New camera quality measurements for optimizing Machine Vision systems

Norman Koren, Founder and CTO, Imatest, LLC

May 22, 2024 Revised June 28, 2024

Motivation: Traditional metrics such as sharpness (MTF or SFR) and noise, taken by themselves, are not adequate for predicting Machine Vision/Artificial Intelligence (MV/AI) system performance.

We describe new metrics, based on camera information capacity, that are superior predictors of system performance.

Outline of the talk

- **Review**: what is information and how is it measured?
- **Define** information metrics and show how they are calculated.
- **Key image information metrics**
  - ***SNRi***: Independent observer SNR (for object detection)
  - ***Edge Location σ***: uncertainty of edge location (for edge detection)
- **Matched filter** for optimizing MV/AI system performance
- **Examples** showing effects of illumination and image processing (filtering)

Auto Sens

We are working on incorporating the new metrics into **ISO 23654, Photography — Digital cameras — Image Information Metrics**. Your participation is encouraged. The next meeting is June 11-14 in New York.

Information concepts

***Information*, defined by Claude Shannon in 1948, is a measure of the resolution of uncertainty. It is the basis of all electronic communications.**

For a system with S possible states, s_1, ..., s_n, with probabilities $p(s_\mathrm{n})$, information can be represented as ***entropy***, $\boldsymbol{H(S) = \sum_{i=1}^{n} p(s_1) \log_2(1/p(s_1))}$. $\mathbf{Log_2}$ is the key.

The number of states S is closely related to the Signal-to-Noise Ratio (S/N or SNR) of a continuous system.

Electronic channels — including cameras — can be characterized by a channel capacity, $\boldsymbol{C}$ (the maximum rate that information that can be transmitted without error), calculated from the Shannon-Hartley equation.

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

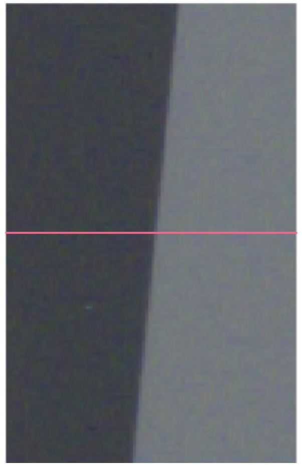
units are bits/pixel or bits/image.
Inputs are *bandwidth, W*, *average signal power, S*, and ***noise power, N*.**

The key performance indicators for MV/AI systems are closely related to *C*.

imatest®

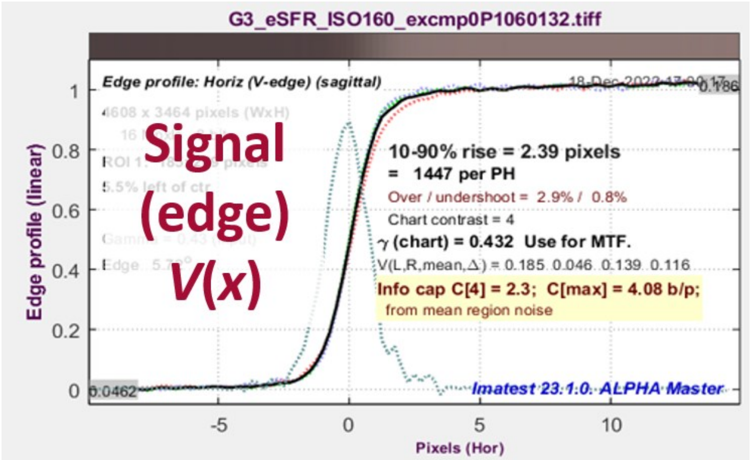
Information capacity from the slanted edge

The key to conveniently calculating information capacity, ***C***,
is to ***measure signal and noise in the same location***.



This can be accomplished with the widely used **slanted edge** test pattern, which is a part of the ISO 12233 standard. It's fast and compact enough to map *MTF* over an entire image. The ISO algorithm

- Linearizes the image,
- Finds the center of each scan line,
- Fits the centers to a polynomial,
- Adds each shifted scan line to one of four bins to obtain a 4x oversampled averaged edge, $V(x) = \mu_s(x)$, shown on the right, which is used to calculate *MTF* and information metrics.



This effectively reduces noise by $\sqrt{\text{samples in each bin}}$.

Best results are obtained when edge ROI length ≥ 100 pixels. [e.g., 100 pixels in 4 bins (25 per bin) reduces noise by $\sqrt{25} \approx 5$.]

imatest®

Measuring noise from the slanted edge

To obtain the spatially dependent noise power for calculating C, $N(x)$, *measured at the same location as the signal,*

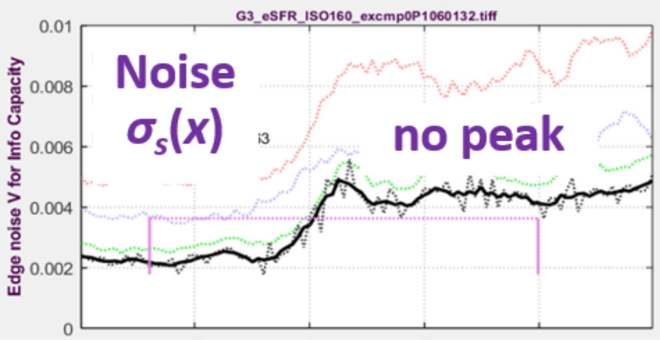
Sum the *squares* of each scan line to find the variance, $\sigma_s{}^2(x) = N(x)$,

$$\text{Spatially dependent noise power } N(x) = \sigma_s^2(\mathrm{x}) = \frac{1}{L}\sum_{l=0}^{L-1} y_l^2(\mathrm{x}) - \mu_s^2(x)\ ^*$$

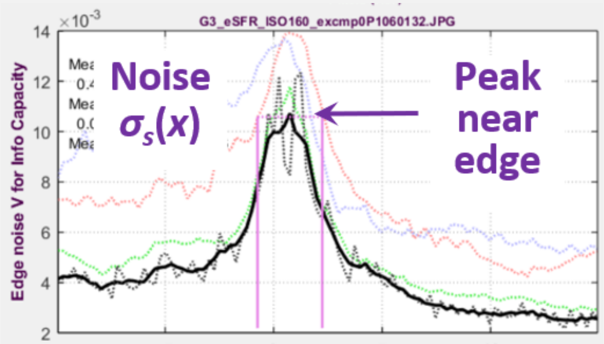
*$N(x)$ is the mean of the squares minus the square of the mean, for each point x.

Noise amplitude $\sigma_s(x) = \sqrt{N(x)}$ can now be viewed. Examples:

Uniformly-processed



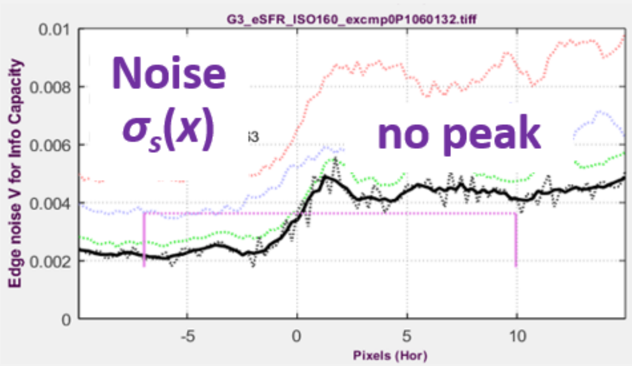
Bilateral-filtered



$N(x)$ can look *very* different for different types of image processing.

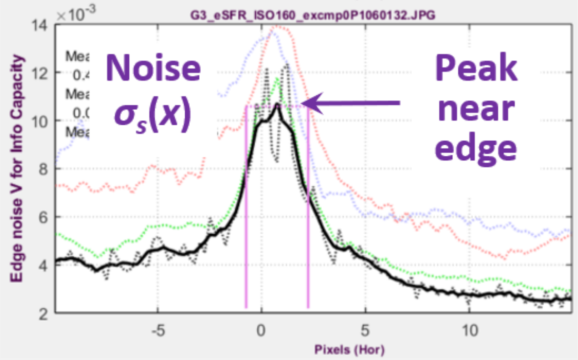
$N(x)$ does not fully characterize the noise. We still need to calculate the *Noise Power (Wiener) Spectrum*.

Noise power $N(x)$ for calculating information capacity, C
depends on the detected image processing type



Noise amplitude $\sigma_s(x) = \sqrt{N(x)}$

A peak in $N(x)$ indicates bilateral (nonuniform) filtering.



Uniformly or minimally processed

Unsharpened or uniformly sharpened. No noise reduction

Little or no noise peak.

C is calculated from $N = \text{mean}\big(N(x)\big)$.

Most accurate calculation; best for camera performance.

Required for calculating image information metrics.

