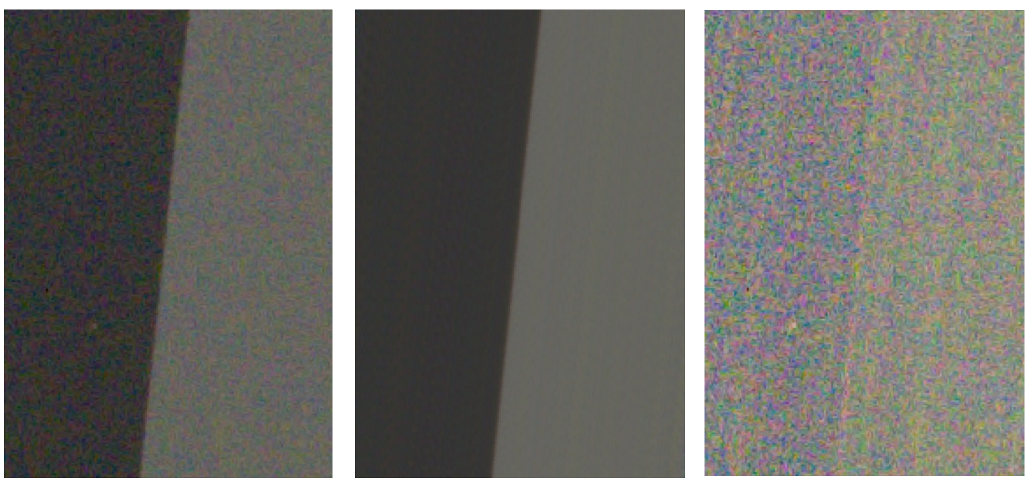
- The Edge Variance method has a limited set of results.
- Produces a useful approximate measurement of C for bilateral-filtered images, but more accurate results are obtained from uniformly/minimally processed images, which should *always* be used when a camera is being evaluated for use in MV/AI systems.

The Noise Image method of calculating information capacity-related metrics

The Noise Image method is the second of two methods for calculating information capacity, along with a rich set of related metrics.

This 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 low-noise interleave (**Inverse binned… ROI**, below). Since the inverse-binned image is a nearly noiseless replica of the original image, we can create a noise image by subtracting it from the original image.

The three images are shown below. The noise image (below-right) has been lightened for display.



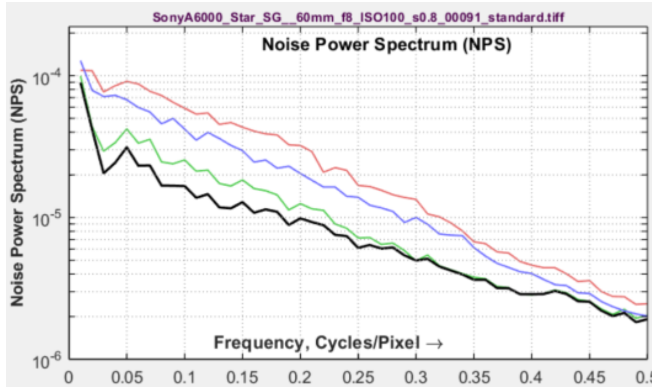
(1) Original ROI ***(2) Inverse-binned / de-interleaved / reverse-projected*** ***(3) Noise image ROI***

This technique allows several additional image quality parameters to be calculated, including an alternative information capacity measurement, C_{NEQ}, derived from NEQ.

Noise Voltage or Power Spectrum (NPS)

The Noise Power and Voltage Spectrum plots have the same shape: the y-axis labels are different because Power is the square of Voltage.

NPS is used for the NEQ and $SNRi$ calculations.



Noise Power Spectrum (NPS)

Noise Equivalent Quanta (NEQ)

NEQ [4] is a frequency-dependent Signal-to-Noise (power) Ratio, in contrast to MTF, which is signal amplitude response-only.

