On a novel technique to quantify local contrast in HDR scenes

Pu Zheng, François-Xavier Thomas, Claudio Greco, Frédéric Guichard
DXOMARK, Boulogne-Billancourt, France

Abstract

This work provides a novel glass-to-glass metric of local contrast, useful in the context of image quality evaluation of HDR content. This metric, called Local-Contrast Gain (LCG), uses the opto-optical transfer function (OOTF) of the imaging system and its first derivative to compute the incremental ratio between contrast in the scene and contrast on the display. In order to be perceptually meaningful, we chose Weber's definition of contrast. In order know the OOTF in analytical form and to make the measurement robust to the uncertainty of measurements of the ground truth, we rely on a model that we propose and that expands upon our previously published work. We provide experimental validation of our metric on a variety of target charts, both reflective and transmissive, both in isolation and within complex setups spanning more than six EVs.

Introduction

High Dynamic Range (HDR) imaging is a technology that enhances the contrast and color range of images, making them appear more vivid and realistic. Several formats exist for both photos (HEIF, AVIF with or without gain maps, *etc*.) and videos (HDR10, HDR10 +, Dolby Vision, etc.) that can encode a luminance greater than 1000 nit with contrasts higher than $1000 : 1$. When viewed on a compatible display, these pictures can reproduce real-life scenes more realistically for human observers. Although HDR provides exciting new possibilities, it also presents novel challenges in terms of evaluation of the performance of an imaging system.

Traditional approaches to evaluating the performance of an imaging system in terms of exposure usually rely on very synthetic statistics, such as the average lightness level in the scene (target exposure) or the global contrast, *i.e*, the contrast between the brighter and darker lightness levels represented in the image. These metrics are usually evaluated in a linear space, but occurrences where the evaluation is carried out in a gamma-encoded space (such as Y'CbCr) are not rare.

These approaches are perfectly suitable for the evaluation of the legacy Standard Dynamic Range (SDR) formats, where the range of luminance that can be captured by the system is limited and the encoding opto-electronic transfer functions (OETF) are generally quite similar to each other and to the inverse of the decoding electro-optic transfer function (EOTF) specified by their format. Unfortunately, the same does not apply to HDR imaging systems.

Given that HDR imaging is *de facto* becoming the norm for consumer electronic imaging, it becomes imperative to define new suitable figures of merit to be able to quantify the user experience.

Our goal is to define a figure of merit for the preservation of contrast in a zone of limited dynamic range contained in a spatial region of a larger HDR image. In order to propose a viable metric, we had to provide a solution to several challenges, listed below.

First and foremost, the metric has to describe perceptually relatable phenomena. In particular, it has to be able to distinguish cases of visible enhancement, preservation, compression, lack, or inversion of contrasts in a fashion similar to that of an expert human observer.

Furthermore, the results must be independent of the encoding format of the image. No unfair advantage should result from a different representation of the same luminance on-screen.

Practical considerations of the measurement process also deserve attention. The results have to be robust with respect to the uncertainty with which the ground truth can realistically be acquired.

Methodology

The proposed LCG metric is, in essence, a metric of preservation of perceptual contrast between a scene and its representation on screen.

Let L_{scene} be the ground truth, *i.e.* the luminance values measurable in the scene (which, in a laboratory environment, we control to a high degree), and let L_{display} be the corresponding values on screen, which depend on how the imaging system responds to the lab setup. Strictly speaking, L_{display} should be measured on the display itself with a light-meter; however this is both impractical and not necessarily useful, since it introduces a dependence on the particular display used. For these reasons, every time we refer to display luminance, henceforth we will implicitly refer to the luminance that the standard decoding of the image prescribes for an ideal display.

The relation between scene luminance and display luminance is given by the Opto-Optical Transfer Function (OOTF) of the system $f : L_{\text{scene}} \mapsto L_{\text{display}}$. This is usually the composition of two functions: an unknown encoding Opto-Electronic Transfer Function (OETF) that maps scene luminance into image luma and an Electro-Optical Transfer Function (EOTF) that maps image luma into display luminance. The EOTF is, in general, inferable from the image format and its metadata. The OETF was originally simply the inverse of the EOTF up to a factor (and thus the OOTF was linear), but camera manufacturers have since learnt to use its shape as a tuning parameter for contrast enhancement.

