https://doi.org/10.2352/ISSN.2470-1173.2020.9.IQSP-242

Correcting Misleading Image Quality Measurements

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

Abstract

We discuss several common image quality measurements that are often misinterpreted, so that bad images are falsely interpreted as good, and we describe how to obtain valid measurements.

Sharpness, which is measured by MTF (Modulation Transfer Function) curves, is frequently summarized by MTF50 (the spatial frequency where MTF falls to half its low frequency value) But because MTF50 strongly rewards excessive sharpening, we recommend other summary metrics, especially MTF50P (the spatial frequency where MTF falls to half its peak value), that provide a more stable indication of system performance.

Camera dynamic range (DR), defined as the range of exposure (scene brightness) where the image has good contrast and Signal-to-Noise Ratio (SNR), Is usually measured with grayscale step charts. We have recently seen several cases where flare light radiating out from bright areas of the image fogs dense patches, causing unreasonably high DR measurements. This situation is difficult to handle with linear test charts, where the flare light is aligned with the patches, but can be handled well in charts with circular patch patterns, where the patch where pixel level ceases to decrease defines the upper DR limit.

Introduction

We discuss several common image quality metrics (often called KPIs—Key Performance Indicators) that are frequently misunderstood or misinterpreted, and we recommend remedies, summarized in a table at the bottom.

MTF (sharpness) summary metrics

MTF (Modulation Transfer Function) is a key indicator of image sharpness, typically expressed as a function of spatial frequency, which can have any of several units (Cycles per Pixel, cycles per mm, Line Widths per Picture Height, etc.) MTF performance is often characterized by one of the following summary metrics.

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

Values other than 50% are sometimes used. For example, MTF10 corresponds to vanishing resolution (from the Rayleigh criterion— the spacing where two points can be distinguished, e.g., 4 pixels for MTF10 = 0.25 C/P), but MTF10 is not recommended because $d\,\mathrm{MTF}(f)/df$ is much lower than at the 50% level, making it highly sensitive to noise.

MTF50 has traditionally been the most widely-used of these metrics. The problem with MTF50 is that it is so sensitive to sharpening that highly sharpened images (with strong overshoot, resulting in visible "halos" at edges) are rewarded with high measured values.

We studied the effects of sharpening on MTF with slanted-edge and star patterns, starting with an image acquired with a high quality 24 Megapixel Micro Four-Thirds camera at ISO 200, converted from RAW to a TIFF using dcraw (set for sRGB color space with no sharpening or noise reduction).

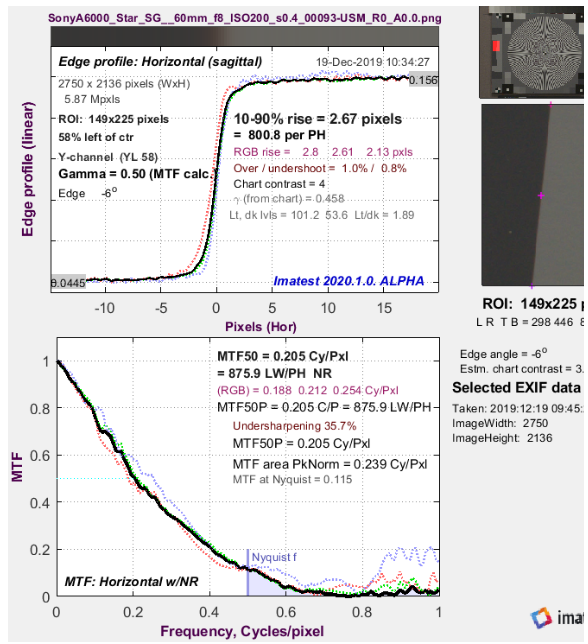


Figure 1. *Average edge and MTF curve for the original image used for the sharpening study.*

We then applied Unsharp Masking (USM) with a variety of parameters (Radius R = 1 and 2; Amount A = 0 to 4) to this image. USM operates by subtracting a gaussian-blurred replica of the image from the image itself. The USM equation is not explicitly stated in the MATLAB documentation [1], but likely has the form,

$$MTF_{USM}(f) = \frac{1-A\exp(-f^2\sigma_x^2/2)/\sqrt{2\pi}}{1-A/\sqrt{2\pi}} \quad \text{where } R = \sigma_x. \quad (1)$$

Note that USM is similar to standard sharpening, which subtracts a spatially-shifted replica of the image, and is commonly used in cameras because it runs faster. Sharpening with $R \cong 2$ is common in cameras. Here is an example of the image used in the study with a moderate amount of sharpening ($R = 2$, $A = 1$).

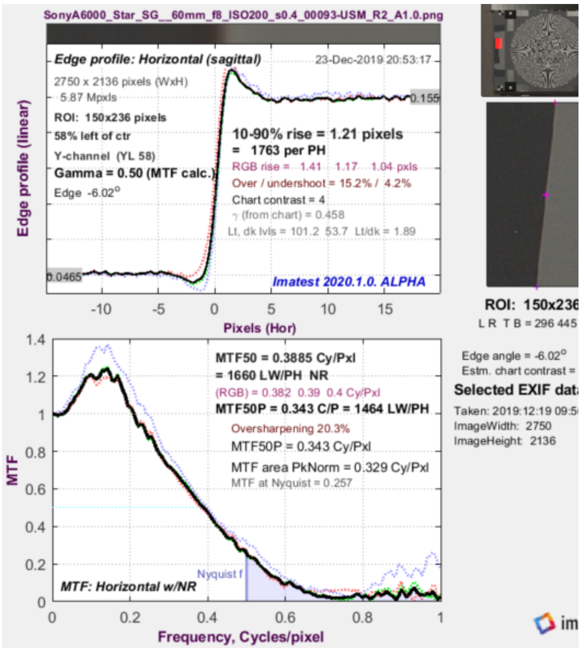


Figure 2. *Average edge and MTF curve for a moderately sharpened image (R = 2; A = 1) from the sharpening study. Note overshoots in the spatial domain (upper plot) and frequency domain (lower plot), corresponding to "Slanted edge overshoot" and "Slant MTF overshoot" in Figure 5.*

The results of the sharpness study, showing the response of the summary metrics to sharpening amount A and radius R, are shown in Figures 3 and 4.