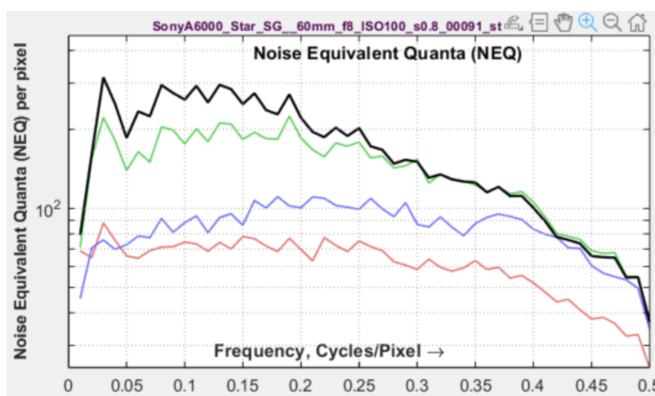
Units are the equivalent number of detected quanta that would generate the measured SNR when photon shot noise is dominant.

$$NEQ(f) = \frac{\mu^2 MTF^2(f)}{NPS(f)}$$

where the mean linear signal, μ, depends on how NEQ is to be interpreted.

In the standard definition of NEQ, where NPS is dominated by photon shot noise, $\mu^2 = V_{mean}^2 = \bar{q}^2$, where $\bar{q}$ is the mean count of the detected quanta, and NEQ is proportional to the count of detected quanta, $\bar{q}$.

The NEQ plot can be made smoother and more consistent using Signal Averaging.



Noise Equivalent Quanta (NEQ)

Information capacity from NEQ, C_{NEQ}

A variant of NEQ, $NEQ_{info}(f)$ (not plotted), calculated using $\mu = V_{P-P}/\sqrt{12}$ (to be consistent with the Edge Variance calculation), is used to calculate information capacity, C_{NEQ}.

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

where bandwidth $W = f_{Nyq} = 0.5$ Cycles/Pixel.

The key results, $C_4(NEQ)$ and $C_{max}(NEQ)$, included in the Results summary, are slightly different from the Edge Variance results, most likely because the calculated Noise Power Spectrum, $NPS(f)$, 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)	=	3.8	4.37	3.94	4.45
Info capacity C_4(NEQ)	=	1.55	2.07	1.65	2.15
Noise image variance	=	1.68e-05	7.82e-06	1.3e-05	6.25e-06
Noise image std dev	=	0.0041	0.0028	0.00361	0.0025
Noise 1D power	=	1.77e-05	8.51e-06	1.37e-05	6.96e-06
Input signal V(P-P)	=	0.105	0.103	0.0879	0.103

Results summary

Ideal Observer SNR (*SNRi*)

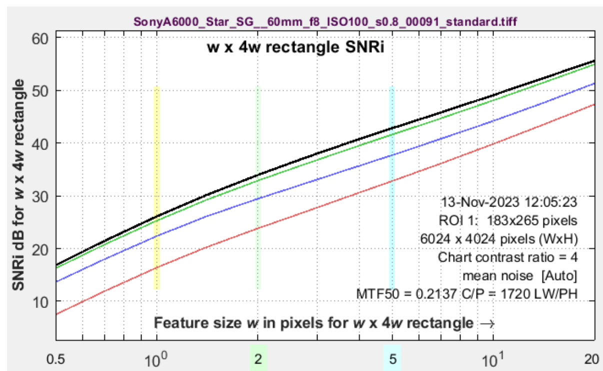
SNRi is a measure of the detectability of objects, described in papers by Paul Kane [8] and Orit Skorka and Paul Kane [9].

$$SNRi^2 = \iint \left(\frac{\left|G(v_x, v_y)\right|^2 MTF^2(v_x, v_y)}{NPS(v_x, v_y)} \right) dv_x \, dv_y$$

$G(v_x, v_y)$ is the Fourier transform of the rectangular object to be detected, defined below. v is spatial frequency in Cycles/Pixel.