The first step is therefore to acquire *a priori* knowledge on a set of contrast values in the scene that will serve to build our ground truth. In practice this is done by evaluating the luminance coming from a set of regions-of-interest (ROI). For the purposes of this metric, ROIs should both be reasonably close in the framing and have low patch-to-patch contrast. Some examples of target charts that provide such ROIs are given in Fig. 2.

Since measuring the luminance from these ROI one by one could be unfeasible (*e.g.*, because of their high number or their small size), in practice for reflective chart one can rely on the

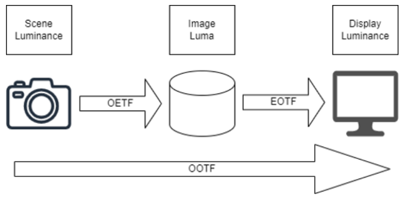

Figure 1. *Opto-Optical Transfer Function as a composition of encoding Opto-Electronic Transfer Function and decoding Electro-Optical Transfer Function*

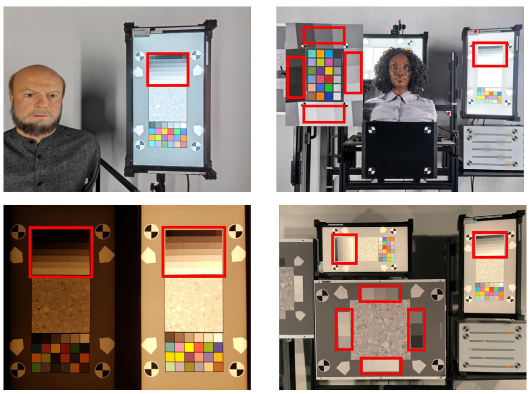

Figure 2. *Examples of target charts suitable for local contrast evaluation.*

overall illuminance on the chart and the reflectance of each ROI, or for transmissive charts on the overall luminance of the backlighting panel and the transmittance of each ROI.

On the picture, we compute the average lightness level for each ROI. These values are linearized in display-referred space, *i.e.*, using the metadata (and the current standards) to estimate the actual luminance in nit that a reference display would emit for that ROI. Linearization into the display space ensures that the results are independent of the image encoding, which was one of our goals.

To find a representation of the local OOTF, we fit a general model to the luminance pairs (scene luminance, display luminance) designed to accurately capture the saturation of the highlights, as well as clipping and contrast inversion of the dark areas.

Let $L_{\text{display}} = f(L_{\text{scene}})$ be the estimated OOTF. In differential form, the Weber contrast between two infinitesimally close luminance values in the scene is described by:

$$W_{\text{scene}} = \frac{\mathrm{d}L_{\text{scene}}}{L_{\text{scene}}}$$

The corresponding Weber contrast on-screen will be:

$$W_{\text{display}} = \frac{\mathrm{d}L_{\text{display}}}{L_{\text{display}}}$$

If we let $L = L_{\text{scene}}$, then the contrast gain, *i.e.*, the ratio between the on-screen contrast and the scene contrast it represents is given by

$$\begin{aligned} \mathrm{LCG}(L) &= \frac{\frac{\mathrm{d}f(L)}{f(L)}}{\frac{\mathrm{d}L}{L}} \\ &= \frac{\frac{f'(L)}{f(L)}}{\frac{1}{L}} \\ &= \frac{\mathrm{d}\log f}{\mathrm{d}\log L}, \end{aligned} \tag{1}$$

which defines our metric.

Notice that LCG(L) indeed satisfies our last goal:

- when $\mathrm{LCG} > 1$, the contrast is perceptually boosted;
- when $\mathrm{LCG} = 1$, the contrast is perceptually preserved;
- when $0 < \mathrm{LCG} < 1$, the contrast is perceptually compressed;
- when $\mathrm{LCG} = 0$, the contrast is lost (saturation);
- when $\mathrm{LCG} < 0$, the contrast is inverted.

