Correcting Misleading Image Quality Measurements

Norman Koren

Imatest LLC, Boulder, Colorado

Electronic Imaging 2020

Many image quality measurements, when accepted uncritically, can be highly misleading. Bad images can be interpreted as good.

- **Sharpness (MTF)**
- **Noise and SNR**
- **Dynamic Range**
- **Texture**
- **Color difference**

Each measurement will be reviewed, and some new material will be presented.

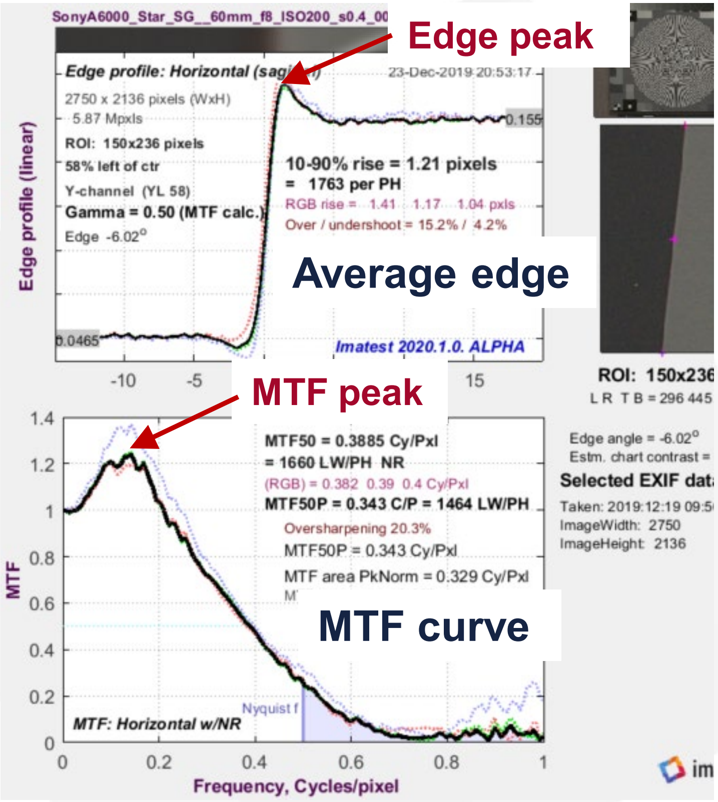
For Dynamic Range, misinterpretations can be especially damaging).

Key advice (which will be repeated): Look carefully at the image and make sure measurements correlate with what you see.

MTF – summary metrics

MTF **curves (right) can be fairly complex. Hence they are often summarized by a number called a "summary metric".**

- **MTF50**, the spatial frequency where MTF is 50% of its **low frequency** value (the most common summary metric),
- **MTF50P**, the spatial frequency where MTF is 50% of its **peak** value, and
- **MTF Area Normalized**, the area under the MTF curve (below f_{Nyq} = 0.5 C/P) normalized to a peak value of 1.

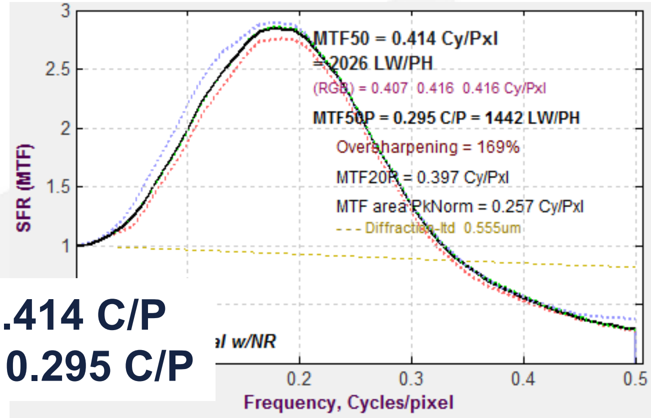


The most common summary metric, MTF50, has a serious drawback. It rewards excessive sharpening with high values.

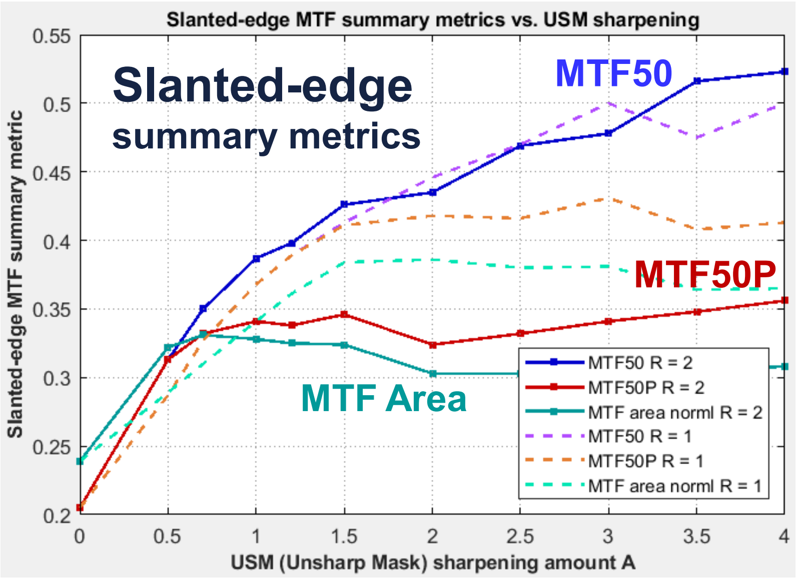
MTF summary metrics vs. sharpening

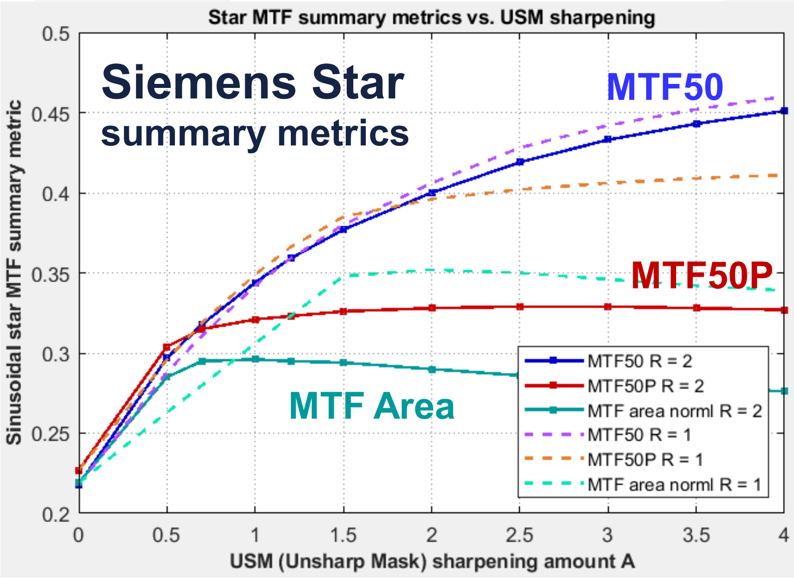
The most common summary metric, MTF50, increases with sharpening, while MTF50P and MTF Area remain relatively constant.