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

The objects to be detected 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 obtained from a chart with a 4:1 contrast ratio. *SNRi* is displayed for each color channel for w from 0.5 to 20.



SNRi for sharp, low-noise (ISO 100) image

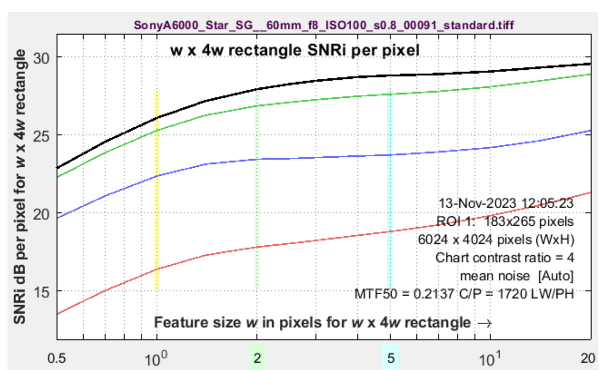
Note that like C_4, $SNRi$ is strongly affected by exposure and chart contrast. But unlike C, *SNRi* is affected by image signal processing (sharpening, etc.).

We currently prefer a closely related measurement, ***Edge SNRi***, for determining the performance of pre-filtering (ISP performed prior to Object Recognition/Machine Vision/AI).

SNRi displayed in dB per pixel²

Because standard *SNRi* plots can be difficult to read, we have added a plot of *SNRi* in dB per pixel², shown on the right. 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.

Tip— Click on Data cursor in the dropdown below the thumbnail on the upper right to get a reading of the actual value.

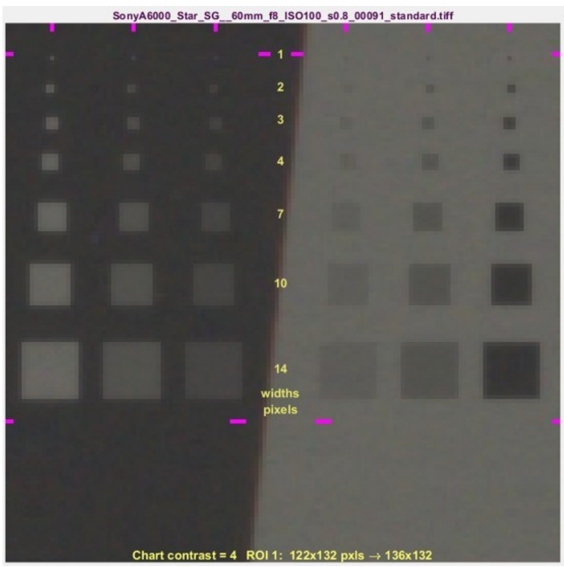


SNRI in dB per pixel² for low-noise (ISO 100) image

Object visibility

The goal of *SNRi* measurements is to predict object visibility for small, low contrast squares or 4:1 rectangles. We have developed an ***Imatest*** display that does this with real slanted-edge images.

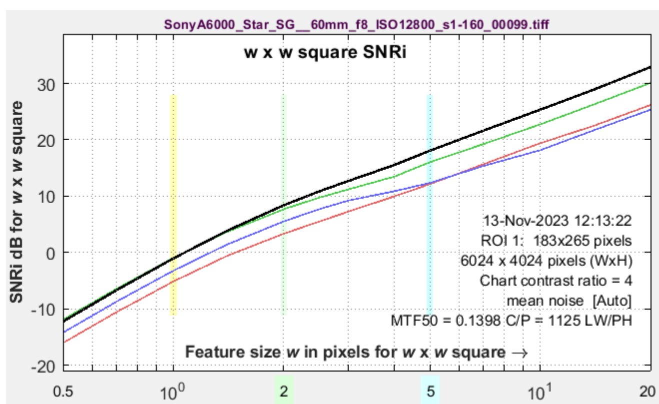
We show two images, below, from a camera with a Micro Four-Thirds sensor. The original chart has a 4:1 contrast ratio (light/dark = 4). The middle and inner squares have reduced contrast. The outer patches correspond to the *SNRi* curves, where, according to the Rose model [6], *SNRi* of 5 (14 dB) should correspond to the threshold of visibility. low noise ISO 100 (left); noisy ISO 12800 (right)