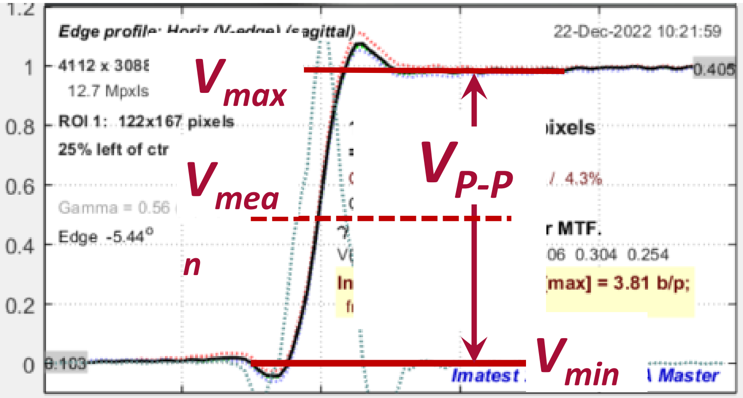
Bilateral-filtered

Sharpened near the edge; noise-reduced elsewhere. Most JPEG images from consumer cameras

Identified by distinct noise peak

C is calculated from the smoothed noise power at the peak, where *MTF* is calculated.
$N = N_{peak\text{-}smooth}$.
Less accurate than uniformly processed.

Signal power for calculating information capacity C



The Shannon-Hartley equation uses the **mean** signal power, $S_{mean}(f)$, to calculate C.

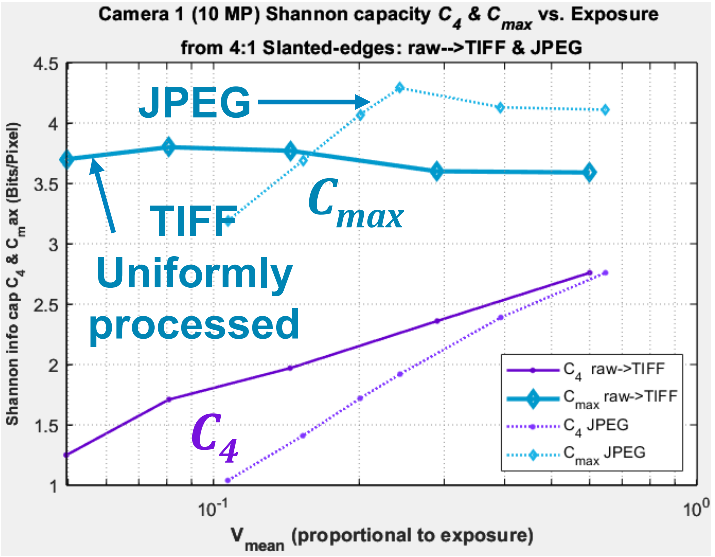
Information capacity is maximum when the signal is uniformly-distributed over $V_{P\text{-}P}$.

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

The three factors, S_{mean}, N, and bandwidth $W = f_{Nyq} = 0.5$ C/P are entered into the Shannon-Hartley equation.

$$\textbf{Information capacity} = C = \int_0^W \log_2\left(1 + \frac{S_{mean}(f)}{N}\right) df \cong \sum_{i=0}^{0.5/\Delta f} \log_2\left(1 + \frac{S_{mean}(i\Delta f)}{N}\right)\Delta f$$

Calculating information capacity C_n and C_{max}



The measured value of C is a strong function of the chart contrast ratio as well as the exposure.

For this reason, we recommend specifying the chart contrast ratio when reporting C, for example, C_4 for widely used ISO standard 4:1 contrast charts.

Since C_4 is strongly dependent on chart contrast ratio and exposure, we have developed a more stable metric, ***Maximum information capacity, C_{max}***, by extrapolating $V_{p\text{-}p}$ to $V_{max} = 1$ (the maximum allowed value) and adjusting the noise, which can be challenging for HDR sensors.

C_{max} is a stable measurement, nearly independent of exposure, that can be used to characterize cameras, but

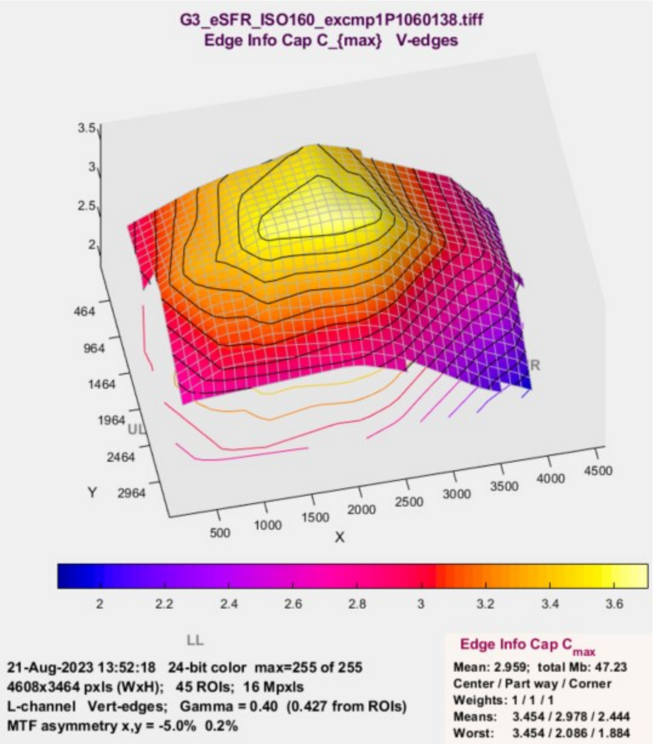
C_4 is useful for characterizing camera performance as a function of exposure.

Information capacity displays for C_4 and C_{max}

The 3D plot shows C_{max} mapped over the image.

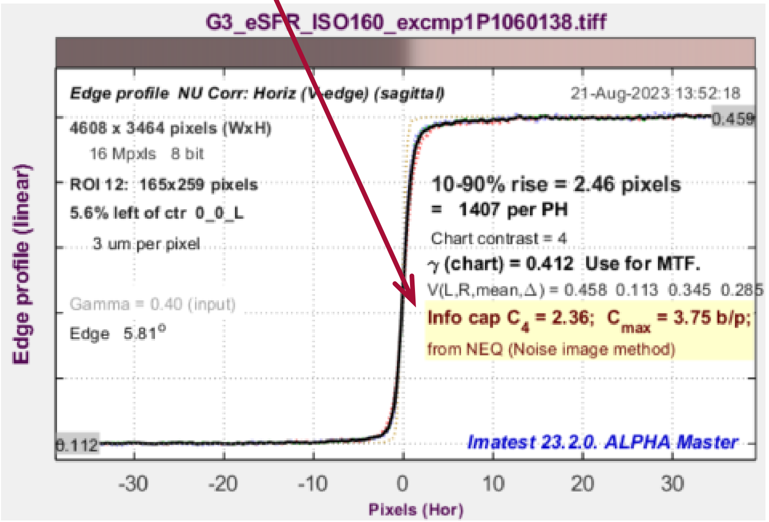
Mean(C_{max}) = 2.96 b/p.

Total info capacity $C_{maxTotal}$ = mean(C_{max}) * number of pixels = 47.23 Mb



C_4 and C_{max} are displayed in the Edge/MTF plot.

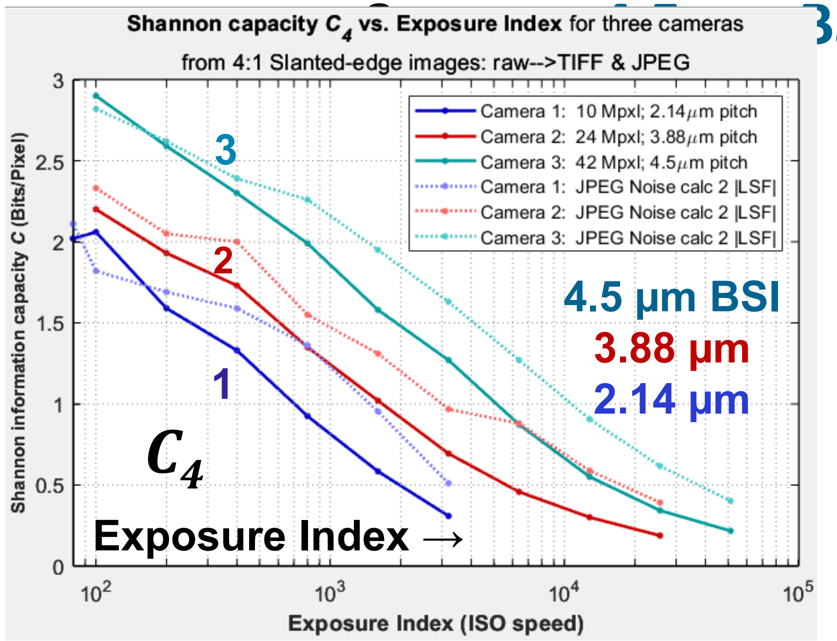
Information capacities
C_4 = 2.36 b/p; C_{max} = 3.75 b/p.