The key thing to observe in these Figures is that increasing sharpening amount A improves all three summary metrics up to the point where overshoot starts. From Figure 5, this happens when $A \cong 1.5$ for $R = 1$ and $A \cong 0.5$ for $R = 2$.

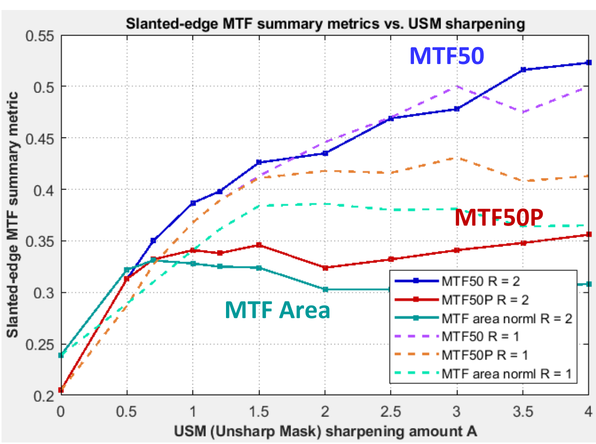


Figure 3. *Summary metrics (MTF50, MTF50P, and MTF area normalized) for the slanted-edge.*

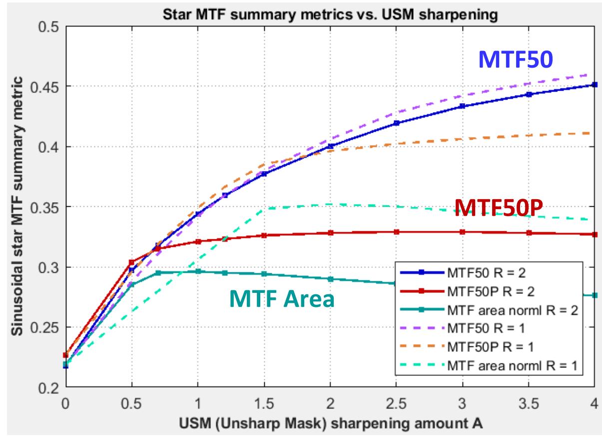


Figure 4. *Summary metrics for the Siemens Star pattern (from the same images as the slanted-edge).*

MTF50 continues to increase beyond the onset of overshoot, but MTF50P and MTF Area Normalized flatten out, increasing only slightly. This means that MTF50P and MTF Area Normalized are better indicators of fundamental imaging system performance, whereas MTF50 is a better indicator of software sharpening. We prefer MTF50P because it is more familiar and tracks MTF50 more closely for sharpening below the onset of overshoot.

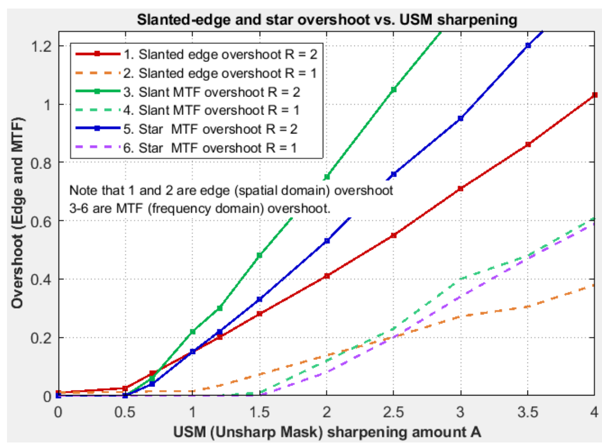


Figure 5. Overshoot *(as a fraction of the asymptotic value).*

Figures 2 and 5 illustrate two distinct types of overshoot: spatial domain (Figure 5, curves 1 and 2) and frequency domain (Figure 5, curves 3-6). The two are highly correlated. For an edge with asymptotic (settling) pixel levels 0 and P_{asymp} and maximum level P_{max},

$$\text{Spatial domain overshoot} = \frac{P_{max} - P_{asymp}}{P_{asymp}} \tag{2}$$

$$\text{Frequency domain overshoot} = \frac{MTF(max) - MTF(0)}{MTF(0)} \tag{3}$$

We have done a study (not shown) of the effects of noise on MTF50P and MTF Area that shows that neither are more sensitive to noise than MTF50, i.e., although random variation of these metrics increases with noise, there is no systematic change.

Sharpness recommendations — Because MTF50 is highly sensitive to sharpening, which degrades visible image quality when applied in excess, it is not a good summary metric for image system performance, and should be avoided with processed images (or images from "black box" cameras). MTF50P is strongly preferred. MTF Area normalized is also of interest, but less familiar, and doesn't track MTF50 quite as closely.

For processed images, an overshoot measurement—either in spatial or frequency domain—should be included whenever a sharpness summary metric is reported. A single number is insufficient to characterize system sharpness.

And remember that summary metrics, no matter how good, never quite tell the whole story.

Noise and SNR measurements

Most noise, SNR, and dynamic range (DR) measurements are made from images of flat patches, for example patches in ISO 14514 and ISO 15739 reflective charts as well as several transmissive Dynamic Range charts.

When these measurements are made from processed images, especially in-camera JPEGs, they are frequently affected by bilateral filters [2], which are edge-preserving noise reduction filters that smooth (lowpass filter) neighborhoods with small variance (uniform or nearly-uniform areas), but do not smooth areas with large variance, such as strong edges. They may, in fact, sharpen edges (boost high frequency content; the opposite of lowpass filtering).

Bilateral filters tend to improve measured SNR and DR, while having little effect on MTF measured from edges. They are nearly universal in the JPEG output of consumer cameras and camera phones. The amount of noise reduction is often increased as the Exposure Index (ISO setting) increases. This improves perceptual image quality, but makes measurements difficult—and removes the information contained in fine texture. Bilateral filters are not applied to raw output (which can be selected in many high-end cameras).

We will show an example of a camera phone with a nasty bilateral filter that caused extreme oversharpening on edges, but obliterated low level detail. This adversely affected texture measurements made with the Deal Leaves (Spilled Coins) chart. We suggest alternative charts for texture measurements.

The accompanying paper (on Information capacity) [3] has a technique for measuring noise in the presence of a Siemens star image. This method is less affected by bilateral filtering, but not completely immune.

Dynamic Range measurements and flare

In an earlier paper [4] we expressed concern about the effect of flare light on dynamic range (DR) measurements. Since it was published, we have seen several cases where flare light caused exaggerated DR measurements, and we have developed an approach for limiting measurement errors caused by flare light.