Low noise ISO 100 (left)
***MTF50* = 0.214 c/p; C_{max} = 4.24 b/p;**

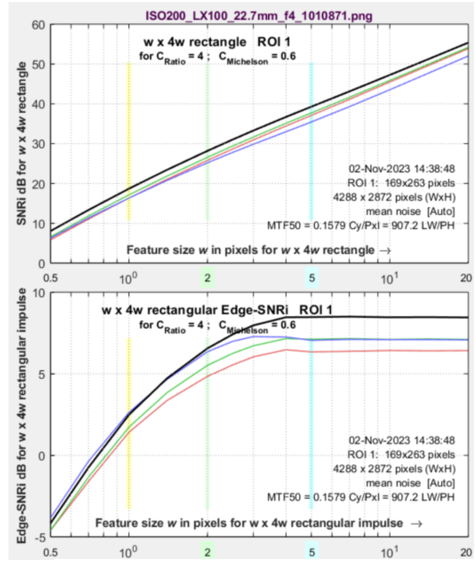
Noisy ISO 12800 (right)
***MTF50* = 0.140 c/p; C_{max} = 1.37 b/p.**

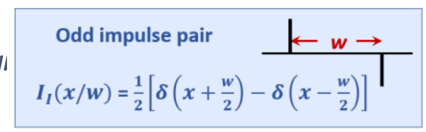
The *SNRi* curve on the right is for the noisy ISO 12800 image on the right, above. The *w* = 1 squares are invisible; the *w* = 2 and 3 squares are only marginally visible, and *w* = 4 squares are clearly visible. In the plot, the Y (luminance) channel *SNRi* at *w* = 2 is 9 dB; it reaches 11 dB for *w* = 3; close to the expectation that the threshold of visibility is around 14 dB.



SNRi for noisy ISO 12800 image (above, right)

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





Edge SNRi is an edge-based measure of the detectability of the edges of small objects, similar to *SNRi*, described above.

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

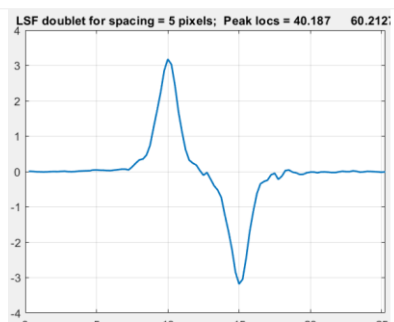
$H(\nu_x,\nu_y)$ is the Fourier transform of the *edges* (the gradient) of the object to be detected.

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, equivalent to a pairs of Dirac delta functions of opposite polarity.

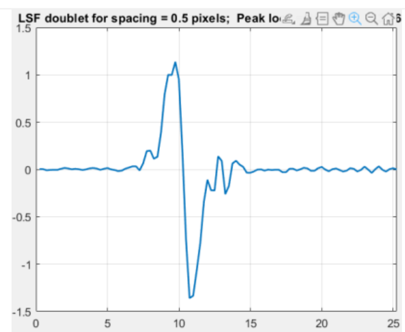
Unlike C, *Edge SNRi* is affected by signal processing (sharpening, etc.), making it useful for evaluating pre-filtering (ISP filtering applied prior to the machine learning/AI blocks). ***Edge SNRi* has become our favorite metric for feature detection.**

Line Spread Function (LSF) doublet results

Edge SNRi is based on pairs of Line Spread Functions of opposite polarity called *LSF doublets*, illustrated for w = 5.0 and 0.5 pixels.