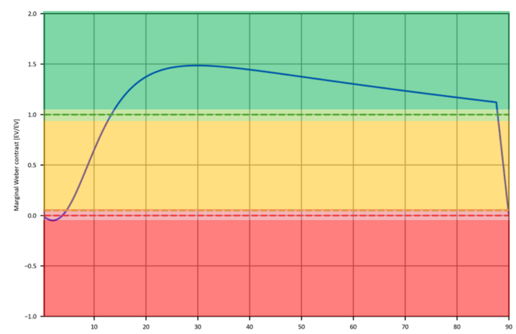

Figure 3. *Regions of LCG values in which the contrast is boosted (green), compressed (yellow), or inverted (red).*

As the ratio of two luminance contrasts in logarithmic scale, we can measure the LCG equivalently in dB/dB or EV/EV; in this article we shall use the latter. The LCG is in practice the *point elasticity* of the OOTF; in this sense, it can be understood as the continuous version of the Tonal Contrast Gain metric defined in the Dynamic Range measurement of the IEEE-SA P2020 Automotive Standard [1].

Estimation of the Opto-Optical Transfer Function

As mentioned above, the OETF is in general not known and needs to be estimated as part of the measurement process.

It is worth asking whether to opt for a fitting of the OOTF or just an interpolation. There are two main reasons to prefer a fitting:

1. Fitting provides robustness to the uncertainty in the measurement of the ground truth
2. Fitting gives the OOTF and its derivative in analytical form, which is much more convenient

The uncertainty is due to how the ground truth can be acquired. For transmissive charts, the acquisition requires either measuring patches individually –which is unpractical, time-consuming, and error prone– or relying on the transmittance of the patches, but this is affected by the vignetting of the light panel. For reflective charts: direct measurement of luminance is impossible; thus, one must rely on reflectance and luminance and is impacted by the uniformity of the lighting.

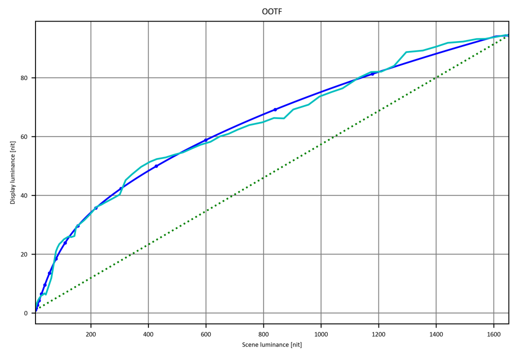

Figure 4. *OOTF reconstructed via fitting (blue) or interpolation (cyan). The unavoidable measurement noise affecting the acquisition of the ground truth makes the fitting approach preferable.*

The model used here is an extension of that used in our previous work [2], *i.e.* the well-known *parametric Naka-Rushton contrast function* [3]:

$$f_{K,n}(L) = \frac{(K^n+1)\,L^n}{K^n+L^n}, \tag{2}$$

where n and K are respectively the slope and highlight roll-off inflection point.

This function provides a way to describe common contrast curves with only two parameters. It typically presents S-curves for $K \in [0,1]$ and $n > 1$, and good approximations of $\gamma = n$ power functions for sufficiently large $K > 1$, as shown in Fig. 5).

This function can describe quite accurately the majority of global OOTF on a normalized scale. To be able to describe local OOTFs in photometric scale, we extend the model thus:

$$f_{\mathbf{t}}(L) = f_{G,S,K,n,L_0,L_{\text{sat}}}(L) = \begin{cases} L_0 + G \cdot f_{K,n}\left(\frac{L}{S}\right) & \forall L \leq L_{\text{sat}} \\ f_{G,S,K,n,L_0,L_{\text{sat}}}(L_{\text{sat}}) & \forall L > L_{\text{sat}} \end{cases} \tag{3}$$

Where G and S are factors to normalize in the interval $(0,1)$ both display and scene luminance respectively, L_0 is a luminance offset, be it due to the encoding function or camera glare, and L_{sat} is the sensor saturation luminance value.