Severely overshar-pened image:



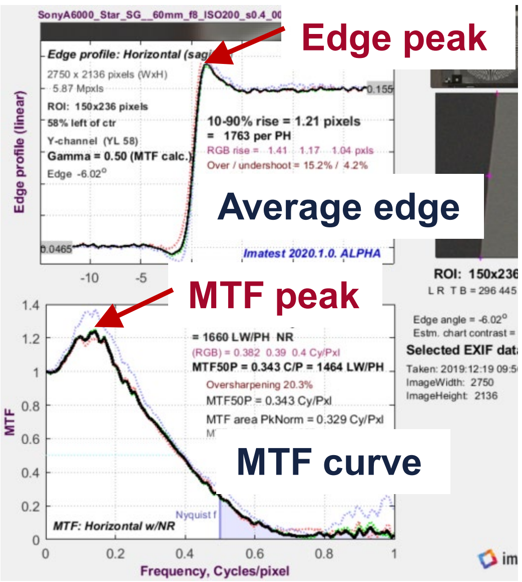
MTF50 = 0.414 C/P
MTF50P = 0.295 C/P

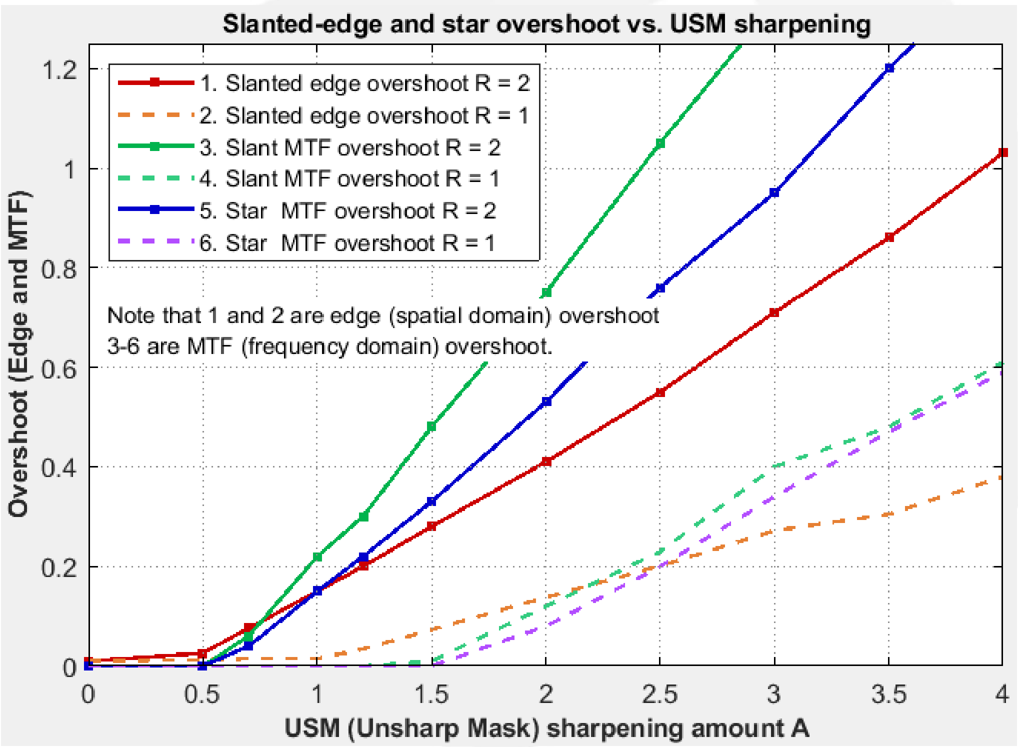




MTF overshoot – spatial & frequency

**Both spatial domain overshoot = $(P_{max} - P_{asymp}) / P_{asymp}$
and frequency domain overshoot = $(MTF(max) - MTF(0)) / MTF(0)$
increase when sharpening is increased beyond a threshold.**





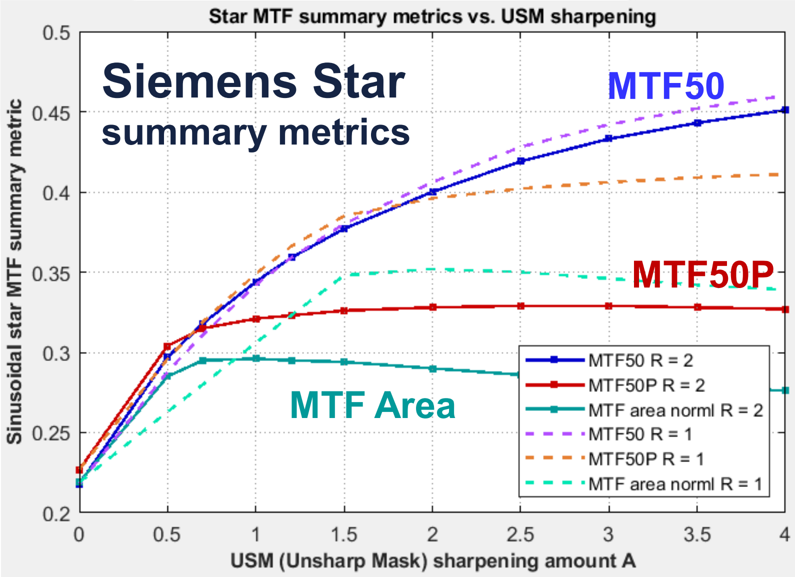
Overshoots are integral to sys-tem performance; hence must be reported along with MTF50P.

MTF recommendations

Avoid **MTF50**, especially for processed images from cameras. It can be highly misleading for strongly sharpened images.

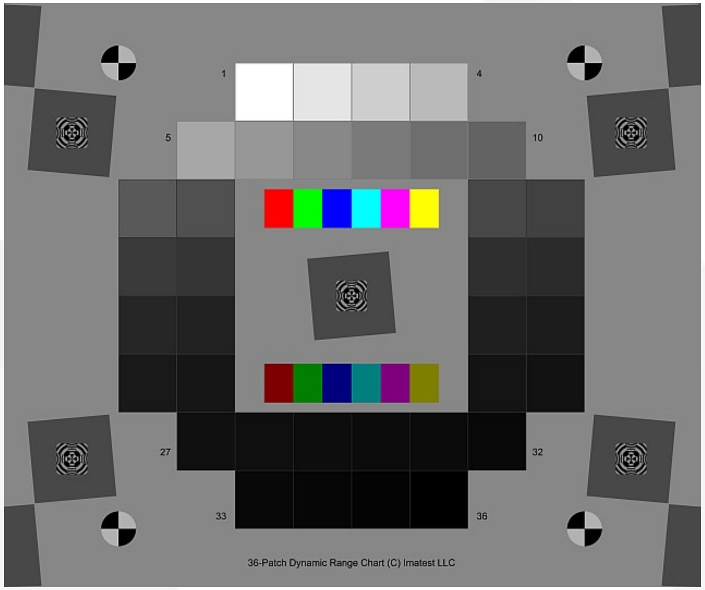
MTF50P is recommended because it is far more stable in the presence of strong sharpening.

Include **overshoot** (either spatial or frequency domain) when reporting on processed images.



Noise and SNR

Noise and SNR (Signal-to-Noise Ratio) measurements are usually made in flat patches of test charts.



In JPEG images from cameras, these patches are often smoothed (noise-reduced) by bilateral filters (which leave edges sharp). Noise reduction is often greater at high ISO speeds.

Raw images (demosaiced) will always give more accurate results.

Improved noise measurements can be made in the presence of a signal with a Siemens star using the technique in the accompanying paper, "Measuring camera Shannon Information Capacity…"
Unfortunately this does not work over a range of tones.

Recommendation: Be cautious of Noise/SNR measurements from camera JPEGs.