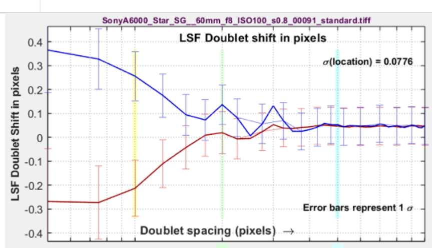


LSF Doublet. w = 5.0 pixels.

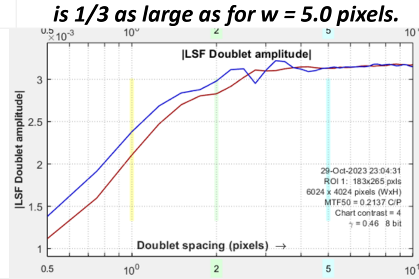


LSF Doublet. w = 0.5 pixels. Amplitude is 1/3 as large as for w = 5.0 pixels.

As spacing w decreases, the peaks are closer (but shifted more from their original locations), and amplitude decreases.



LSF Doublet shift as a function of spacing w



LSF Doublet amplitudes as a function of spacing w

Summary of the Noise Image method

- The Noise Image method uses a 2D image of the noise to calculate several image quality metrics.

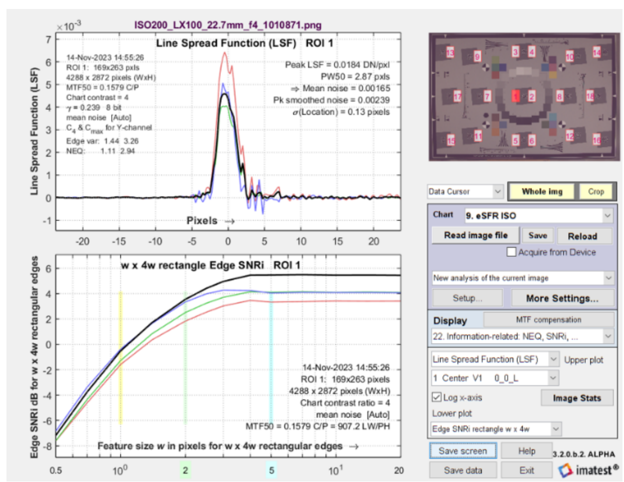
- It only gives reliable results with uniformly or minimally processed images, which can be distinguished from bilateral-filtered images by the absence of a peak in $\sigma_s{}^2(x)$ or $\sigma_s(x)$ displays. It should not be used with bilateral-filtered images.
- It produces a rich set of related results, including deal observer SNR ($SNRi$), $Edge\ SNRi$, Noise Equivalent Quanta (NEQ), and a second information capacity measurement, derived from NEQ, that can be compared with the Edge Variance results.

Image Signal Processing (ISP)

Several recent papers [10],[11],[12] state that appropriate image processing prior to Object Recognition, Machine Vision, or AI algorithms may improve system performance (accuracy, speed, and power consumption). Because information capacity is relatively independent of Image Signal Processing, it provides little guidance about how to design optimal image processing.

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 its edges. In practice, ISP needs to perform well over a range of tasks: detecting objects of edges greater than a minimum size and limiting interference from neighboring objects.

SNRi has some drawbacks as an object detection metric. Plots of *SNRi* are challenging to interpret because *SNRi* increases with feature size. And there is the problem of object color. What If the object has the same color as the background (e.g., gray cars in front of gray concrete)? In such cases it is the *edge* that matters. Because of these shortcomings, we prefer ***Edge SNRi***.



Results with no filtering. LSF (top), Edge SNRi (bottom)

Pre-filtering: effects of ISP filtering

Starting with an unsharpened image, we applied sharpening and/or lowpass filtering (blurring).