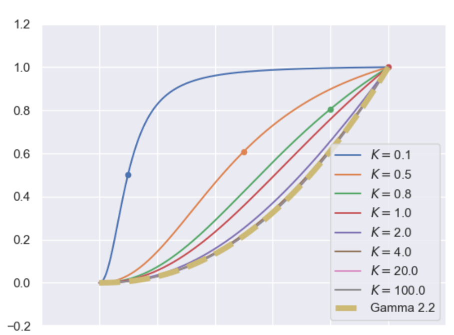

Figure 5. *Naka-Rushton curve with n=2.2 and increasing values of K, compared to a γ=2.2 power function.*

Figure 6. *Example of tone inversion: the patch that is darker in the real scene appears brighter in the picture (likely because of the interaction of lens glare and local tone-mapping algorithms).*

This model is quite effective at describing global monotonic OOTFs. However, we observed experimentally that it does not always fit well local OOTFs, especially in the dark parts of the image. Specifically, we know analytically that this model cannot handle tone inversions, because its derivative is always nonnegative, while cases of tone inversion do happen in real-world situations (see an example in Fig. 6).

We therefore augmented our model with a term $g_{\mathbf{p}}$ such that its derivative can be negative. We obtained the best results with a quadratic polynomial $g_{\mathbf{p}} = p_2 L^2 + p_1 L + p_0$.

In order to preserve the differentiability of the function, we want this term to blend continuously with Eq. 3. The blending term must be able to vanish the polynomial terms, which naturally suggests an exponential $\alpha(L) = e^{-\frac{L}{\lambda}}$.

Both new terms are of class C^{∞} , so differentiability is not an issue. In total, this introduces four new parameters to the model: p_0, p_1, p_2, and λ. However, we can reduce this number to three by imposing one of the roots at the maximum scene luminance: $g_{\mathbf{p}}(L) = p_{\text{A}}\left(\frac{L-p_{\text{r}}}{S}\right)\left(\frac{L}{S} - 1\right)$.

This gives a complete model defined as:

$$f_{\mathbf{t},\mathbf{p},\lambda}(L) = e^{-\frac{L}{\lambda}} \cdot g_{\mathbf{p}}(L) + \left(1 - e^{-\frac{L}{\lambda}}\right) f_{\mathbf{t}}(L) \tag{4}$$

There are nine parameters in total. Two of them can be determined analytically without optimization, all the others can be optimized together in a single joint model. Note that the scale factors G and S can be estimated analytically and do not need to be optimized. Furthermore, we are only interested in evaluating this over the range of luminance covered by the target chart, so no inference is made about the behavior in other parts of the dynamic.

Viewing Glare

Notice that Eq. 1, when implemented in a programming language, is developed as $\frac{L}{f(L)} \cdot f'(L)$, which presents an obvious risk of division by zero if $f(L) = 0$, *i.e.*, if there is clipping in the dark parts of the image. The natural solution for a programmer would be to add a small positive constant to stabilize the fraction:

$$\text{LCG}(L) = \frac{L}{f(L) + \nu} f'(L)$$

It is worth asking whether the value ν, which being homogeneous with $f(L)$ is measured in nit, also has a physical interpretation. If $f(L)$ is the light emitted by the display, then ν would be an additional constant light that the user perceives. This is consistent with the notion of *viewing glare*, the reflection of the environment light on the screen, illustrated in Fig. 7.

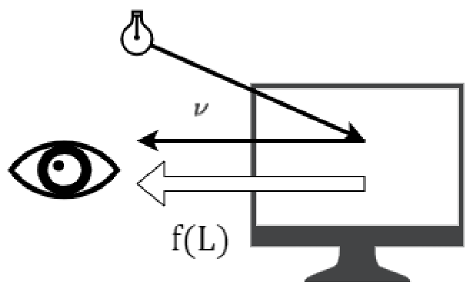

Figure 7. *Except in pitch darkness, an observer will always perceive a fraction of the ambient light as coming from the display, due to the reflectance of the screen.*

Interpretation of the LCG values

In order to provide an intuition of how to interpret the results of a device, let us analytically find what the LCG would look like for a few special cases.