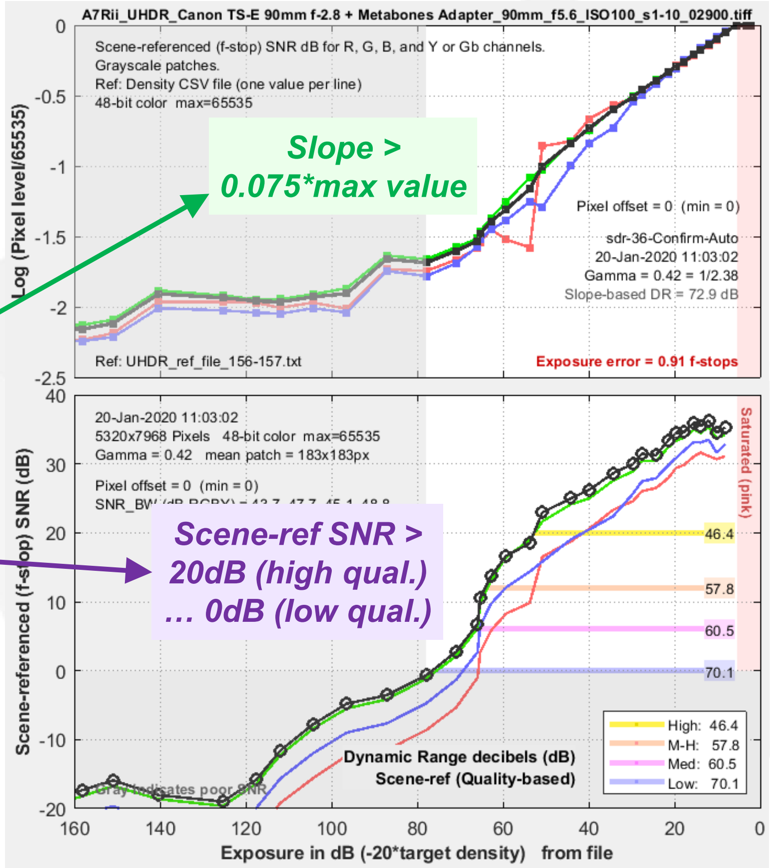
Dynamic Range

Dynamic Range (DR) is defined as the range of exposure (scene illumination) where the camera responds with

A. ***good contrast*** (the slope of log(pixel level) vs. log(exposure) must be > 0.075 of its max value), *and*

B. ***good Signal-to-Noise Ratio (SNR)*** (scene-referenced SNR must be greater than DR).

Both criteria must be met. Neither is sufficient by itself.



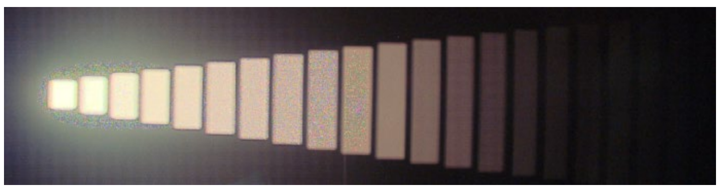
Flare light (stray light from lens surfaces) can affect DR measurements. When it is mistaken for a real signal (from the chart), the DR measurement can be exaggerated (and erroneous). We will show an example.

Dynamic Range issues

The problem with Dynamic Range measurements

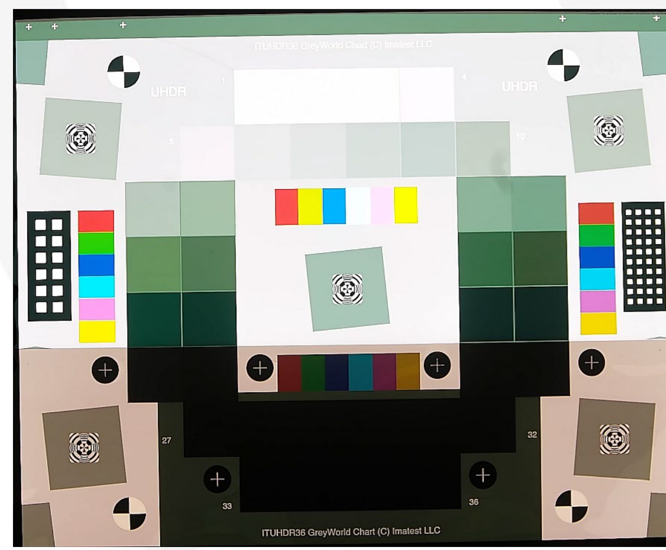
is that recent High Dynamic Range sensors have exceptional dynamic ranges: 120-150dB, and many engineers expect to measure similarly high numbers in cameras.

But such numbers cannot be achieved in cameras, which have a lens between the chart and sensor that causes ***flare light*** (stray light from lens surfaces) to diffuse from bright to dark areas of the image.



So engineers are tempted to cheat– to do anything to get the 120dB their marketing departments expect. Educating them is a tough job.

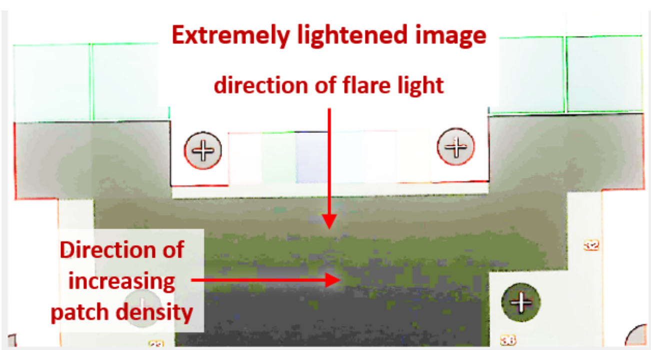
The image on the right has an issue with flare light (not obvious). It will be described in the next two slides.