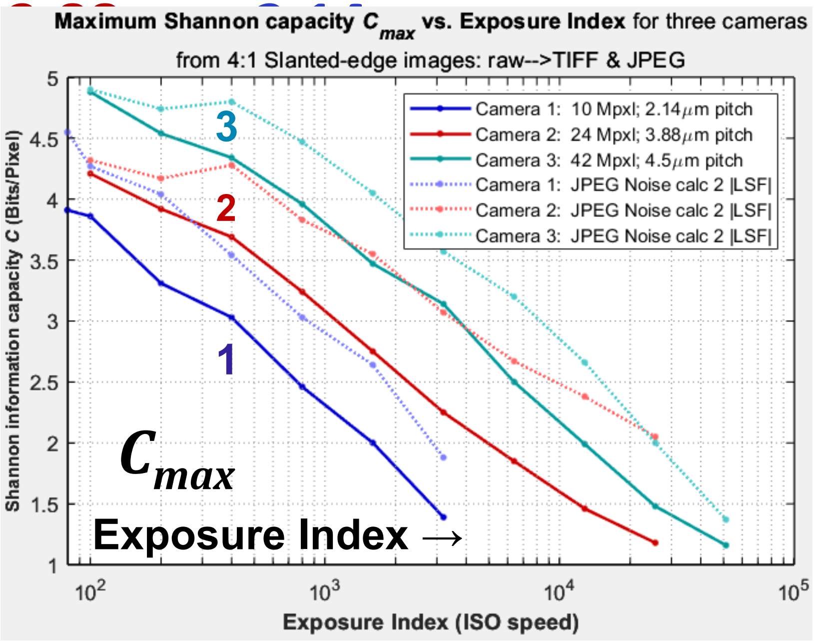


C_4 and C_{max} results for three cameras

In auto-exposure cameras that keep image Digital Numbers (DNs) constant, Exposure Index (EI) (sometimes called ISO speed) is proportional to analog gain, Hence **illumination and SNR decreases as Exposure Index (EI) increases**.

C_4 and C_{max} decrease with Exposure Index (EI).
C_4 and C_{max} increase with pixel size, as expected.





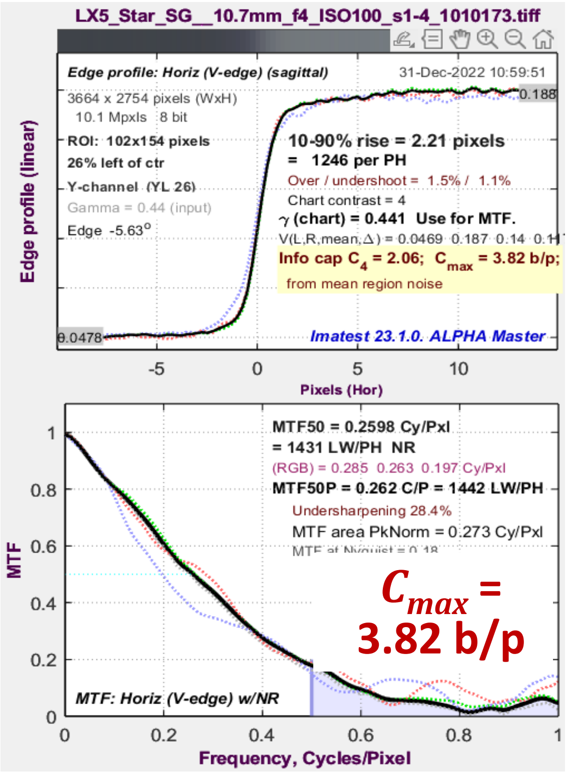
C_{max} is larger than C_4 by roughly 2 bits/pixel.

imatest®

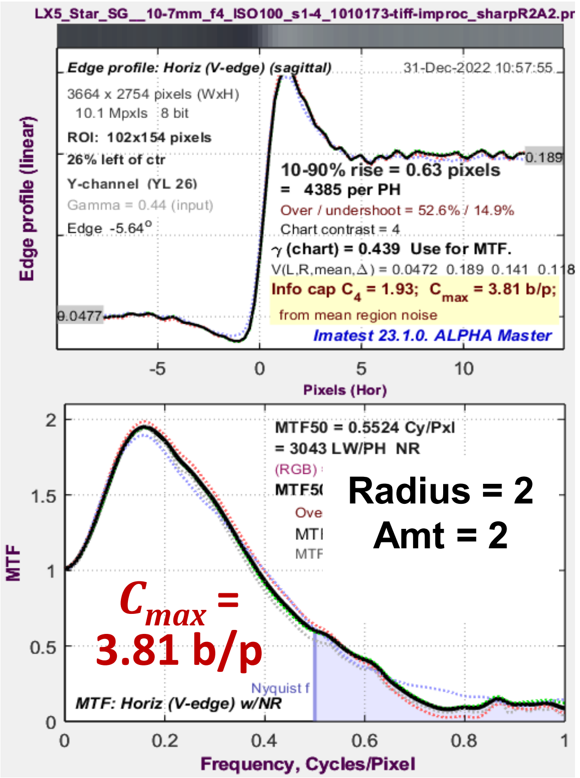
Sharpening and information capacity

Uniform Sharpening has little effect on C because it boosts the high frequency signal and noise by the same amount.

Minimally processed TIFF



USM-sharpened TIFF



C is unaffected by linear, reversable image processing.

For this reason, it is not useful for finding optimum image processing.

The image information metrics, to be described in the following slides (especially *SNRi* and *Edge Location σ*), serve this purpose because

- They are sensitive to image processing, and
- they measure how well objects and edges are detected.

imatest®

The key image information metrics

are derived from *MTF* and the Noise Power (Wiener) spectrum of the *noise image*.

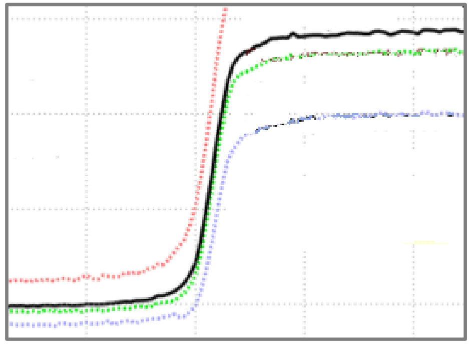
To obtain the noise image

Note that the oversampled image consists of four averaged interleaves from the original bins of the ISO 12233 calculation.

De-bin the image by moving the low-noise contents of each interleave back to their original locations.

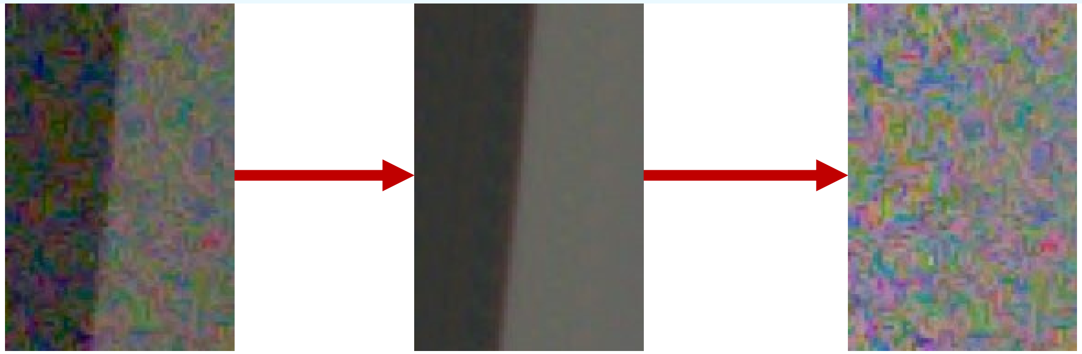
The de-binned image (2) has much lower noise than the original (1).

Noise image (3) = original image (1) – de-binned image (2).



Micro 4/3 camera @ ISO 12800

(1) Original	(2) De-binned	(3) Noise image = original – de-binned



The noise (which has a mean of 0) image is shown lightened.

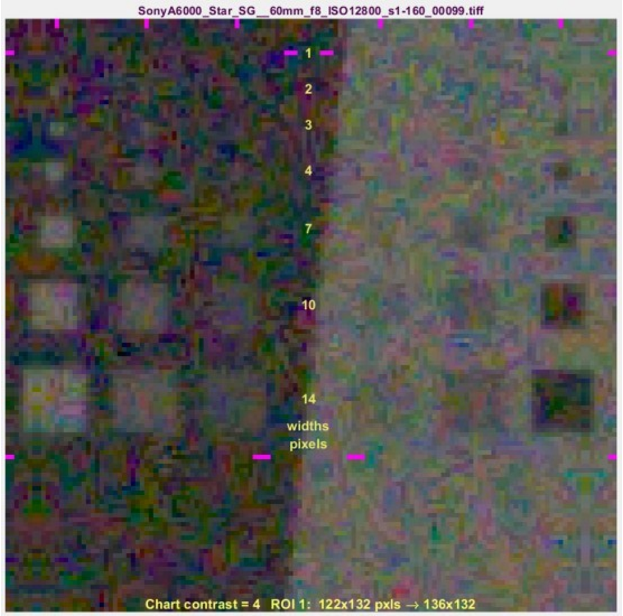
imatest®

Key image information metrics are derived from the noise image

- ***Noise Power Spectrum (NPS)***
- ***Noise Equivalent Quanta (NEQ)*** — a frequency-dependent SNR used in metical imaging
- ***Information capacity, C_{NEQ}, derived from NEQ***
- ***Ideal observer Signal-to-Noise Ratio (SNRi)*** — detectability of small objects (whether it is present).
- ***Edge SNRi & Edge Location σ (standard deviation)*** — accuracy of object ***location*** (shape and position).