A review of the context is necessary to fully grasp the issue. Dynamic Range is defined as the range of exposure (scene illumination) where the camera responds with good contrast and good Signal-to-Noise Ratio (SNR). In practice, this means that

A. the contrast (slope of the tonal response curve, log(*pixel level*) vs. log(*exposure*)) is greater than a specified minimum (we use 0.075× the maximum slope), *and*

B. the scene-referenced SNR is also above a specified minimum, at least 1 (0dB) for low image quality (10 (20dB) for fairly high quality).

We have named the results of these two criteria, A. slope-based DR, and B. quality-based DR, and we have measured them separately. We have learned from painful experience that the two measurements cannot be separated.

Both criteria must be satisfied for a DR measurement to be valid. If the contrast is lower than the minimum, no visible image is present (the 0.075× criterion may be a little too lax). If the scene-referenced SNR is lower than 0dB, the noise is so severe that no image detail is visible, even though mean patch density continues to decrease.

Camera sensor manufacturers now offer High Dynamic Range sensors with Dynamic Ranges specified at 120-150dB (10^6:1 to 3×10^7:1). Sensor DR measurements are made with a sequence of flat-field images, i.e., of images where DR is zero.

But everything changes when real HDR scenes (or HDR test chart images) are captured with cameras, which have a lens is between the object and the sensor. We have found that the primary effect the lens is the addition of flare light to the image—light that diffuses from bright to dark areas of the image, and that this diffusion can extend long distances from the source. The primary cause of this flare light is multiple secondary reflections in the lens ($M(2M\text{-}1)$ secondary reflections for an M-component lens). ISO 18844 measurements only characterize short-range flare; they provide no information on long-range flare that can strongly affect DR measurements.

The crux of the issue is that many manufacturers want to claim high dynamic ranges (≥120dB) for their cameras, and pressure is put on engineers to come up with corresponding measurements. To this end we have found that some engineers pick the measurement (either A. slope-based or B. quality-based) that gives them the highest number without considering that *both* criteria must be met.

The image below is apparently from a low-cost camera intended for the automotive industry (we don't know for sure).

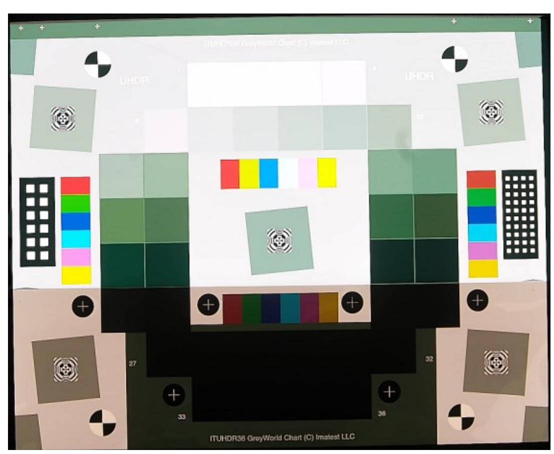


Figure 6. *Image of 36-patch High Dynamic Range chart that had a dynamic range problem caused by flare light (not visible in the image).*

The results of analyzing this image (the old way) are shown in Figure 7. Quality-based DR measurements are exceptionally high: 68.5dB for "High quality" (Scene-referenced SNR = 20dB) to a completely unreasonable 144dB for "Low quality" (0dB SNR).

On the other hand, the slope-based Dynamic Range is 70.3dB. The explanation for these results is based on Figure 8.

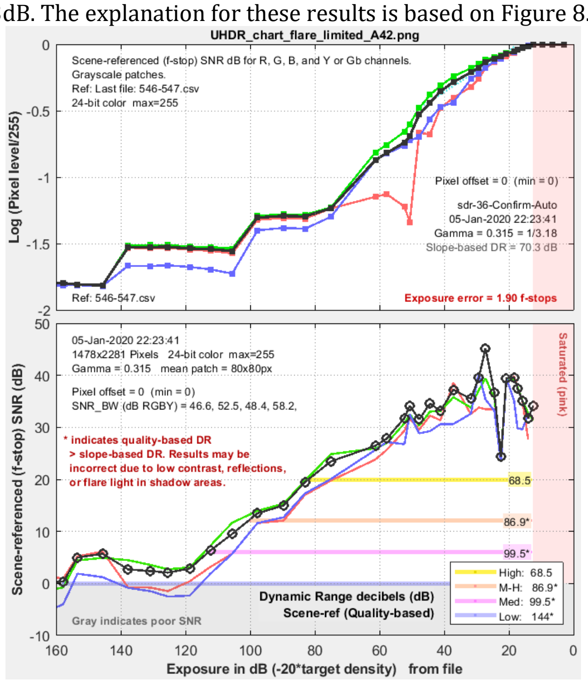


Figure 7. Tonal response and scene-referenced SNR for the chart in Figure 6, showing (old) measurements of slope-based and quality-based DR.

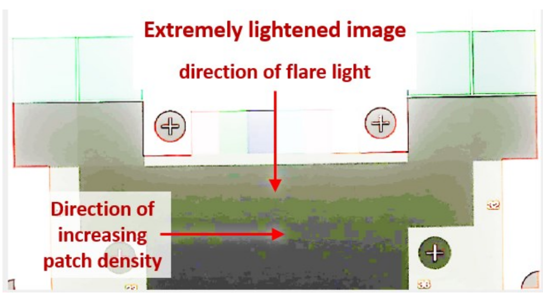


Figure 8. The lower portion of Figure 8, shown extremely lightened, illustrating how the direction of increasing (chart) patch density is orthogonal to the direction of decreasing flare light (originating from the top of Figure 6.)

Figure 8 shows the bottom half of the chart image, extremely lightened. Patch densities increase from left to right in each row. Flare light radiating from the light patches at the top of the image completely dominates the bottom three rows the image. The effect of flare light is plainly visible in the upper plot in Figure 7. Response flattens out in three steps: exposure between 75 and 98dB (row 4), 115 and 138dB (row 5), and beyond 145dB (row 6). The slope-based Dynamic Range (70.3dB) is the range of exposure between saturation and the patch were the response flattens out.

In current versions of *Imatest* we limit the quality-based DR to the slope-based DR, i.e., we no longer report quality-based dynamic ranges beyond the slope-based limit. The results are shown in Figure 9.