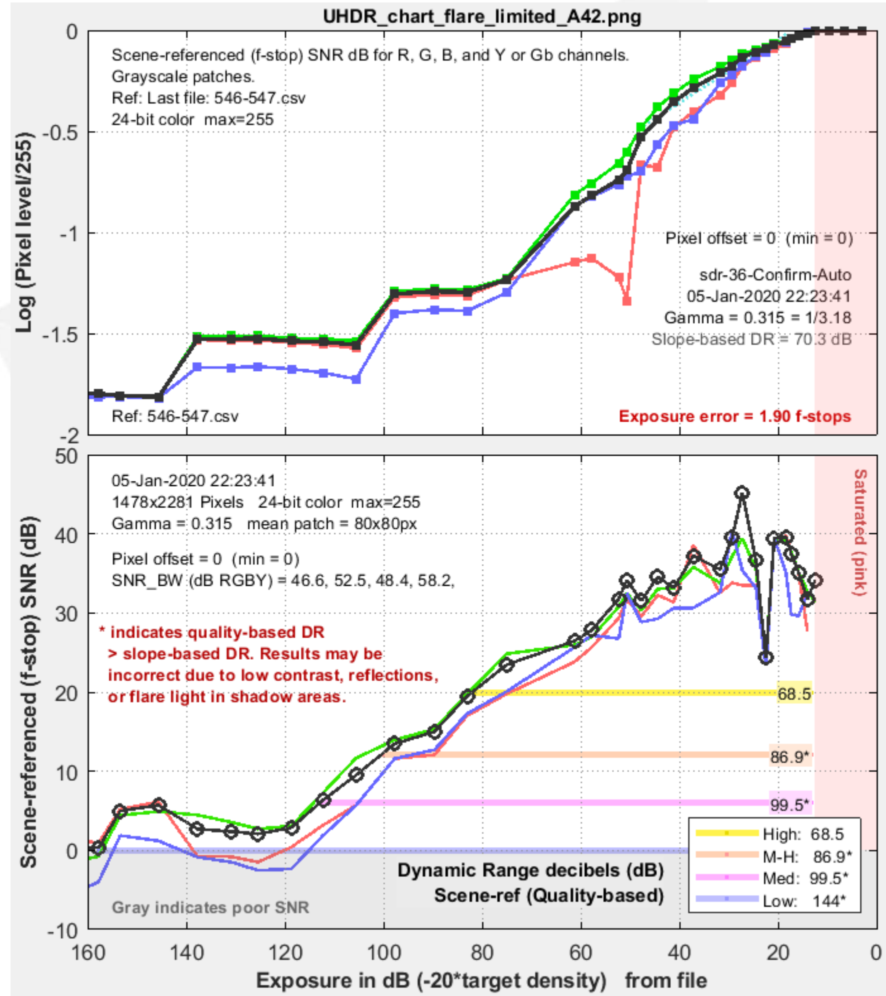
A misleading Dynamic Range measurement

The image on the previous page has significant flare light (not obvious). When the DR measurement is based on only scene-referenced SNR (ignoring slope, i.e., contrast), the measured DR is unreasonably high: **99.5dB for SNR = 6dB**; **144dB for SNR = 0dB**.

What is happening is the “signal” in the lower part of the chart is flare light diffusing from the top. Lightening the image,



Flare light was mistaken for real signal.



Fixing the misleading DR measurement

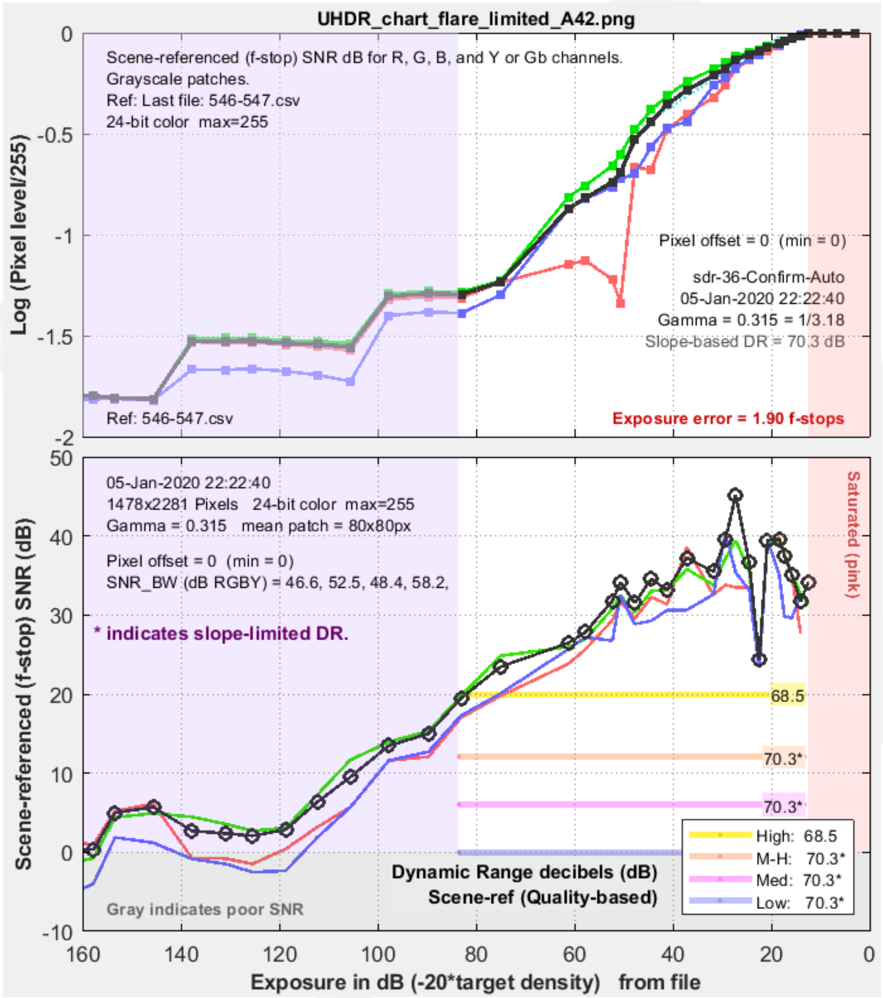
To obtain a correct measurement, DR is limited to tones where the slope (upper curve; log(pixel level) vs. log(exposure)) is greater than 0.075 of its maximum value.

Regions beyond this are shaded in blue.

DR, limited by the zone where slope drops below 0.075, is now **70.3dB**.

This only works in charts with a circular patch arrangement, not for linear charts.

We have seen cameras where the slope extends beyond the point where scene-referenced SNR is zero (i.e., where noise is so bad that there will be no detectable signal detail).



Dynamic Range Recommendations

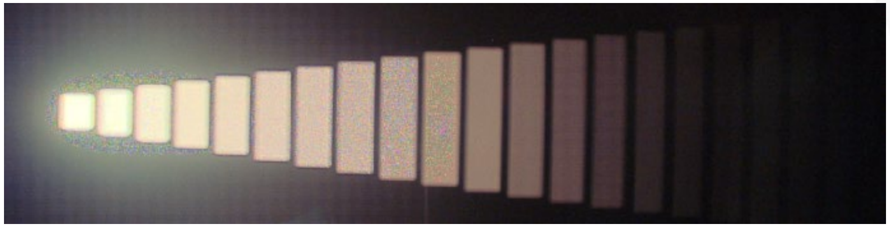
Returning to the definition of Dynamic Range (DR),

DR is defined as the range of exposure (scene illumination) where the camera responds with ***good contrast*** *and* ***good Signal-to-Noise Ratio (SNR)***.

Both criteria must be met. Neither is sufficient by itself.

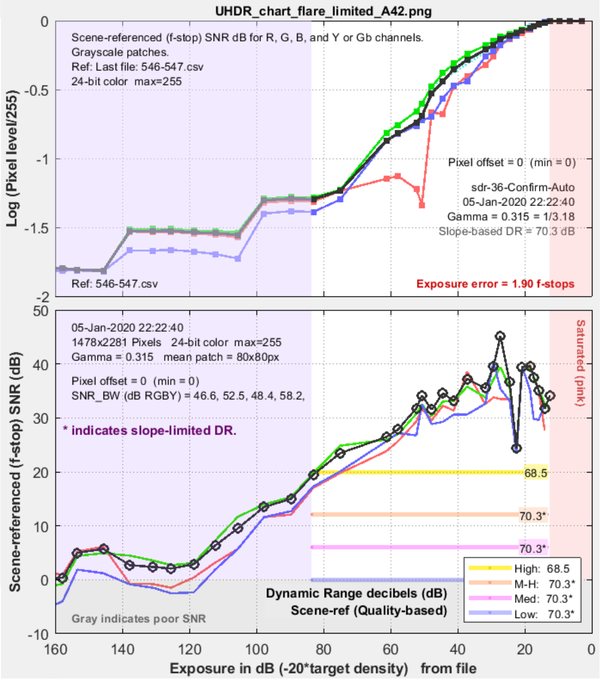
Flare light can cause erroneous measurements. It is especially hard to deal with in linear charts, where flare light and chart signal vary in the *same* direction.

Direction of decreasing flare light *and* chart signal →



Charts with circular patch arrangements are recommended because it is easy to distinguish flare from real patch signals, which vary in orthogonal directions.