Additional metrics (will not be covered in detail)

- *Detective Quantum Efficiency (DQE) — derived from NEQ*
- *Noise Autocorrelation — may indicate sensor crosstalk*
- *Object visibility* — of small/low contrast objects, shown on the right. Derived from *SNRi*.



Noise Power (Wiener) Spectrum $NPS(f)$

The 2D Fourier Transform (FFT) of the noise image must be transformed into 1D.

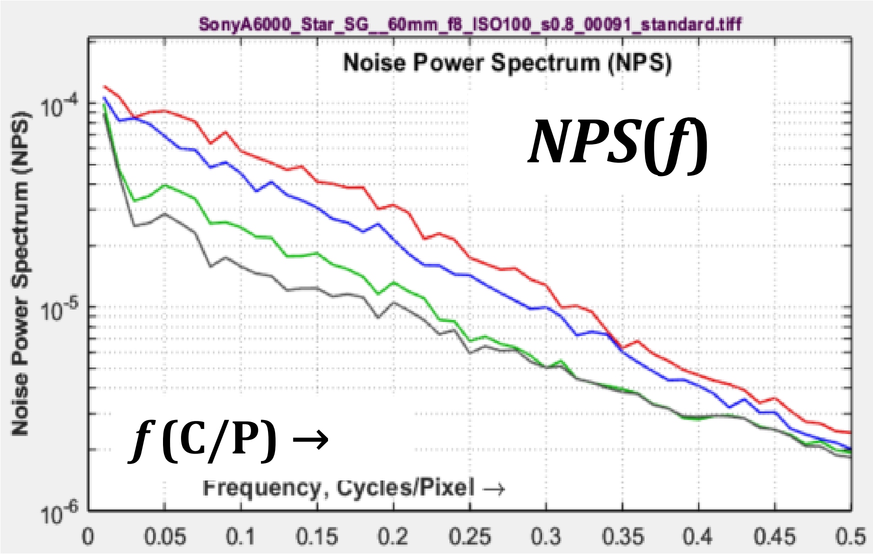
- Noting that $f = 0$ at the center of the 2D FFT image (from MATLAB fft2 and fftshift), divide it into several annular regions, and find the average noise power for each region.
- Because this procedure does not maintain the invariance in energy between the spatial and frequency domains implied by Parseval's theorem,

Normalize $NPS(f)$ so that $\int NPS(f)\, df = \int \sigma^2(x)\, dx = \int N(x)\, dx$

The noise amplitude (voltage) spectrum is

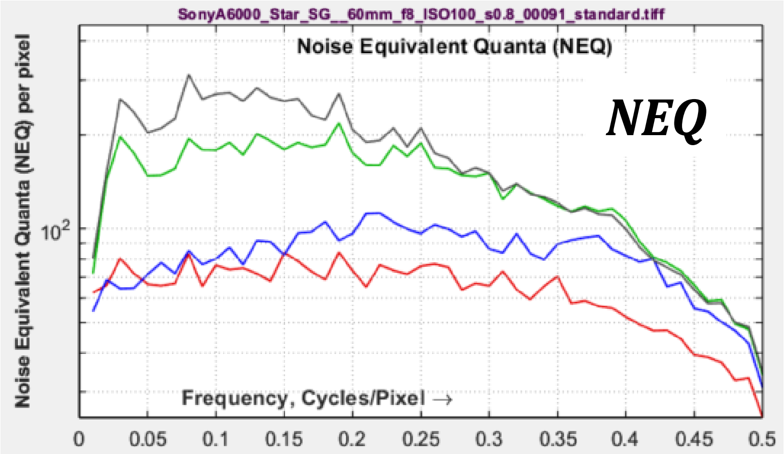
$$N_V(f) = \sqrt{NPS(f)}$$

$NPS(f)$ is a part of the kernel that defines image information metrics, $K(f) = MTF^2(f)/NPS(f)$.



imatest®

Noise Equivalent Quanta $NEQ(f)$



$NEQ(f)$ Frequency-dependent Signal-to-Noise (power) Ratio, equivalent to the number of quanta that would generate the measured SNR when photon shot noise is dominant. Used in medical imaging.

$$NEQ(f) = \frac{V_{mean}^2\, MTF^2(f)}{NPS(f)} = V_{mean}^2\, K(f)$$

$K(f) = MTF^2(f)/NPS(f)$ is the *kernel* (the defining factor) of the image information metrics to be introduced.

Because uniform filtering affects $MTF^2(f)$ and $NPS(f)$ identically, $NEQ(f)$ and $K(f)$ are not affected by uniform, reversable filtering such as sharpening or lowpass filtering.

Calculations derived from $NEQ(f)$

An information capacity, C_{NEQ}, can be calculated from $NEQ(f)$ by substituting $V_{P-P}/\sqrt{12}$ (for a uniform distribution) for V_{mean}.

$$C_{NEQ} = \int_0^{f_{Nyq}} log_2(1 + NEQ_{info}(f))\, df$$

C_{NEQ} can be thought of as a summary metric for $NEQ(f)$.

Results are close to C from edge variance; they differ because C_{NEQ} includes the noise spectrum.

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

Detective Quantum Efficiency, *DQE(f)*, is the ratio of $NEQ(f)$ (the number of quanta equivalent to the measured SNR) to the mean number of incident quanta. It has maximum value of 1.

$$DQE(f) = \frac{NEQ(f)}{\bar{q}}$$

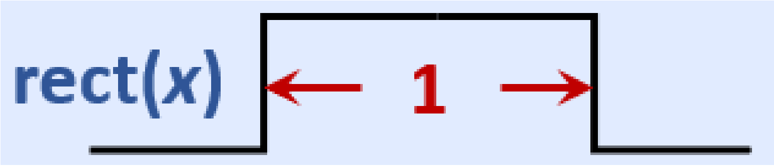
Under development.

imatest®

Ideal Observer Signal-to-Noise Ratio *SNRi*

SNRi* is metric for the detectability of *objects, calculated for w × kw rectangles.

For $\Delta g(x,y) = \Delta Q \cdot \mathrm{rect}(x/w) \cdot \mathrm{rect}(y/kw)$,

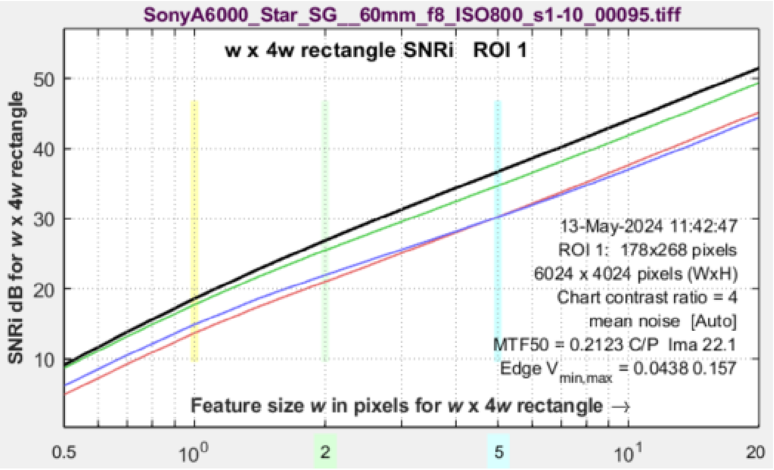


The Fourier transform of $\Delta g(x,y)$ is

$$FFT\big(\Delta g(x,y)\big) = G\big(f_x, f_y\big) = kw^2 \Delta Q \frac{\sin(\pi w f_x)}{\pi w f_x} \frac{\sin\big(\pi k w f_y\big)}{\pi k w f_y}$$

$$SNRi^2 = \int_0^{f_{yNyq}} \int_0^{f_{xNyq}} |G\big(f_x, f_y\big)|^2 \boldsymbol{K(f)} \; df_x \, df_y \;\; \text{where} \;\; f = \sqrt{f_x^2 + f_y^2}$$

Rescued by Paul Kane from [ICRU Report 54] (an obscure medical imaging document that correlates *SNRi* with Bayesian detection statistics).