The Line Spread Function and $Edge\ SNRi$ for a $w{\times}4w$ rectangle are shown on the right. We selected smoothed peak noise for the calculations.

The key results ($Edge\ SNRI$ and $SNRi$ in dB per pixel^2) for a w×4w object are shown in the Table.

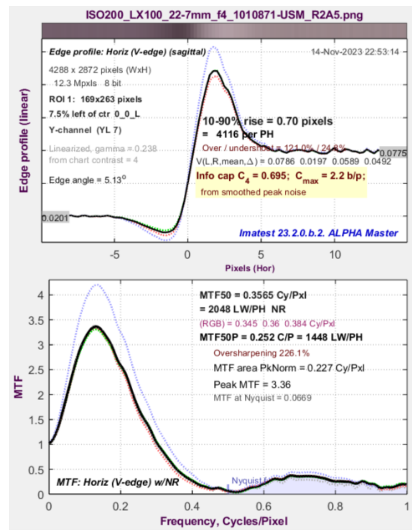
Filter	*MTF50 C/P*	*Edge SNRi* w = 1	*Edge SNRi* large w	*SNRi* dB/pxl^2 w = 1	*SNRi* dB/pxl^2 w = 5	C_{max} (NEQ)	σ(loc.) pixels
None	0.158	-0.52	5.49	21.7	28.3	2.94	0.13
USM R2A3	0.294	-1.5	4.7	20.7	26.5	2.73	0.219
USM R2A3 + σ = 1 Gaussian LPF	0.243	6.1	9.8	24.7	30.0	3.44	0.105
σ = 1 Gaussian LPF	0.122	1.56	8.5	24.9	33.7	2.72	0.89
USM R2A5 (*extreme oversharpening*)	0.357	-6.8	-1.1	14.7	20.1	2.02	0.26

Good news! Edge SNRi **(9.8 dB for large *w*; 6.1 at *w* = 1) was better for the USM + Gaussian lowpass filter than either the unfiltered or USM-only filtered image.** This is an extremely significant result. It shows that correctly chosen filtering can improve the performance of a key task (edge detection) before the image is sent to the Object Recognition/Machine Vision/AI processing block.

This important result shows that filtering can improve object detection, indicating that it may be able to improve Object Recognition, Machine Vision, and Artificial Intelligence system performance.

Edge SNRi appears to be slightly more sensitive than *SNRi* dB per pixel^2 (showing greater differences for different filtering). Sharpening + lowpass filtering gives the best result.

The excessively oversharpened USM R2A5 image, plotted on the right is illustrated because it's all too common, and we do our best to discourage it: it is a cheap way of improving *MTF50* measurements and image appearance on tiny displays (phones), but it creates "halos" (peaks near edges) that degrade appearance in large displays. The poor *Edge SNRi* and other results are additional reasons to avoid this type of image processing, as have described in [5].



Edge/MTF plot for extremely oversharpened image. MTF50 correlates poorly with performance.

Matched filter

In the above section, we discussed a applying a filter $\mathcal{F}(f)$ to optimize either *SNRi* or *Edge SNRi*.

An optimum filter can be determined for a specific task (for example, detecting an edge of a certain size, with no interference from neighbors). Such a filter is called a *matched filter*. Since real-world cameras must perform a multitude of tasks, exact matched filters are rarely practical. They are discussed in more detail in "**Image Information Metrics and Applications: Reference**," linked from www.imatest.com/solutions/image-information-metrics.

We were fortunate that the filter parameters we found by trial-and-error were reasonably close to the matched filter calculated from the image properties (sharpness and noise).

Exposure sensitivity

Like C_4, *SNRi* and *Edge SNRi* vary with exposure. Because there is no consistent relationship between exposure and noise in HDR sensors, we will need a **Standard exposure** for comparing cameras (and it will need to be in the nascent ISO 23654 standard). For images encoded with $gamma \approx 0.454 = 1/2.2$ (sRGB, etc.), $V_{max} \approx 0.5$ is reasonable. For linear ($gamma = 1$) images, the equivalent exposure results in $V_{max} = 0.5^{2.2} = 0.22$ (where V_{max} is normalized to a maximum of 1).