Results from chart with circular patch arrangement



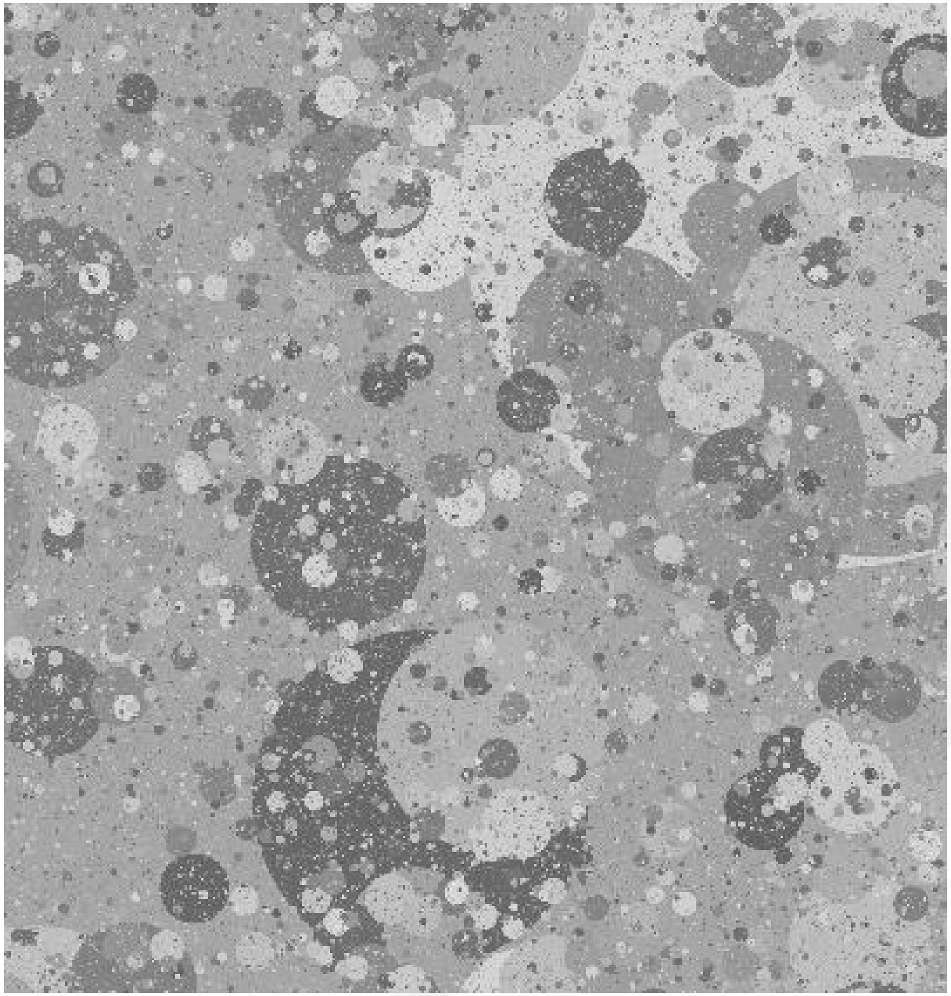
Texture – Dead Leaves 1

Texture is fine detail, often with medium to low contrast.

It needs to be measured with special patterns because it can be removed by bilateral filtering, which smooths (lowpass-filters) low contrast detail while leaving sharp edges untouched.

Bilateral filtering is common in JPEG images from consumer cameras and camera phones.

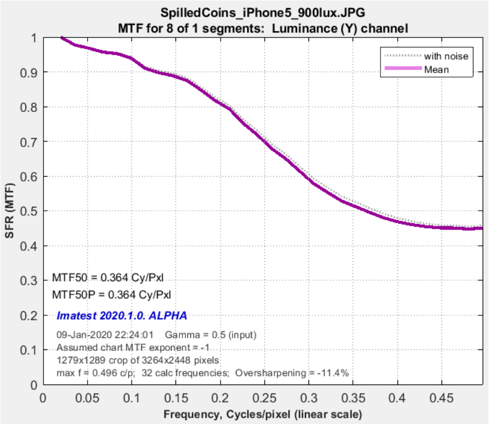
The Dead Leaves (Spilled coins) chart– original shown on the right– is scale-invariant and has a frequency spectrum similar to common scenes.

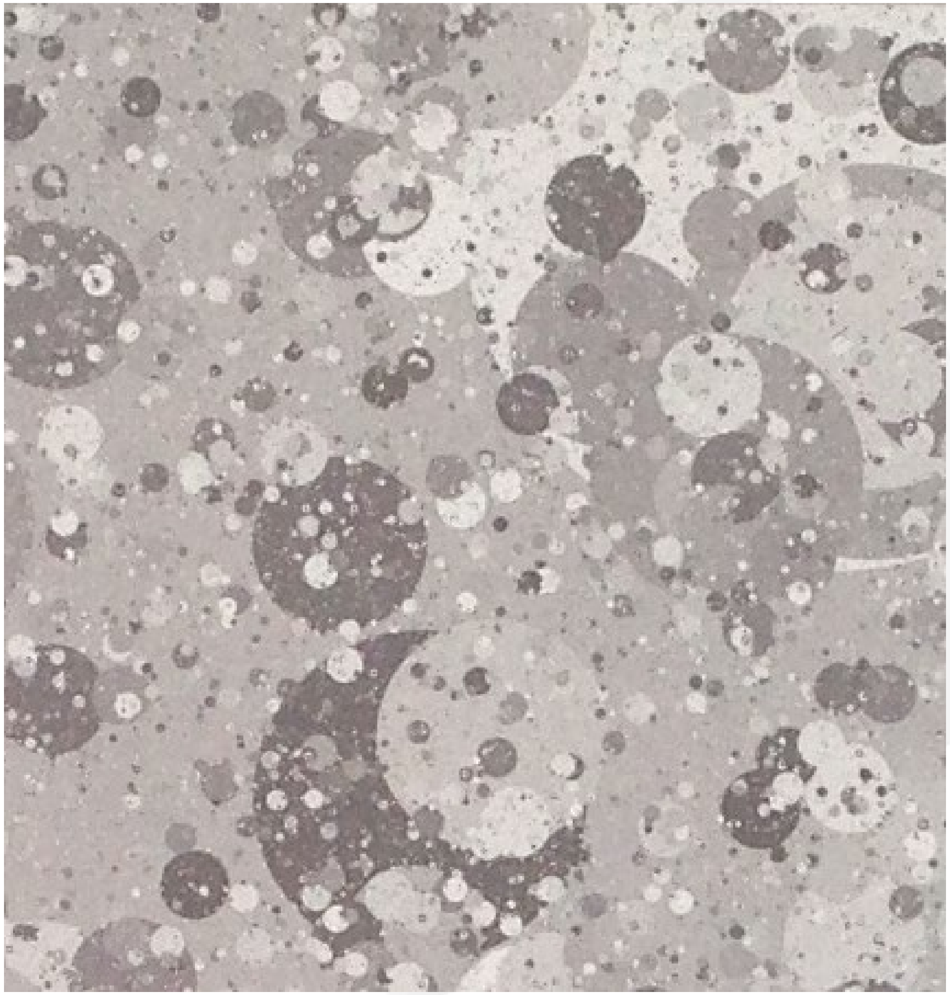


Texture – Dead Leaves 2

The iPhone 5 reproduces the chart very well (compare with the previous image). Some loss of detail is visible. It has very conservative image processing.

The MTF curve looks good. The response at high frequencies may be due to noise.