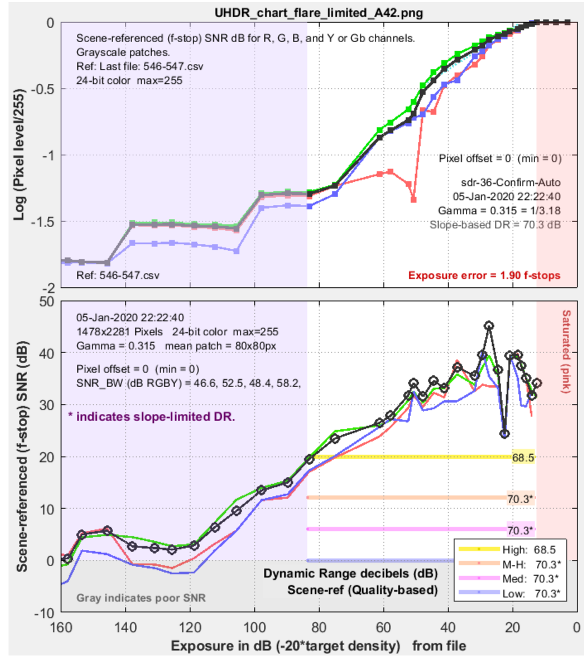


Figure 9. Tonal response and scene-referenced SNR for the chart in Figure 6, showing the new display where the quality-based DR is limited to the slope-based DR.

We need to mention that it is common for slope-based DR to exceed quality-based DR, especially in mirrorless or DSLR cameras with high-quality lenses that have well-controlled flare light.

Dynamic Range recommendations — Neither slope-based nor quality-based Dynamic range are sufficient by themselves to characterize a camera's dynamic range. ***Both must be taken into account.*** Total dynamic range is the lower of the slope and quality-based measurements.

In many recent cameras— especially low-cost cameras intended for the automotive industry—DR measurements are strongly affected by flare light, which can be mistaken for the test chart image. We have shown how flare light can be distinguished from image in charts with a circular patch arrangement. We do not recommend the use of linear test charts. The Contrast Resolution chart [4] and analysis is also highly effective in removing the effects of flare light.

Texture measurements

Both the Dead Leaves [5][6] and Spilled coins charts consist of circles of random size and density (with maximum contrast of 3:1) and both are scale-invariant (which implies a $1/f^2$ Power Spectral Density (PSD)). The authors of [5] and [6] point out that that a power-law PSD is typical of common scenes. We verified this with a variety of images including travel and grandchildren, where the average PSD was $1/f^{2.2}$.

Dead Leaves/Spilled Coins charts have become an industry standard for measuring texture. Unfortunately, they are subject to some particularly egregious errors from extreme bilateral filtering. This is a paradox because bilateral filtering, which reduces texture (fine detail with medium to low contrast) below the value expected from slanted-edge MTF measurements, is the primary reason for separate texture measurements.

Before we get into depth with the texture issue, we should mention the other issue with Dead Leaves texture measurements: noise. It is very difficult to distinguish the chart pattern from noise, but there is a straightforward solution: signal averaging. Averaging n images improves the Signal-to-Noise Ratio (SNR) by a factor of $\sqrt{n}$, i.e., doubling the number of samples increases SNR by 3dB. While this slows down measurements, it's seldom a serious problem because texture is typically measured during camera development and evaluation, rarely during production.

The best way to illustrate where Dead Leaves texture measurements can go wrong is to show an extreme example. This kind of ugly image processing is uncommon, but not as rare as we'd like. We see it from time-to-time, and anyone testing for texture should be aware of it.

First, Figure 10 contains the original (Spilled Coins) pattern.

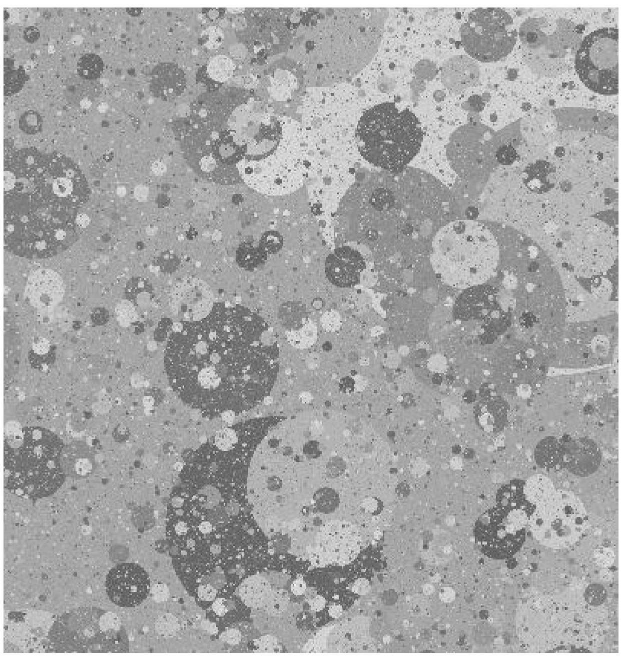


Figure 10. *Original spilled coins pattern (crop)*

Figures 11 and 12 illustrate the pattern reproduced by the iPhone 5, which has conservative sharpening with little visible noise reduction, and by "Phone B", which has extreme sharpening and noise reduction extreme enough to completely remove fine detail in smooth areas.

The visible image quality of "Phone B" is clearly far inferior to the iPhone, but its summary metrics from the MTF plots in Figure 13 (especially MTF50, whose shortcomings were pointed out earlier) are not that different. In particular, the MTF curve for "Phone B" gives little indication of how badly fine detail has been destroyed.

What is happening in Figure 12 is that the threshold where the bilateral filter transitions from extreme smoothing to extreme sharpening is lower than the maximum Dead Leaves contrast of 3. It must be stressed that this is quite unusual; it's definitely not "good behavior". Not only are the noise reduction and sharpening far more extreme than usual, but the transition is abrupt.