Summary

We have developed a powerful toolkit of new measurements — Figures of Merit for imaging systems that combine sharpness and noise — that are especially applicable to Object Recognition, Machine Vision, and Artificial Intelligence systems. The key measurement is ***information capacity***, which can be used to predict a camera's potential performance for MV/AI systems. We have developed additional metrics related to specific tasks, most importantly object and edge detection, that are potentially useful for designing ISP filters that optimize OR/MV/AI system performance.

Using ***Edge SNRi***, which is closely related to the more traditional object-based *SNRi*, we have shown an example of image processing (sharpening + lowpass filtering) that improves object detection and is likely to improve MV/AI system performance. This needs to be tested.

As we become more familiar with information capacity and determine the requirements for effectively performing tasks, we should be able to select cameras with the minimum number of pixels to meet the spec, resulting in faster calculations, lower power consumption, and reduced cost.

The key takeaway from this document is that

Information capacity *C* is a key measurement for predicting Object Recognition/Machine Vision/Artificial Intelligence system performance.

Several additional metrics based on *C*, most importantly *Edge SNRi* (for edge detection), can be used to design filters to optimize OR/MV/AI system performance.

Standard MTF measurements are insufficient for this purpose.

To do — Better understand the numeric results for *SNRi* and *Edge SNRi*.

Partner with researchers in industry and academia to determine the correlation between information capacity and related metrics and MV/AI system performance.

Find better ways of characterizing information capacity in High Dynamic Range (HDR) sensors, where noise is not a simple function of signal.

Continue working on the new **ISO 23654** standard for camera 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] ISO 12233:2017: https://www.iso.org/standard/71696.html

[4] Brian W. Keelan, "Imaging Applications of Noise Equivalent Quanta" in *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.

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

[6] 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

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

[8] 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.

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

[10] 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

[11] 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

[12] 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/

Appendix I. Information theory background

Because concepts of information theory are unfamiliar to most imaging engineers, we present a 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 classic 1948 and 1949 papers [1],[2] are highly readable.

What is information?

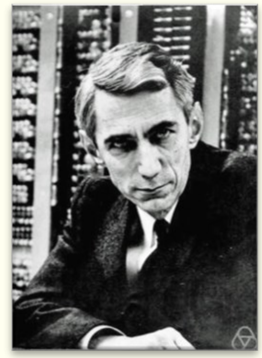
Information is a measure of the resolution of uncertainty. The classic example is a coin flip. For a "fair" coin, which has a probability of 0.5 for either a head or tail outcome (which we can designate 1 or 0), the result of such a flip contains one bit of information. Two coin flips have four possible 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. But for the warped coin, where $p(x_1) = 0.95$ and $p(x_2) = 0.05$, $H = 0.286$ bits. If the results of the warped coin toss were transmitted without coding, each symbol would contain 0.0286 information bits. That would be extremely inefficient.



Claude Shannon

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, 11 Life Lessons From History's Most Underrated Genius, is a great read. (Perhaps Shannon is considered "underrated" because history's most famous genius lived in the same town.) There are also nice articles in The New Yorker and Scientific American. And IEEE has an article connecting Shannon with the development of Machine Learning and AI. The 29-minute video "Claude Shannon – Father of the Information Age" is of particular interest to the author of this report because it was produced by the UCSD Center for Memory and Recording Research, which I visited frequently in my previous career.

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 without error. 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 0.286.

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 Gaussian noise power, N, is

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

This equation is challenging to 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. Slanted-edge analysis is one-dimensional. We have developed methods for calculating C for both the Siemens star and slanted edge test patterns.

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 coding is important 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, there were no convenient methods of measuring it.