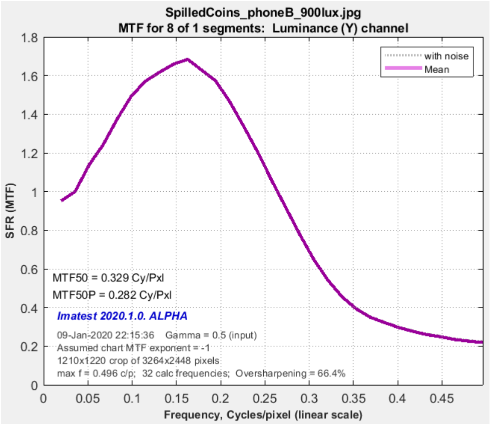


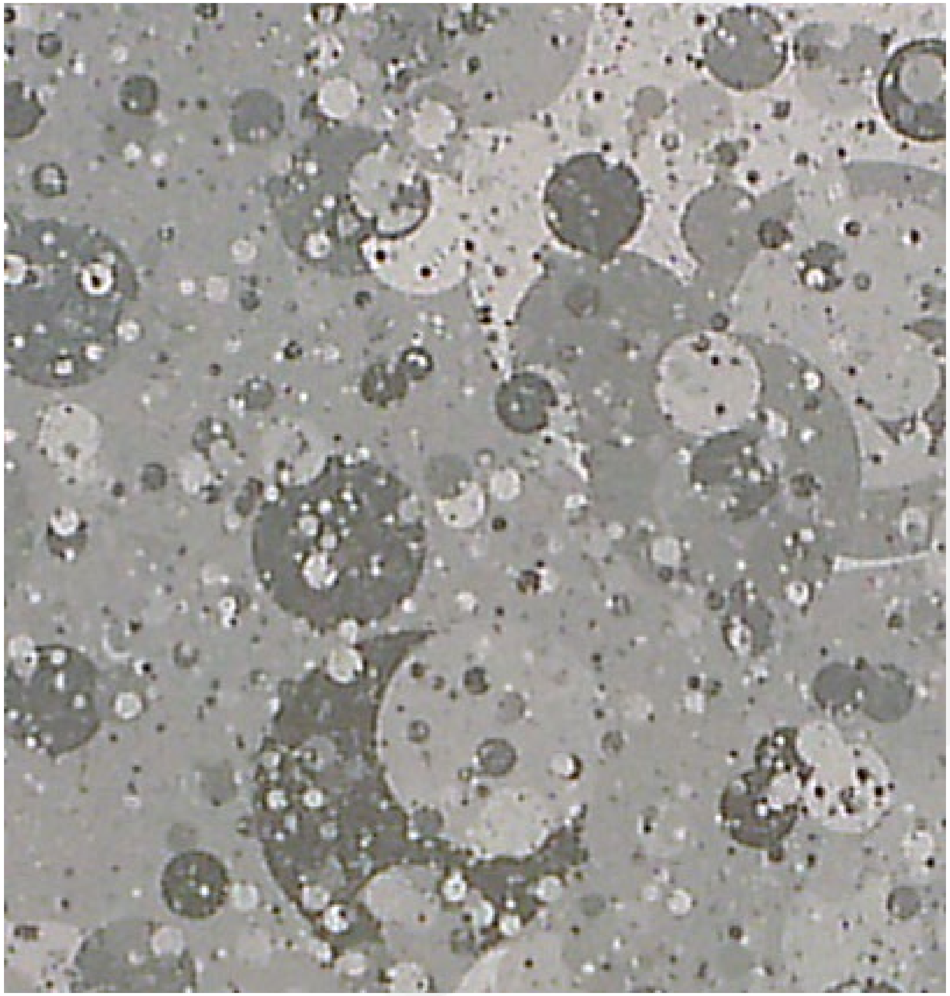


Texture – Dead Leaves 3

"Phone B" produces an ugly image, with sharpening "halos" near contrasty edges and severe loss of detail. The bilateral filter switches from extreme noise reduction to extreme sharpening near the maximum edge contrast (3:1).

The MTF curve shows the sharpening peak, but understates the contrast loss.





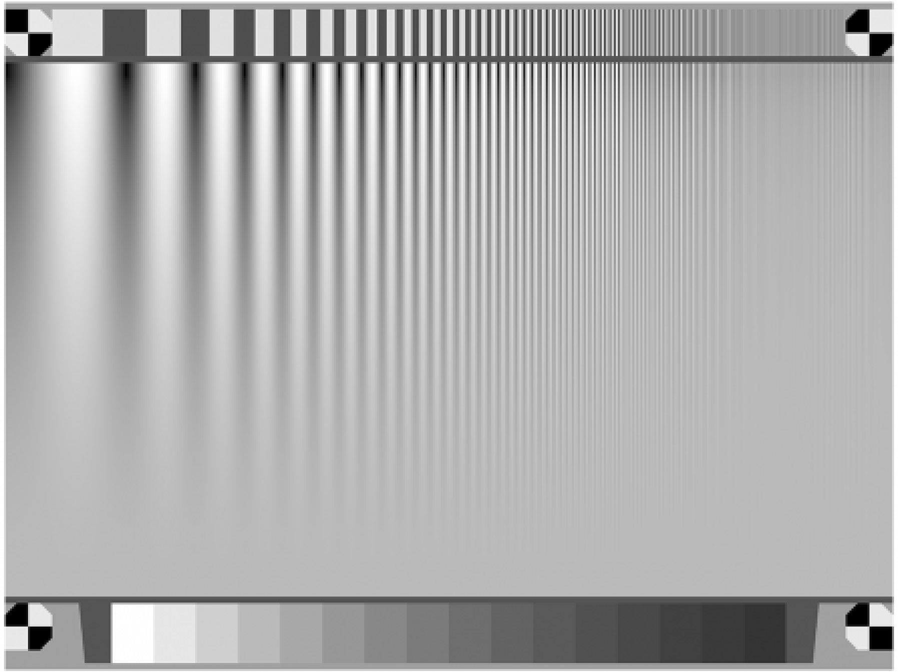
This is an unusual situation, but we must be alert to the possibility.

Texture – Log F-Contrast 1

The Log F-Contrast chart increases in spatial frequency along the x-axis and decreases in contrast from top to bottom along the y-axis.

It is sensitive to noise, and results definitely benefit from signal averaging.

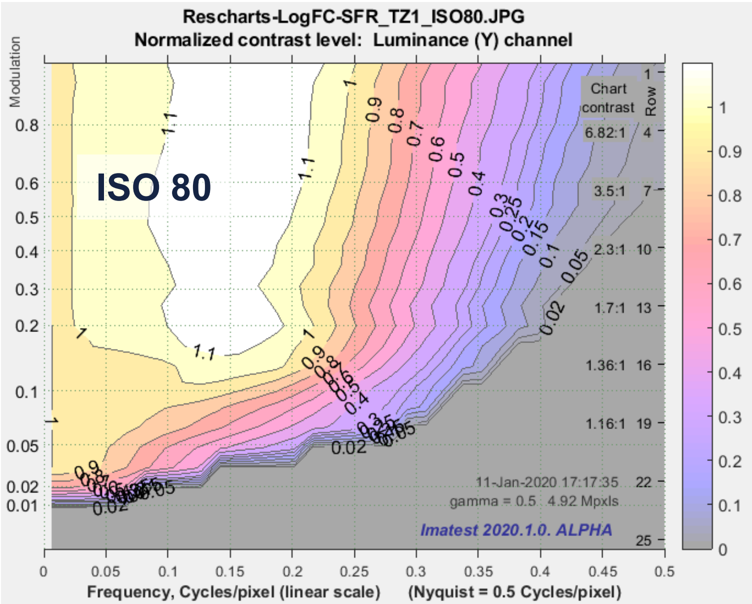
It provides detailed information about the dependence of texture on image contrast or modulation.