***Feature size w* in pixels →**

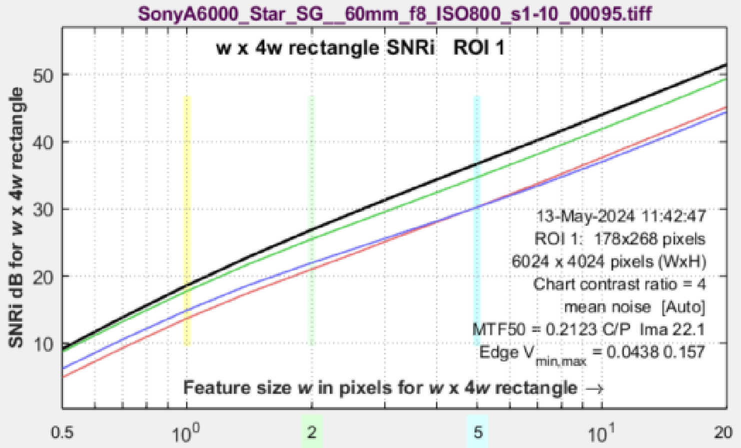
In spatial domain, *SNRi*² is the total energy of the object S/N: related to *object visibility*.

***SNRi* is proportional to the Michelson contrast of the chart ((*lt*–*dk*)/(*lt*+*dk*))** (0.6 for 4:1 contrast ratio).

***SNRi* plots can be difficult to interpret because they strongly increase with *w*.**

SNRi per pixel
a better way of displaying *SNRi*

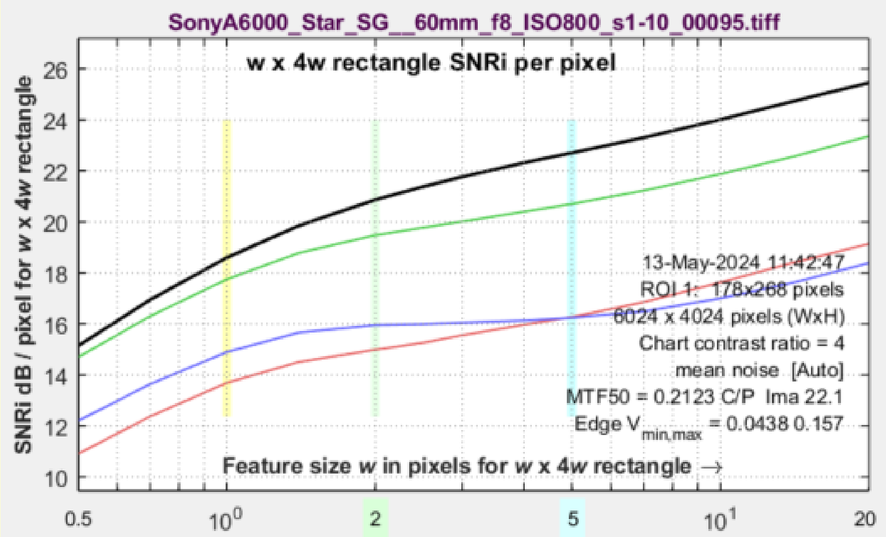
SNRi **— the metric for the detectability of *objects* — is difficult to interpret because it increases with object size.**



***Feature size w* in pixels →**

Note that the y-axis scales are very different.

SNRi **in units of SNR per distance (pixels) is easier to grasp because it approaches a limit.**



***Feature size w* in pixels →**

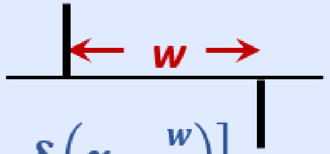
We expect ***SNRi*** to be predictive of the key machine vision performance metric, [***Mean Average Precision, mAP***].

imatest®

Edge SNRi

Edge SNRi, **is new metric for the**
detectability of edge location or object shape.

Odd impulse pair



$$I_I(x/w) = \frac{1}{2}\left[\delta\left(x+\frac{w}{2}\right) - \delta\left(x-\frac{w}{2}\right)\right]$$

Similar to *SNRi*, with the object replaced by the edges (the gradient of the object), which forms ***Line Spread Function doublets*** (pairs opposite-polarity δ-functions spaced by *w*).

$$\Delta h(x,y) = V_{P-P} \cdot I_I(x/w) \cdot I_I(y/kw);$$

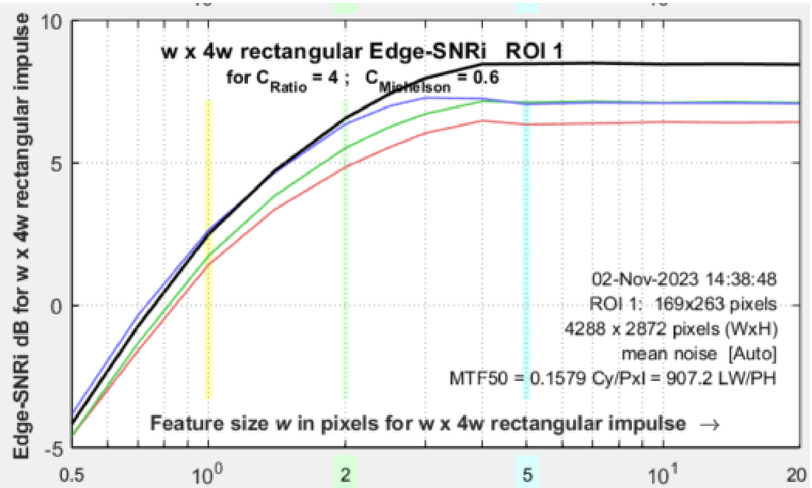
$$FFT\big(\Delta h(x,y)\big) = H\big(f_x, f_y\big) = \pi^2 f_x\, f_y\, G\big(f_x, f_y\big) = 2\, V_{P-P} \sin(\pi w f_x) \sin\big(\pi k w f_y\big)$$

$$Edge\ SNRi^2 = \iint \left|H\big(f_x, f_y\big)\right|^2 \boldsymbol{K}(\boldsymbol{f})\ df_x\, df_y$$

Edge Location σ, **derived from** ***Edge SNRi***, **is our preferred metric for evaluating system performance (next slide).**

In spatial domain, ***Edge SNRi***^{2} **is the energy of the LSF doublets.**

Affected by filtering (ISP).



Feature size w **in pixels →**

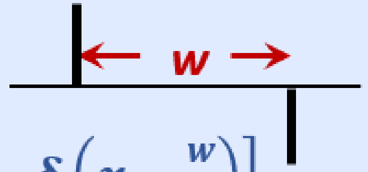
imatest®

Edge Location σ

***Edge Location Standard Deviation* (*σ*) is metric for the detectability of *edge location or object shape*. Lower is better.**

$$\boldsymbol{Edge\ Location\ \sigma} = \frac{\mathbf{1}}{\boldsymbol{Edge\ SNRi}}$$

Odd impulse pair

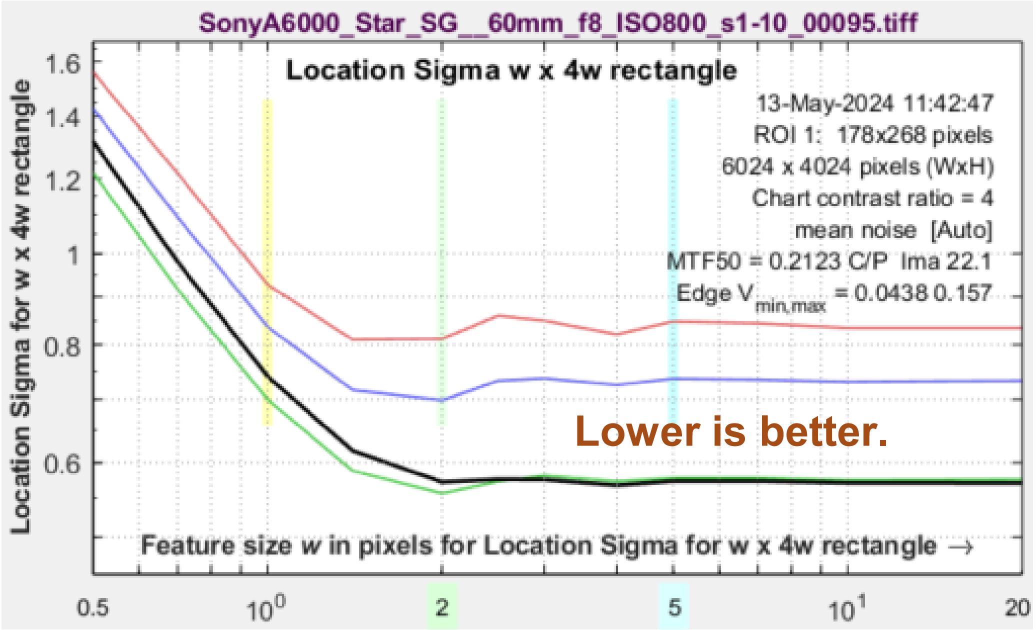


$$I_I(x/w) = \frac{1}{2}\left[\delta\left(x+\frac{w}{2}\right) - \delta\left(x-\frac{w}{2}\right)\right]$$

Edge Location σ has units of **pixels** (but can be converted to object distance, angle, etc.). **Affected by filtering (ISP).** Can be used to design matched filters to optimize location (shape) detection.