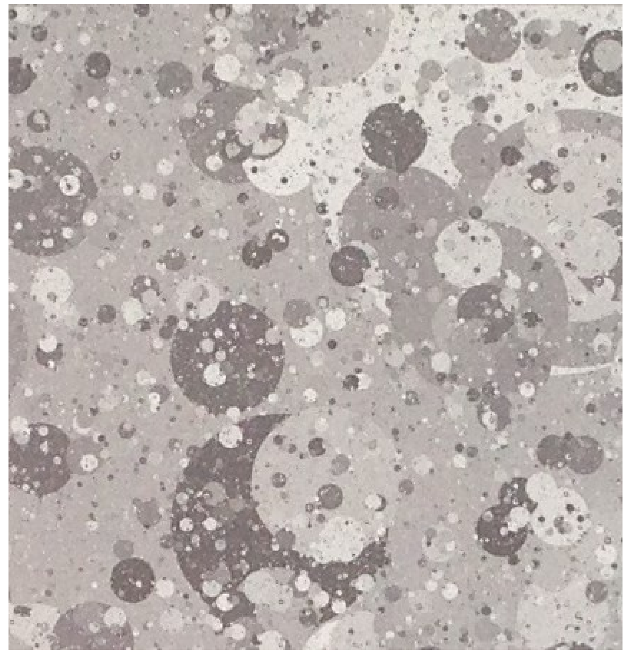


***Figure 11**. Spilled coins pattern reproduced with the iPhone 5, which has conservative sharpening and little apparent noise reduction*

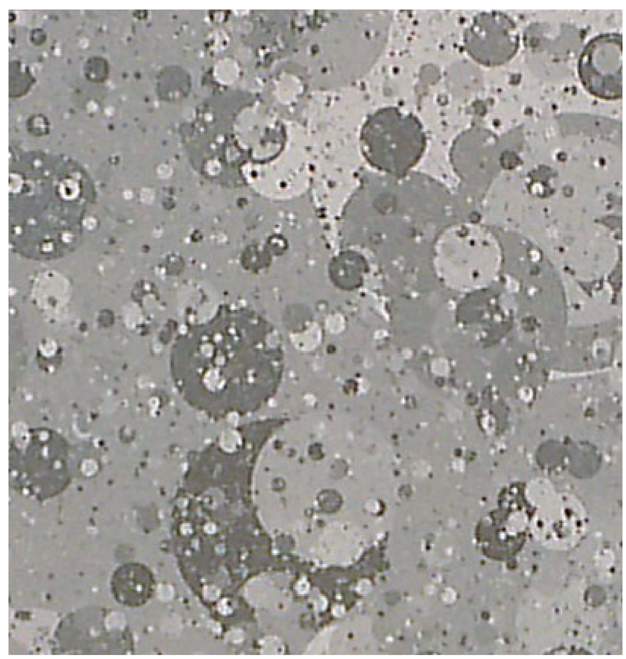


Figure 12. *Spilled coins pattern reproduced with "Phone B", which has strong sharpening visible on contrast edges and strong noise reduction that removes fine detail in smooth regions.*

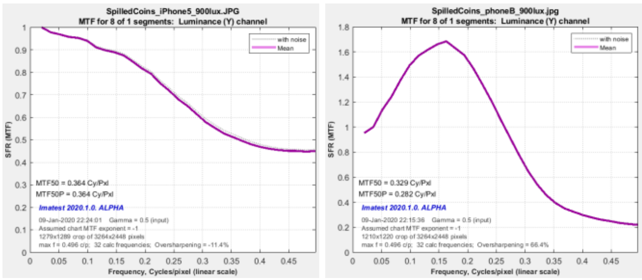


***Figure 13.** MTF curves for Figures 11 (iPhone 5) and 12 ("Phone B")*
For the iPhone 5 (left), MtF50 = MTF50P = 0.364 C/P.
For "Phone B" (right), MTF50 = 0.329 C/P; MTF50P = 0.282 C/P. The slanted-edge measurement was even more extreme: MTF50 = 0.414 C/P; MtF50P = 0.295 C/P; peak MTF = 2.8; overshoot = 86%.

Most image quality degradations can be easily measured, but in this case the measurement— the MTF curve— doesn't correlate well with the obseravtion that low level detail has been completely destroyed. The visibly oversharpened edges with relatively high contrast cominate the MTF curve. We haven't (yet) figured out an easy workaround with the Dead Leaves/Spilled Coins chart.

At least three charts other than Dead Leaves/Spilled coins can be used to measure texture: Log F-Contrast, Random Scale-Invariant, and the (low contrast) Siemens star.

Log F-Contrast chart

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.

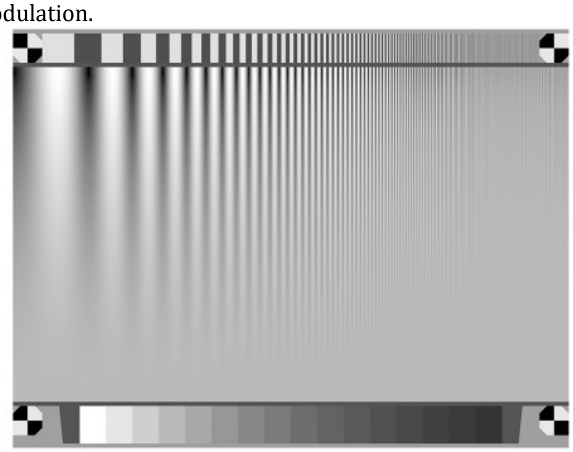
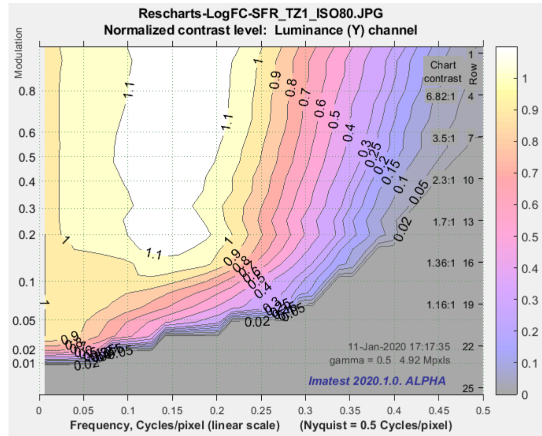


***Figure 14.** Log F-Contrast chart*

The Log F-Contrast chart is designed to fill only a portion of the frame, typically about 1,000 pixels high. The most useful result is the normalized contour plot, which clearly shows how image processing changes with chart contrast. The differences between a camera's response at high and low ISO speeds can be impressive.