Texture – Log F-Contrast 2

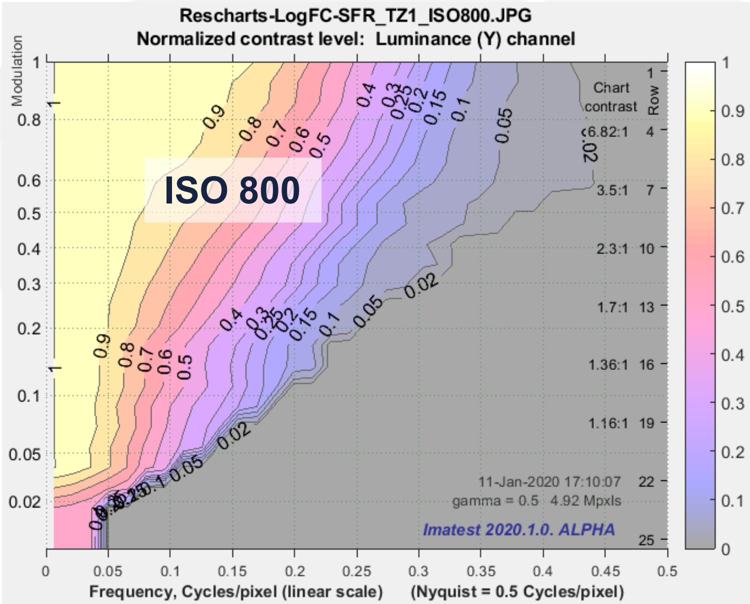
Log F-Contrast results for an older compact camera at ISO 80 and 800, where image processing is very different.

Moderate sharpening, noise reduction



Row	Modulation	Contrast	MTF50	MTF50P	MTF20	MTF20P	MTF10	MTF10P	Peak
1	0.995	>40:1	0.353	0.332	0.426	0.416	0.464	0.457	1.2
4	0.800	6.82:1	0.34	0.319	0.413	0.4	0.45	0.443	1.18
7	0.590	3.5:1	0.332	0.312	0.397	0.385	0.429	0.422	1.2
10	0.413	2.3:1	0.314	0.302	0.376	0.367	0.41	0.403	1.16
13	0.250	1.7:1	0.293	0.282	0.353	0.345	0.385	0.379	1.17
16	0.128	1.36:1	0.273	0.265	0.329	0.327	0.373	0.37	1.07
19	0.063	1.16:1	0.199	0.199	0.298	0.298	0.356	0.356	1

No sharpening, strong noise reduction



Row	Modulation	Contrast	MTF50	MTF50P	MTF20	MTF20P	MTF10	MTF10P	Peak
1	0.999	>40:1	0.246	0.246	0.308	0.308	0.346	0.346	1
4	0.788	6.82:1	0.229	0.229	0.292	0.292	0.333	0.333	1
7	0.589	3.5:1	0.198	0.198	0.264	0.264	0.313	0.313	1
10	0.409	2.3:1	0.17	0.17	0.231	0.231	0.272	0.272	1
13	0.245	1.7:1	0.139	0.139	0.201	0.201	0.235	0.235	1
16	0.132	1.36:1	0.113	0.113	0.174	0.174	0.231	0.231	1
19	0.065	1.16:1	0.0965	0.0965	0.151	0.151			1.01

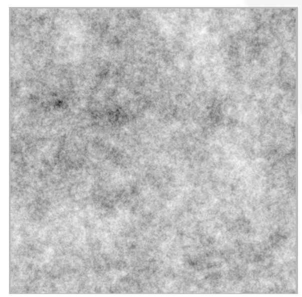
Texture – Summary

Summary

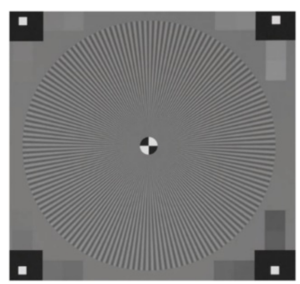
Dead leaves is standard, but results may get confused by unusual image processing.

Log F-Contrast produces an excellent picture of how image processing varies with contrast.

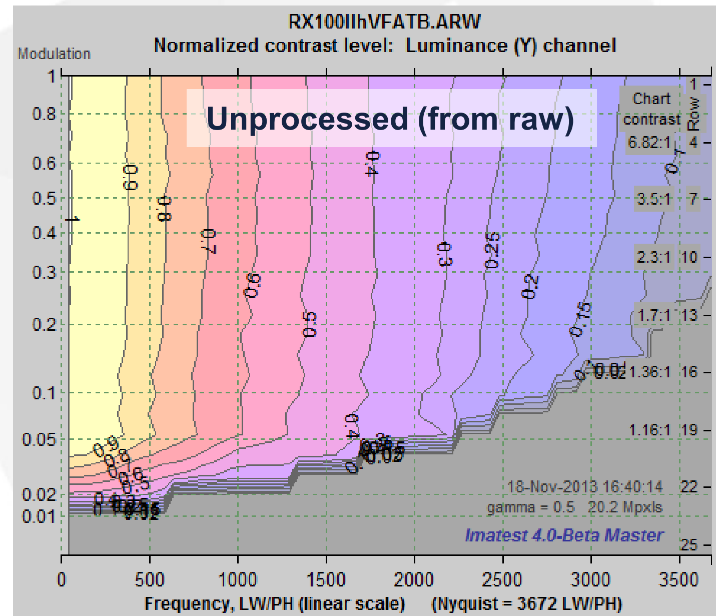
Random 1/f accurately measures texture, but lack of edges makes focusing, visual analysis difficult.



Low-contrast Siemens Star used in ISO/TS 19567-1:2016(E). Only one contrast level; much less information than Log F-Contrast.



Log F-Contrast results from raw image (Sony RX-100 II); no image processing; vertical contours indicate no noise reduction.



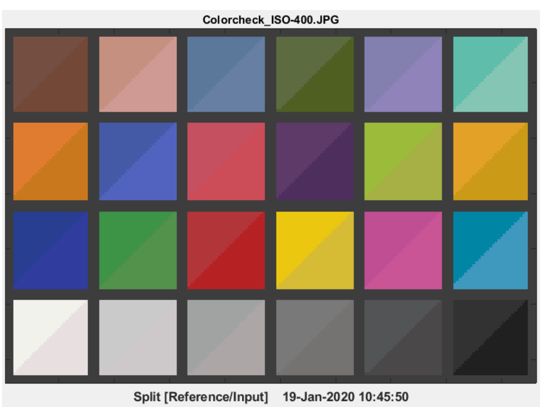
Row	Modulation	Contrast	MTF50	MTF50P	MTF20	MTF20P	MTF10	MTF10P	Peak
1	0.999	>40:1	1393	1393	2837	2837	3538	3538	1
4	0.698	6.82:1	1422	1422	2791	2791	3465	3465	1
7	0.508	3.5:1	1410	1410	2745	2745	3465	3465	1
10	0.335	2.3:1	1410	1410	2699	2699	3324	3324	1
13	0.216	1.7:1	1393	1393	2633	2633	3228	3228	1
16	0.120	1.36:1	1381	1381	2557	2557	3020	3020	1
19	0.053	1.16:1	1276	1276	2515	2515	3552	3538	1