It is our preferred metric for evaluating system performance.

We expect it to be predictive of machine vision performance metric, ***IoU***.



***Feature size w* in pixels →**

imatest®

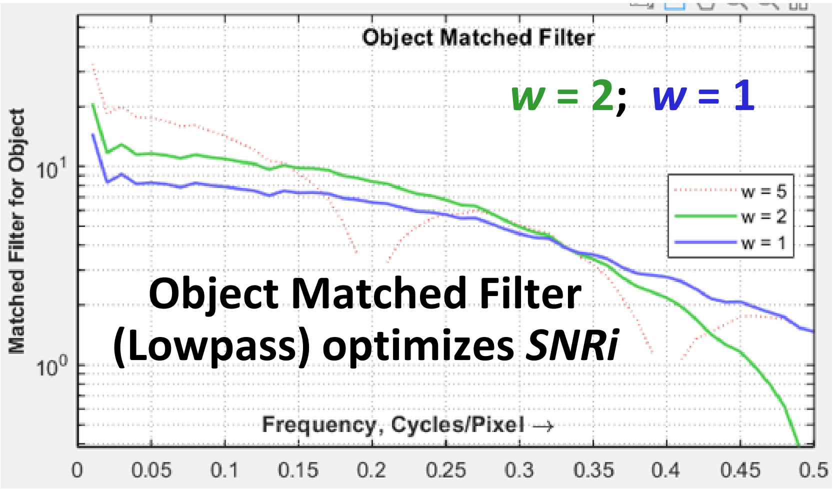
Optimum filtering: the matched filter

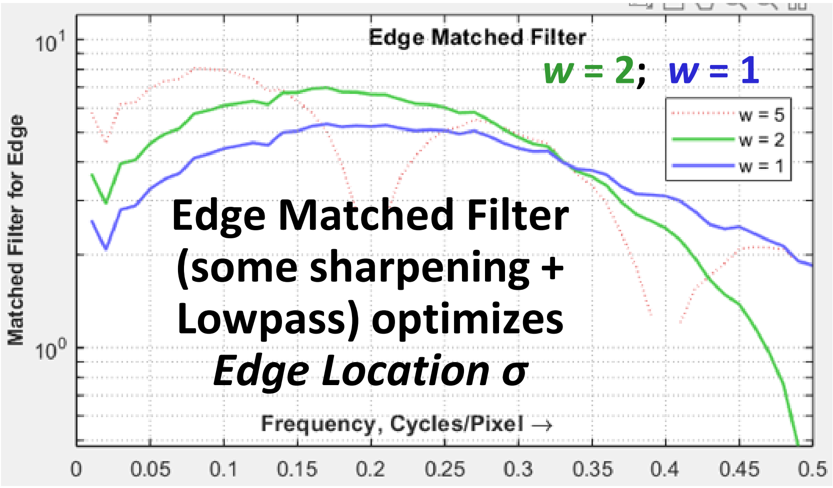
A custom filter that maximizes the object or edge detection performance for a specific system. Originally developed for radar. Described in [ICRU Report 54] (an obscure medical imaging document that connects SNRi with Bayesian detection statistics).

Matched filters optimize a single metric: *SNRi* or ***Edge Location σ*** for a specific object width ***w***.

If the matched filter transfer function (below) is known, it can be approximated by a lowpass filter (Bessel, Butterworth, etc.), and, if needed, sharpening filter. The filter must perform well for a variety of conditions, including interference from neighboring objects. This requires a tradeoff (not severe).

Best practices are needed for designing practical matched filters.





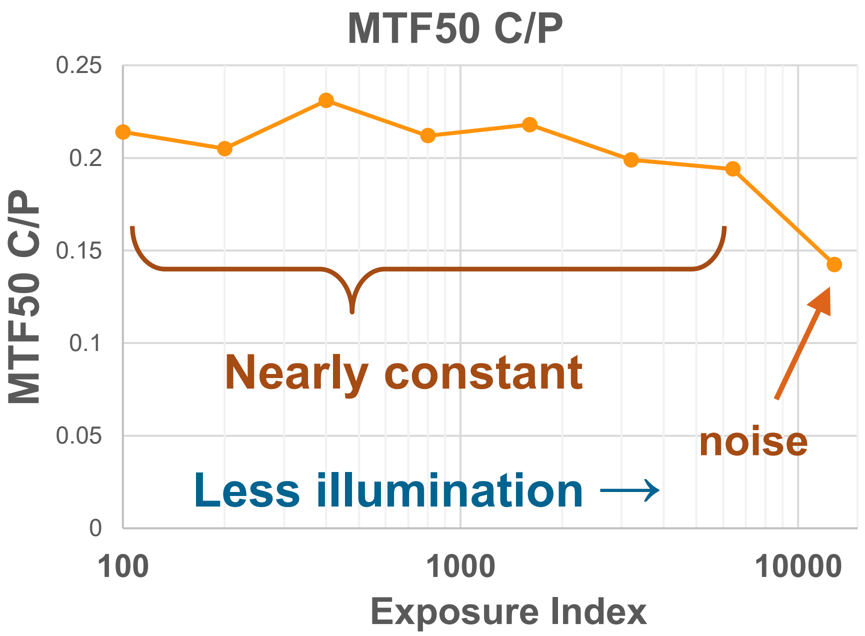
imatest®

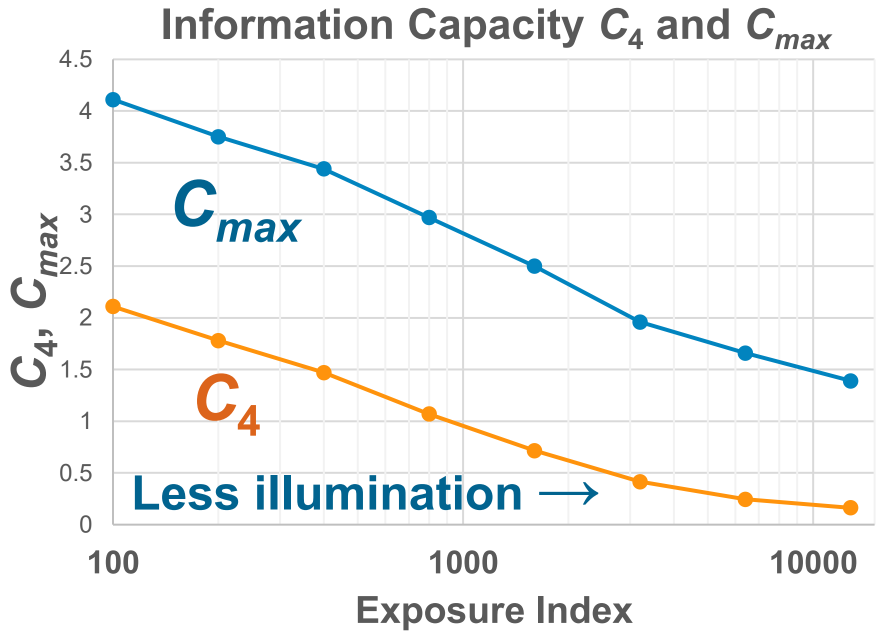
Example 1a: Exposure Index

24 MP Micro Four-Thirds mirrorless camera

Vary Exposure Index (EI; proportional to analog gain) from 100 to 12800.

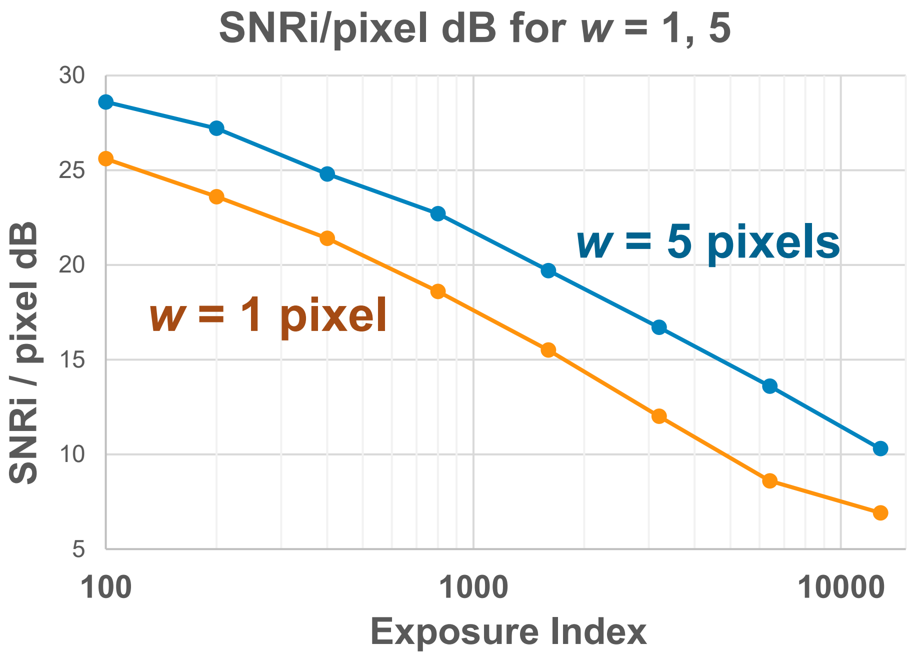
With auto-exposure, increasing EI decreases the light reaching the sensor, but keeps the image Digital Numbers (DNs) relatively constant.