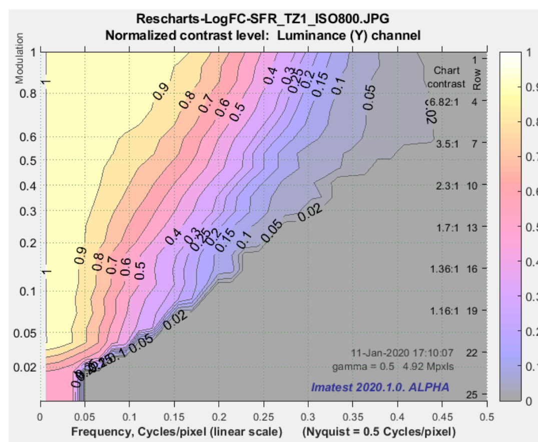
Figures 15 and 16 contain results for a compact digital camera at ISO 80 and 800.



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

***Figure 15.** Log F-Contrast results for compact digital camera (older model) exposed at ISO 80. Moderate sharpening and noise reduction.*

Note that for a pattern where the lightest level is L_{max} and the darkest is L_{min}, Contrast Ratio $= L_{max} / L_{min}$, Weber Contrast $= (L_{max} - L_{min})/L_{min}$, and Michelson Contrast = Modulation $= (L_{max} - L_{min})/(L_{max} + L_{min})$.

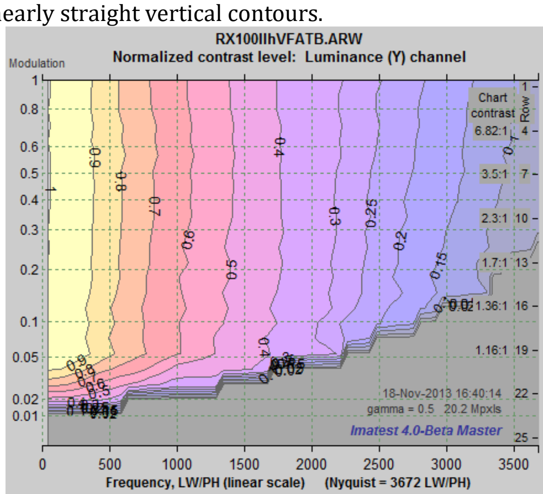


Row	Modulation	Contrast	MTF50	MTF50P	MTF20	MTF20P	MTF10	MTF10P	Peak
1	0.999	>40:1	0.246	0.246	0.308	0.308	0.346	0.346	1
4	0.788	6.82:1	0.229	0.229	0.292	0.292	0.333	0.333	1
7	0.589	3.5:1	0.198	0.198	0.264	0.264	0.313	0.313	1
10	0.409	2.3:1	0.17	0.17	0.231	0.231	0.272	0.272	1
13	0.245	1.7:1	0.139	0.139	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

***Figure 16.** Log F-Contrast results for compact digital camera (older model) exposed at ISO 800. Same conditions as Figure 15. Sharpening is gone; noise reduction is much stronger.*

The differences in the two results are quite dramatic. The moderate sharpening visible at ISO 80 (normalized contrast > 1.1) is gone at ISO 800 and the response rolloff for modulation < 0.5 drops off much more rapidly—clear evidence of increased noise reduction at low contrast levels. We know of no other test chart that measures MTF response as a function of chart contrast (or modulation).

For many years imaging-resource.com has included images of the Log F-Contrast chart, without analysis, in its test reports. The results for most cameras resemble Figure 15. But they sometimes post raw images, which can be converted with minimal processing. Here is an example from a Sony RX100 II camera (an excellent compact design with both raw and JPEG output). Results from this image are typical of images with no sharpening and no noise reduction, i.e., no bilateral filtering, as indicated by the nearly straight vertical contours.



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

***Figure 17.** Log F-Contrast results for a raw image from the Sony RX100 II, converted with minimal processing (no sharpening or noise reduction)*

Random 1/f pattern

This chart is similar to the Dead Leaves/Spilled Coins chart except that starts as a totally random pattern, then is shaped to have a $1/f^2$ PSD. It maximizes noise reduction (lowpass filtering), and, because it has no sharp edges, is not afflicted by the extreme sharpening that affected the Spilled Coins results in Figures 12 and 13.

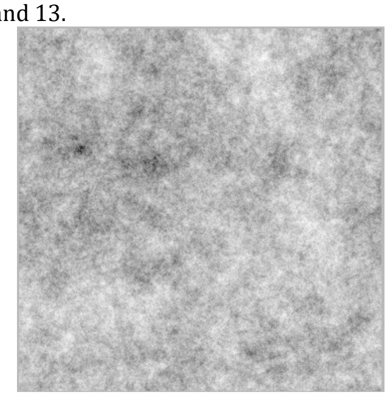


***Figure 18.** Random 1/f chart*

Although the Random 1/f chart maximizes noise reduction (with minimum sharpening), it has two striking disadvantages.

1. It is impossible to analyze visually because there are no sharp features for the eye to lock on to.
2. It works poorly with camera autofocus (or manual focus). Focusing must be done carefully outside the random area.

For these reasons, we are hesitant to recommend it for most purposes.

Low Contrast Siemens Star

Finally, a low contrast Siemens star is included in somewhat obscure standard, ISO/TS 19567-1:2016(E), "Photography — Digital cameras — Texture reproduction measurements — Part 1: Frequency characteristics measurements using cyclic pattern". [7] With sinusoidal star patterns there is no problem with oversharpened edges throwing off MTF measurements.

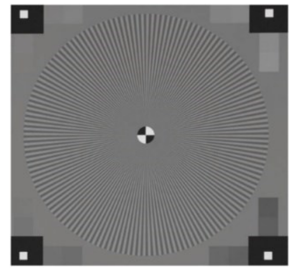


***Figure 19.** Low contrast Siemens star*

Contrast ratio is 3:1; Weber contrast is 2. Michelson contrast (modulation) is (3-1)/(3+1) = 0.5. The standard shows moderate angular dependence in Figure 6 and has several examples of MTF plots for a range of ISO speeds (Appendix C and D). Results are similar to what would be expected with Log F-Contrast, but they are limited because only one contrast level is available.