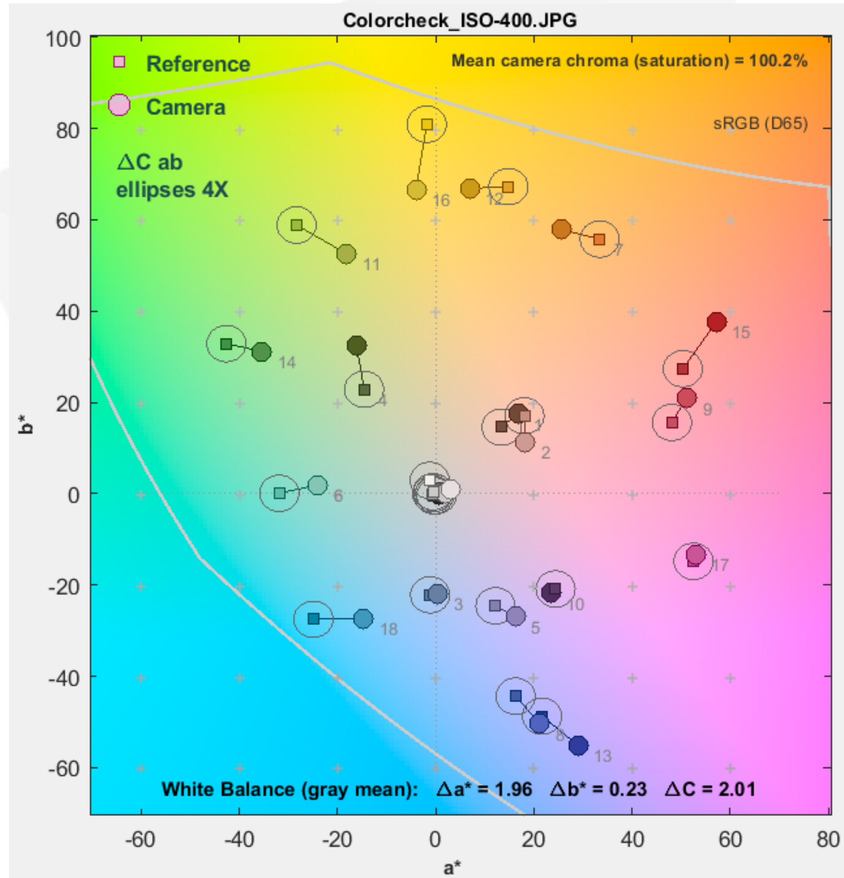
Color difference 1

$\Delta E_{ab} = \sqrt{\Delta L^{*2} + \Delta a^{*2} + \Delta b^{*2}}$ and $\Delta C_{ab} = \sqrt{\Delta a^{*2} + \Delta b^{*2}}$ are the most common, familiar measures of color difference,

They are simple geometric distance in CIELAB color space, but they do not correspond well to human perception

The circles (or ellipses) on the a*b* color difference plot (derived from the Colorchecker image below), represent $\Delta C = 4$ (more than one JND).





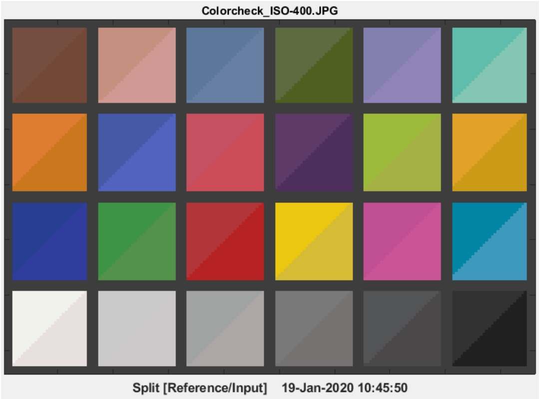
Color difference 2

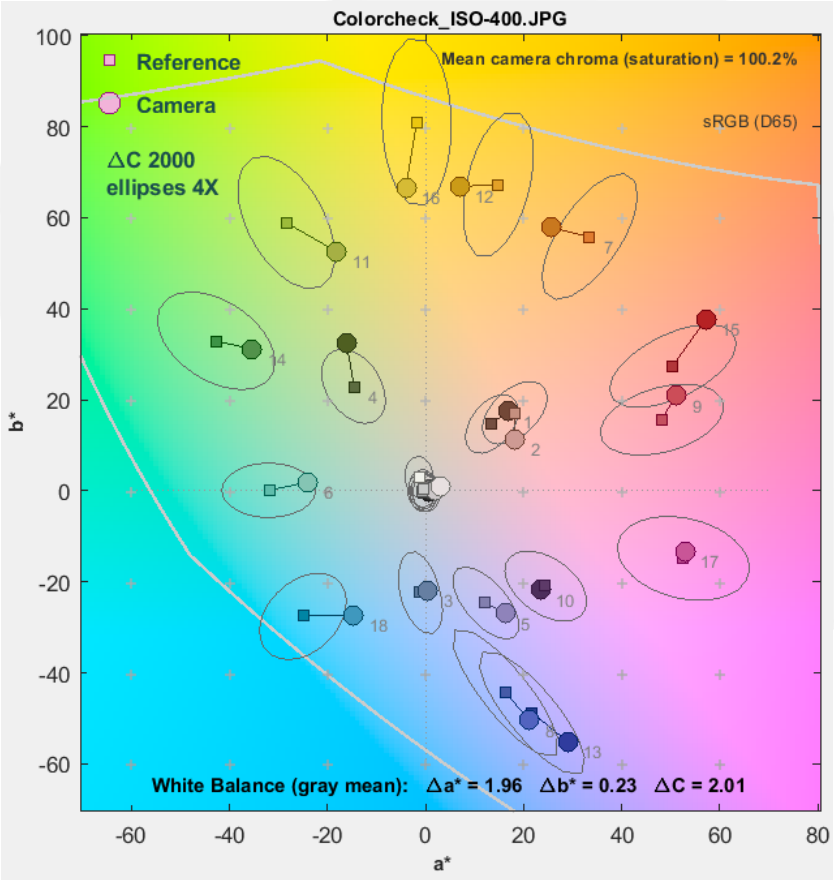
ΔC_{2000} and ΔE_{2000} are much better indicators of perceptual color difference.

Their equations are complex, but easy to evaluate on computers.

The circles (or ellipses) on the a*b* color difference plot, represent ΔC_{2000} = 4 (more than one JND).

Summary — Always use ΔC_{2000} and ΔE_{2000} for describing color difference. ΔC_{ab} and ΔE_{ab} are familiar, but not accurate.





Summary

Sharpness: Avoid **MTF50**. Use **MTF50P** and **overshoot** if needed.

Noise/SNR: Be aware of noise reduction from bilateral filtering (in JPEGs from cameras), which can lead to exaggerated SNR (and DR) measurements.

Dynamic Range (DR): Be aware of the effects of flare light. Remember that DR is the lower of two measurements: **quality (SNR)-based DR** and **slope-based DR**. ***Both*** must be measured. Consider using the **Contrast-Resolution chart**.

Texture: Look at the Dead Leaves image to be sure it is consistent with the MTF curve. Consider using the Log F-Contrast chart.

Color difference: Use ΔE_{2000} and ΔC_{2000} instead of ΔE_{ab} and ΔC_{ab}. The ellipses in the a*b* plot are good indicators of color differences.

Key advice (again!): Look carefully at the image and make sure measurements correlate with what you see.

Thank you