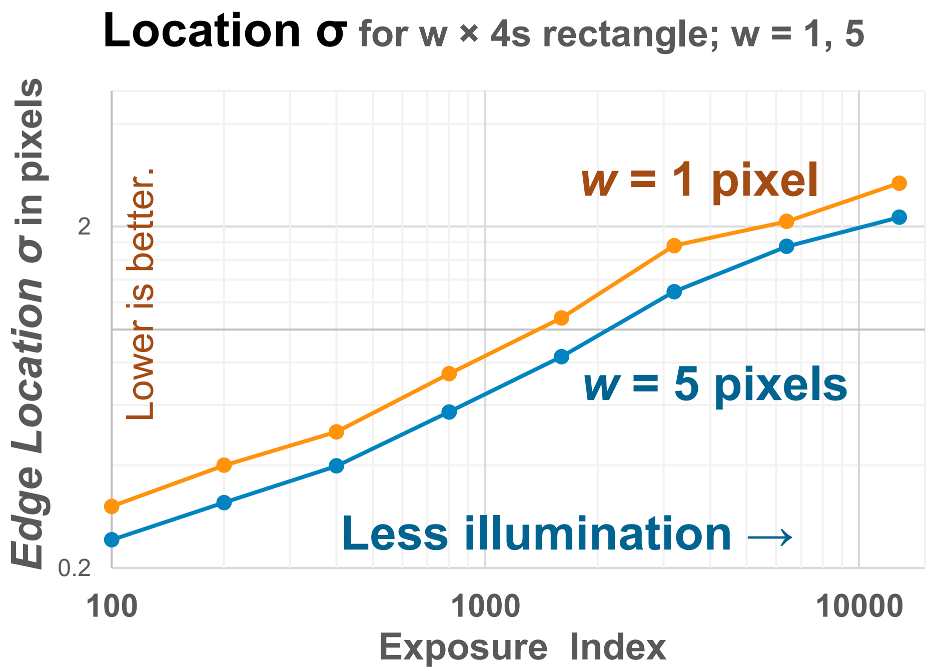




Example 1b: Exposure Index

24 MP Micro Four-Thirds mirrorless camera, EI 100-12800



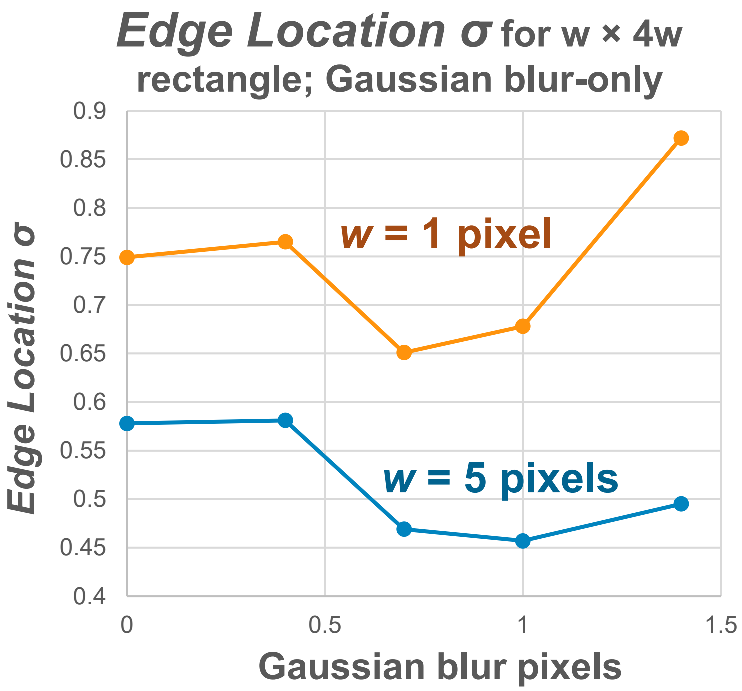


As expected, performance improves with more illumination (lower EI).

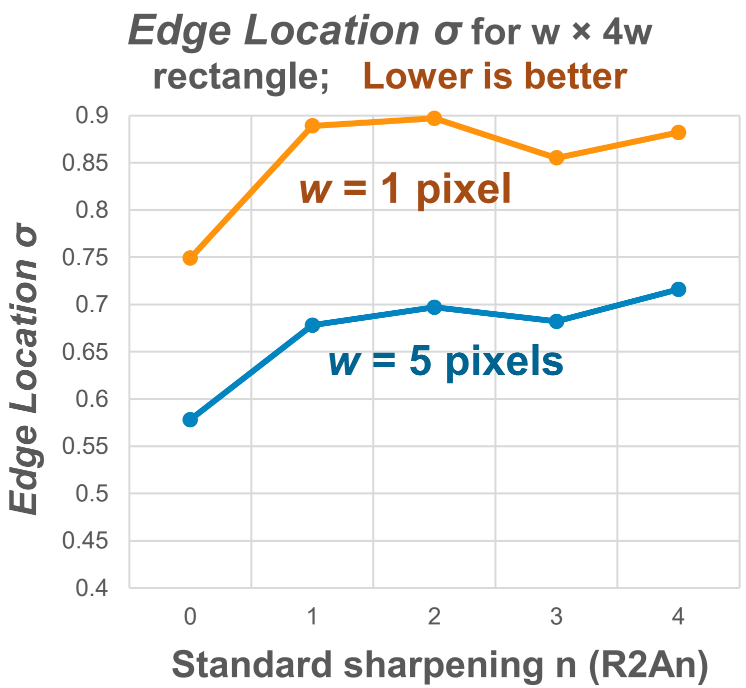
Example 2: Image processing

24 MP Micro Four-Thirds mirrorless camera, EI 800: ***Edge Location σ* (lower is better)**

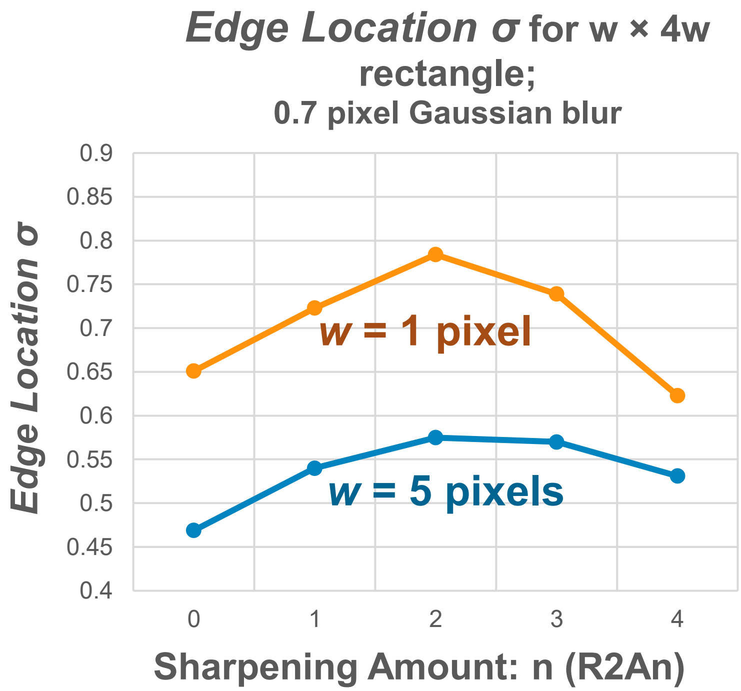
Gaussian blur (LPF)