The first case we can consider is that of a perfectly linear OOTF: $f(L) = A \cdot L$. While a linear OOTF is not necessarily ideal in the sense of user preference, it is a useful reference, as it represents contrast fidelity with respect to the ground truth. Ignoring the viewing glare term ν for the sake of simplicity, we can observe that:

$$\begin{aligned}\mathrm{LCG}(L) &= L \cdot \frac{f'(L)}{f(L)} \\ &= L \cdot \frac{\mathrm{D}[A \cdot L]}{A \cdot L} \\ &= L \cdot \frac{A}{A \cdot L} \\ &= 1\end{aligned}$$

That is, for a linear OOTF, the LCG is constant and equal to one. Conversely, we can say that in the neighborhood of any point such that $\mathrm{LCG}(L) = 1$, the OOTF behaves linearly. This is consistent with our previous observation that when $\mathrm{LCG}(L) = 1$ the contrast is preserved.

Another case worth considering is that of a gamma function: $f(L) = L^\gamma$. While $f(L)$ is an OOTF not an OETF, a gamma functions are still somewhat typical. We can determine that:

$$\begin{aligned}\mathrm{LCG}(L) &= L \cdot \frac{\mathrm{D}[L^\gamma]}{L^\gamma} \\ &= L \cdot \frac{\gamma L^{\gamma-1}}{L^\gamma} \\ &= \gamma\end{aligned}$$

That is, for a gamma OOTF, the LCG is constant and equal to gamma. Conversely, we can say that in the neighborhood of any point the OOTF behaves like a gamma function with the value of gamma given by the value of the LCG, which generalizes our previous result.

Scalar performance indicators

When the goal is to compare two devices, it is useful to define scalar metrics that correlate with some definition of performance defined with regard to the target application.

An intuitive solution would be to simply average the value of the LCG over its domain. However, we have observed experimentally that for this metric to correlate with a user's experience, the boosting of contrasts in one region should not compensate for compressions and inversions elsewhere. We thus define the *Average Contrast Compression* as:

$$C = \frac{1}{L_{\max} - L_{\min}} \int_{L_{\min}}^{L_{\max}} [\mathrm{LCG}(L)]_{-1}^{+1} \, \mathrm{d}L$$

As mentioned in the previous section, the value of the LCG in a point can be interpreted as the exponent of a local gamma function. This means that an Average Contrast Compression of C implies that the imaging system loses on average over its domain (*i.e.*, locally to where it was measured) as much perceptual contrast as an OOTF $f(L) = A \cdot L^C$.

A different approach would be to consider the range of luminance values in which the captured image is exploitable. To do so, let θ be the smallest acceptable value of LCG, chosen with regard to the target application, and let $\mathbb{I}$ be the largest interval such that $\mathrm{LCG}(L) \geq \theta \forall L \in \mathbb{I}$. We define the *Local Contrast Dynamic Range* as:

$$R = \log_2 \frac{\sup \mathbb{I}}{\inf \mathbb{I}}$$

Notice that because the LCG is computed locally, this metric only makes sense when compared to the dynamic in the ground truth of the chart over which the OOTF was estimated.

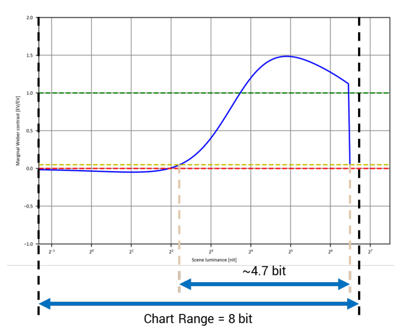

Figure 8. *The Local Contrast Dynamic Range is the range in bit of the largest interval over which the LCG is always above threshold. A reasonable example of threshold is θ=5%.*

Experimental results

We provided extensive experimental validation of our metric. While the metric can be meaningfully applied to a setup containing a single element, it is particularly pertinent in setups with multiple charts in different parts of the dynamic (see examples in Fig. 2.

In all situations, the results of the metrics match with the observation of expert analysts in terms of preservation, compression, lack, or inversion of contrast.