Texture recommendations — Dead Leaves/Spilled coins image are adequate in most instances, but you need to look carefully at the image to be sure details is not suppressed. There are infrequent cases where measurements fail, and bad cameras might be passed. The underrated (and underutilized) Log F-Contrast chart provides valuable information about how contrast affects image processing. We have discussed the attributes and limitations of the two additional charts.

Color difference measurements

It is well-known among color scientists that ΔE and ΔC (also called ΔE_{ab} and ΔC_{ab}, dating from 1976) are poor representations of perceptual color differences. But they are familiar and widely-used, especially by newcomers to image quality evaluation.

ΔE_{ab} is the simple geometric distance between points in CIELAB L*a*b* space: $\Delta E_{ab} = \sqrt{\Delta L^{*2} + \Delta a^{*2} + \Delta b^{*2}}$. $\Delta C_{ab} = \sqrt{\Delta a^{*2} + \Delta b^{*2}}$ (ΔE_{ab} with ΔL^* removed). Newer color difference formulas, ΔE_{94}, ΔC_{94}, ΔE_{2000}, and ΔC_{2000}, have much more complex equations [8], and graphic representations of these color differences have been very limited until recently.

The problem with ΔC_{ab} is that the eye is much less sensitive to chroma differences for highly saturated (chromatic) colors than for low-saturation colors (where $\sqrt{a^{*2} + b^{*2}} < 10$). For this reason, ΔC_{ab} can greatly exaggerate perceptual color differences.

Figures 20 and 21 are derived from an image of the X-Rite Colorchecker (Figure 22) captured with a good quality mirrorless camera. They show circles or ellipses representing $\Delta C_{ab} = 4$ and $\Delta C_{2000} = 4$, which is somewhat more than one Just Noticeable Difference (JND). The L*a*b* values of the acquired image are displayed as circles (○), while the reference values, supplied by X-Rite, are displayed as squares (□).

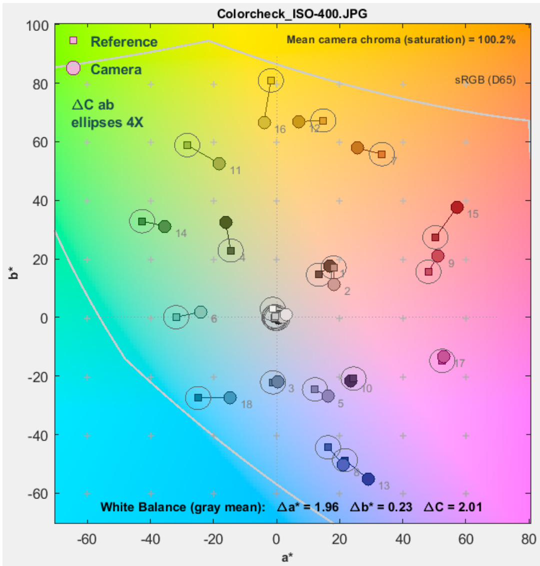


Figure 20. *Reference and camera a*b* values, showing ΔC_{ab}, circles*

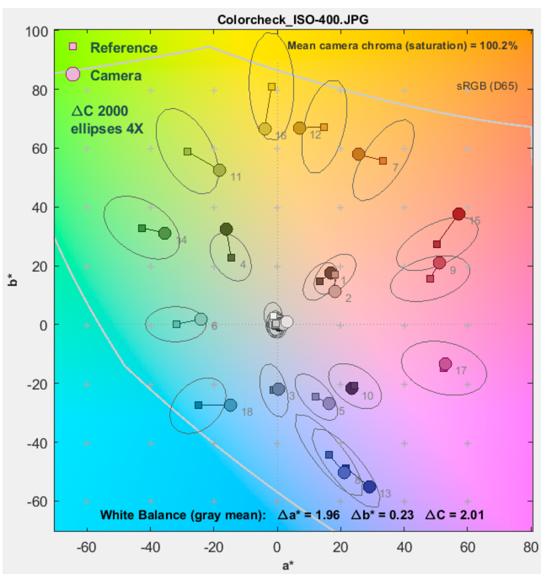


Figure 21. *Reference and camera a*b* values, showing ΔC_{2000}, circles*

Although ΔC_{ab} and ΔC_{2000} are similar for neutral colors (the bottom row of the Colorchecker, where a* ≈ b* ≈ 0 in Figures 20 and 21), they diverge strongly for large values of $\sqrt{a^{*2} + b^{*2}}$. Yellow patch 16 (near the top of the Figures) is a good example. $\Delta C_{ab} = 14.4$ would indicate a severe color error. $\Delta C_{2000} = 3.65$ (still significant) is more reasonable.

Figure 22 shows the split-color display for this image (reference on upper-left; input on lower-right). What you see will be limited by your display (printed versions may be less accurate). Also keep in mind that the image includes ΔL^* (luminance difference), which is not included in ΔC_{ab} or ΔC_{2000} calculations.

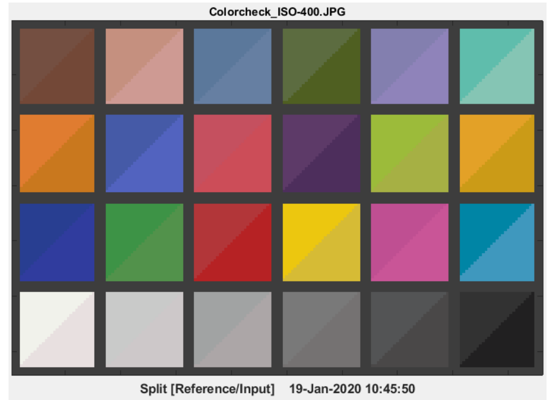


Figure22. *Split view (Reference/Input) of Colorchecker image used for Figures 20 and 21.*

Color difference recommendations — Always use ΔC_{2000} (or ΔE_{2000}) to specify color differences. ΔC_{94}, which hasn't been discussed, is actually very close to ΔC_{2000}, differing primarily in the blues, where the ΔC_{2000} ellipses are distinctly narrower.

References

[1] MathWorks (MATLAB) documentation on imsharpen (USM). [Online] www.mathworks.com/help/images/ref/imsharpen.html

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

[3] N. Koren, "Measuring camera Shannon Information Capacity with a Siemens Star Image," Electronic Imaging 2020, Society for Imaging Science and Technology.

[4] N. Koren, "Measuring the impact of flare light on Dynamic Range," Electronic Imaging 2018, Society for Imaging Science and Technology.