Sharpening-only



LPF 0.7 + Sharpening



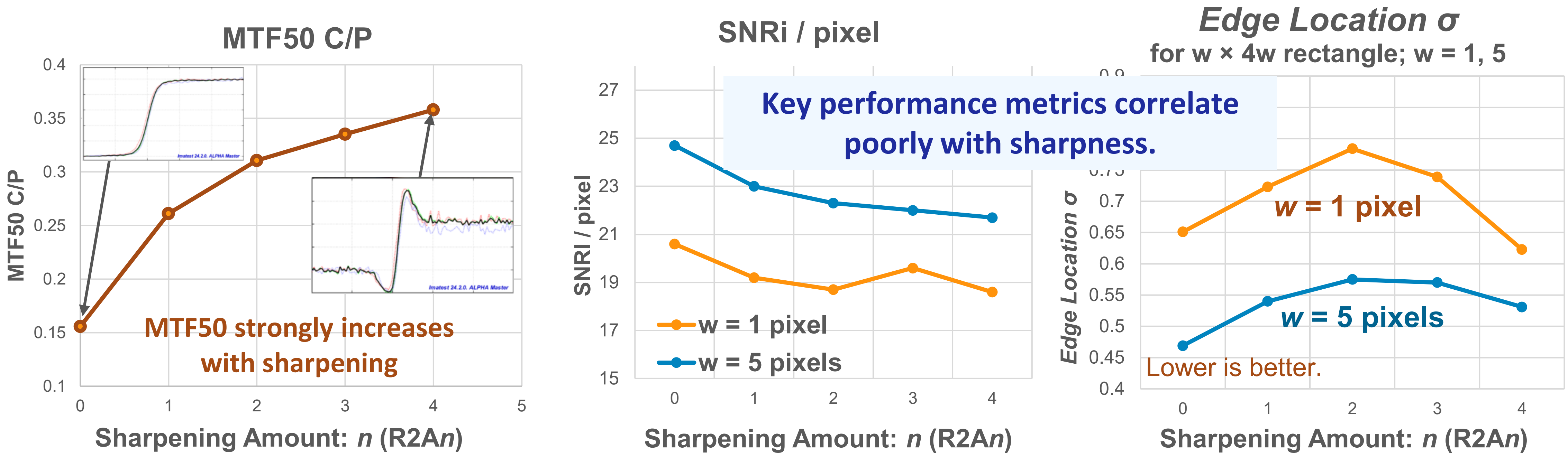
Lowpass filtering (Gaussian blur = 0.7 & 1) makes some improvement. Sharpening-only causes some degradation. LPF + Sharpening shows no clear trend.

The effects of image processing are not dramatic, perhaps because the original edge was very high quality.

Example 3: Sharpening

24 MP Micro Four-Thirds mirrorless camera, EI 800

0.7 pixel Gaussian blur + Sharpening



Key performance metrics *SNRi* and *Location σ* are poorly correlated with sharpness (MTF50, etc.).
They and may even *decrease*.

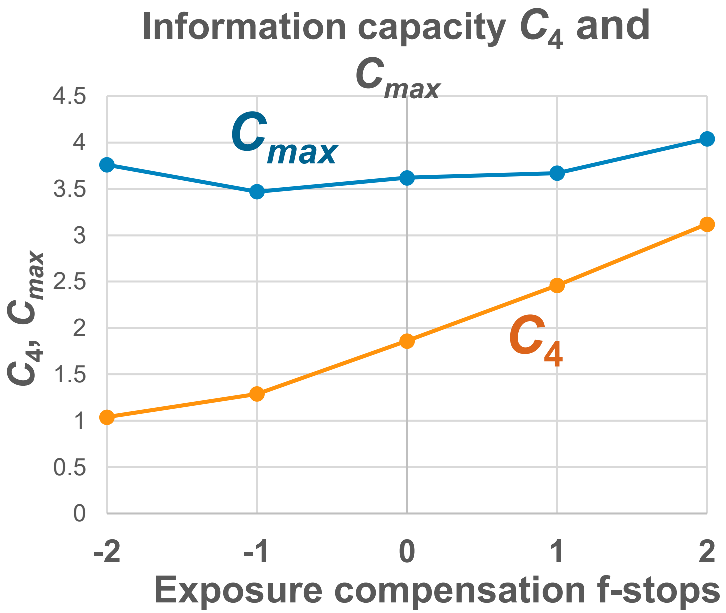
Sharpness metrics (MTF50, etc.) are *not* good indicators of system performance.
Extreme oversharpening, which boosts noise, should be avoided.

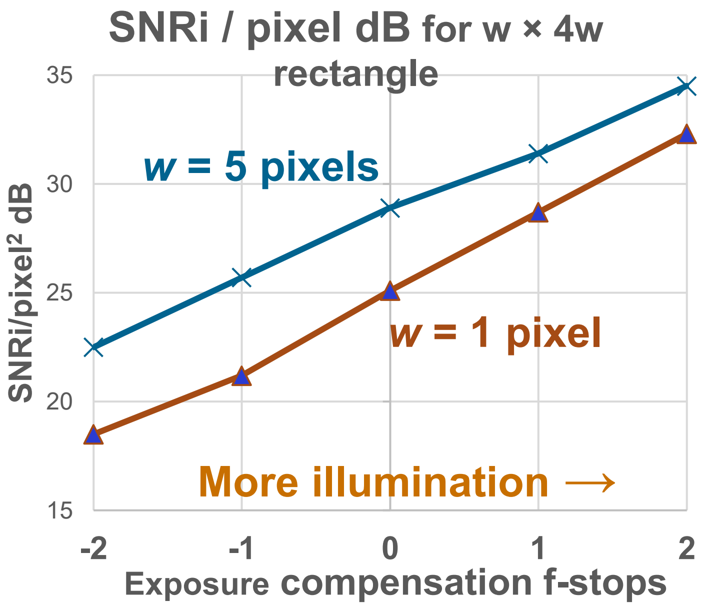
Example 4: Exposure compensation

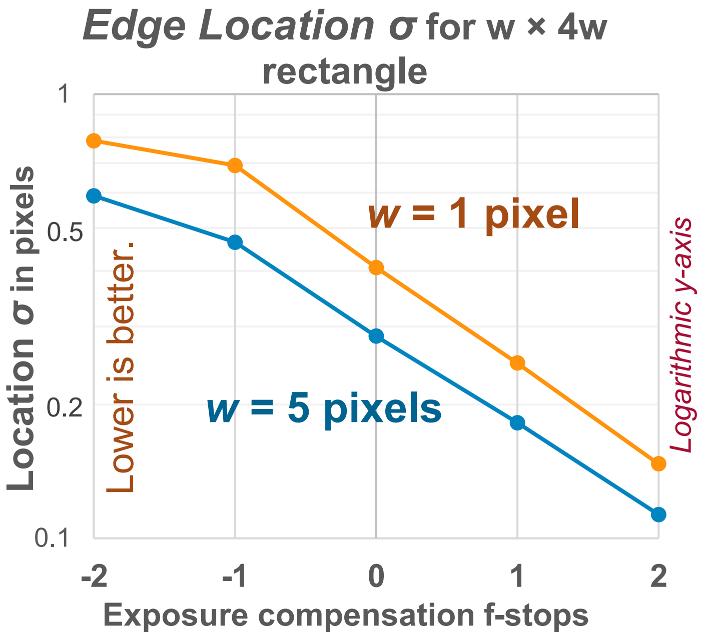
Lower is better.

16MP Micro Four-Thirds mirrorless camera. EI 160, f/5.6

Exposure compensation from -2 to 2 f-stops (dark to light).
Each step of 1 f-stop doubles the illumination, improving the performance.







Exposure compensation f-stops

-2

G3_eSFR_ISO160_excmp-2P1060134.tiff

0

G3_eSFR_ISO160_excmp0P1060137.tiff

+2

G3_eSFR_ISO160_excmp2P1060139.tiff

Exposure compensation from -2 to 2 f-stops (dark to light)

Summary

We have introduced new information-based image quality metrics, most importantly information capacity, *SNRi*, and *Location σ*, that are

- **Closely related to each other, sharing the *kernel*, $K(f) = MTF^2(f)/NPS(f)$,**
- **predict object and edge detection performance,**
- **Should be better than traditional sharpness and noise measurements for predicting Machine Vision system performance (mAP and IoU).**

Information capacity *C*, can be used to specify camera performance.

Once the required value of *C* has been determined, a camera can be selected with the *minimum* number of pixels needed to meet the requirement, and then image processing (filtering) can be designed.

This should

- **Maximize speed**
- **Minimize power consumption, and**
- **Minimize cost**

Notes

***Signal averaging* —** N identical images can be averaged to improve the consistency (Signal-to-Noise Ratio) of the results, which is improved by $\sqrt{N}$ (3 dB for every doubling of N). Noise is increased by $\sqrt{N}$ to keep results unchanged.

To do (a few of many)

- **Verify the correlation between image information metrics, especially $SNRi$ and $Edge\ Location\ \sigma$, and Machine Vision/Artificial Intelligence (MV/AI) performance metrics, such as Mean Average Precision (mAP) and Intersection over Union (IoU). Accuracy, speed, and power consumption are all critical.**
- **We look forward to working with researchers on this topic.**
  Grad students: There could be several PhD theses lurking here.
- Determine best practices for designing matched filters.

Thank you.

Documentation for image information metrics is linked from

Auto Sens

www.imatest.com/solutions/image-information-metrics/

Please visit *Imatest* at booth at AutoSens booth 223.