[5] F. Cao, F. Guichard, H. Hornung, "Measuring texture sharpness of a digital camera", Proc. SPIE 7250, 72500H, (2009).

[6] F. Cao, F. Guichard, H. Hornung, "Dead leaves model for measuring texture quality on a digital camera", Proc. SPIE 7537, 75370E (2010)

[7] ISO/TS 19567-1:2016(E), "Photography — Digital cameras — Texture reproduction measurements — Part 1: Frequency characteristics measurements using cyclic pattern" [Online], www.iso.org/standard/65356.html

[8] en.wikipedia.org/wiki/Color_difference [Online]

Author Biography

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

Appendix: Summary recommendations

We have three general recommendations

1. ***LOOK*** at the image. If the image looks good but the measurements are bad— or vice-versa— be suspicious of the results, which may be adversely affected by one of the factors (flare light, bilateral filtering, etc.) discussed above— or by an unexpected factor (perhaps an odd image processing artifact), not described here. Always be alert for new and different issues.
2. Be aware that bilateral filtering affects results from JPEG images acquired from cameras. It may be mild (minimal noise reduction) at low ISO speeds, but it often increases at ISO high speeds. It can affect all the measurements listed below except color difference. Signs bilateral filtering include: (a) a different MTF curve for raw and processed (images), (b) reduced MTF at low contrasts in Log F-Contrast results (Figures 15-17), or (c) different MTF curves for slanted-edges and Dead Leaves/Spilled Coins images.
3. Be especially aware of the effects of flare light during Dynamic Range measurements. Remember that sensor DR measurements, some of which exceed 120dB, cannot be attained with real cameras that have glass between the test chart and image sensor. And do not forget that the true DR is the lower of the quality-based and slope-based DR measurements. Neither measurement is sufficient by itself.

Summary Table of misleading measurements and recommendations

Measurement	Issue	Recommendation
Sharpness (MTF)	MTF50 is overly sensitive to sharpening. It can be high for a poor-quality camera with extreme sharpening.	Use MTF50P and always report overshoot (in spatial or frequency domain) for processed images. Note that MTF50P = MTF50 for unsharpened or slightly sharpened images.
Noise/SNR	Noise is often measured in flat patches of test charts, but bilateral filtering in processed images (JPEGs from cameras, especially at high ISO speeds) may reduce measured noise and increase SNR, and leading to unrealistic Dynamic Range measurements.	Use raw (unprocessed) images where possible. Be cautious in accepting results from processed images. Noise can now be measured in the presence of a signal using a Siemens star chart, as described in [3].
Dynamic Range (DR)	Flare light causes erroneous measurements, which can be unrealistically high.	DR the range of exposure (scene illumination) where the camera responds (A) with good contrast (slope-based DR), *and* (B) good SNR (quality-based DR). Customers have sometimes chosen the one that gives best results. This leads to exaggerated and erroneous results. Both must be measured, and the smaller of the two must be used.
Texture	Dead leaves (spilled coins) measurements give valid results most of the time, but can be confused by noise and bilateral filtering.	Look at the image to see if the results make sense. If the image has both sharpening "halos" and loss of detail for lower contrast detail, the MTF may not be accurate. Consider using the Log F-Contrast chart to view the effects of image contrast on image processing. The Random 1/f chart gives an accurate measurement of loos of detail due to noise reduction, but it is not good for visual analysis.
Color difference	ΔE_{ab} and ΔC_{ab} (= ΔE and ΔC) are familiar and widely-used, but are not a good approximation to human perception.	Use ΔE_{2000} and ΔC_{2000} (or ΔE_{94} and ΔC_{94}; very similar) instead of ΔE_{ab} and ΔC_{ab}. The ellipses in the a*b* plot are good indicators of color difference.

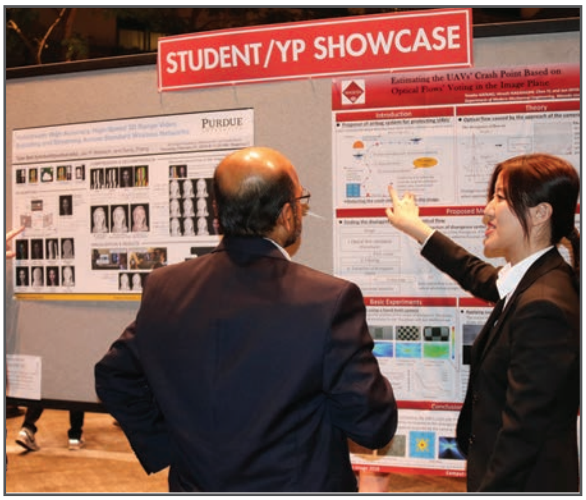
JOIN US AT THE NEXT EI!

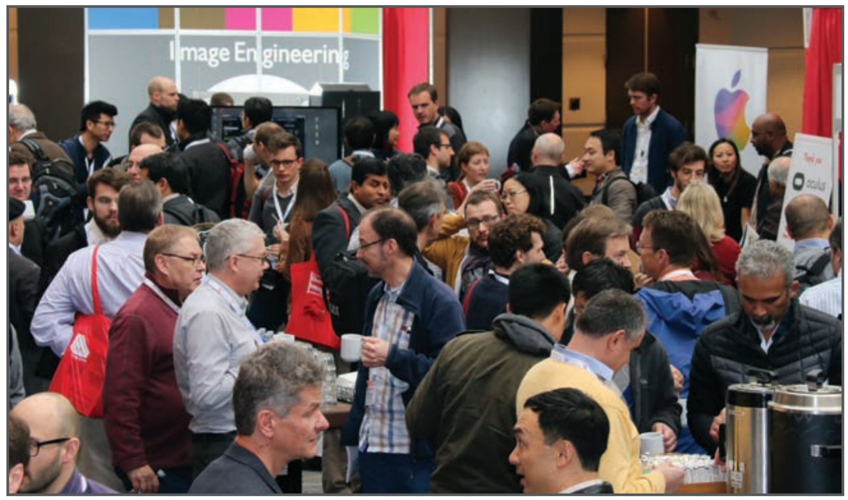
IS&T International Symposium on

Electronic Imaging

SCIENCE AND TECHNOLOGY

Imaging across applications . . . Where industry and academia meet!





• SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •
• INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •

www.electronicimaging.org

IS&T
imaging.org