In Fig. 9, we present the results for a Portrait HDR setup. This setup is comprised of a Realistic Mannequin (on the left)

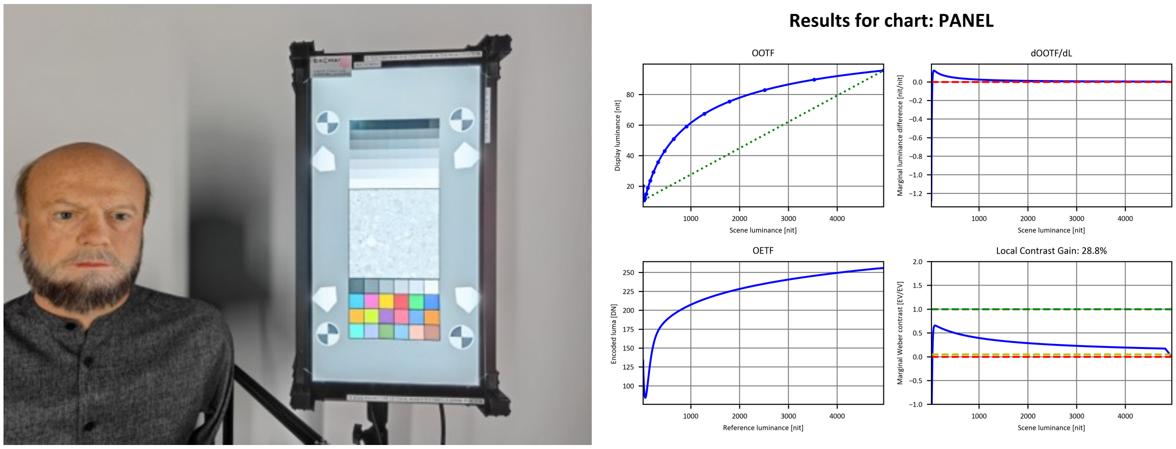

Figure 9. *Left: Portrait HDR setup. Right: Results measured on the Composite chart.*

and a back-lit Composite chart (on the right). The illuminance on the forehead of the realistic mannequin is 1000 lx, while the luminance of the light panel is 5500 cd/m^2. The measurement is performed on the gray scale of the Composite chart. We can see that the fitting of the OOTF has effectively captured the inversion of tones in the darkest patches (this is particularly visible when observing the OEFT). However, this only affects the first two patches, so it does not have a major effect on the overall LCG. In practice, the LCG peaks at a value around 50 % for very dark values, then decreases slightly but consistently, before saturating in the last patch. Overall, this gives an average contrast compression of 28.8 %.

In Fig. 10, we present the results for an HDR Composite setup. This setup is comprised of two back-lit Composite charts. The luminance of the left panel is 100 cd/m^2, while the luminance on the right panes is 6500 cd/m^2, which gives an approximate dynamic of 6 EV between them. Here, we can see that the two local OOTF are very different. On the darkest light panel, the first few patches are clipped, but then the contrast is considerably boosted, up to 150 %. Conversely, on the brightest light panel the OOTF is roughtly linear (LCG= 1) for the first few patches, but then the LCG linearly decreases until the OOTF reaches saturation.

In Fig. 11, we present the results for an Autofocus HDR setup. This setup is comprised, among other things, of a DeadLeaves chart and two back-lit Composite charts. The illuminance on the DeadLeaves chart is 1000 lx, while the luminance of both light panels is 1850 cd/m^2, which gives an dynamic of approximately 2 EV between the reflective and transmissive charts. Here, we can see that the two panels give very similar results to each other, while their behavior is noticeably different than that on the DeadLeaves chart. In particular, the DeadLeaves shows a contrast boost over most of its luminance range, while the tones over the light panels are generally compressed. Some clipping of the brightest parts of the DeadLeaves is visible.

Conclusion

In this article, we present a novel technique for quantifying the performance of an imaging system in terms of rendering of the local contrast.

The main novelty of our approach is its local nature. Several measurements have been proposed that provide some figure of merit associated to a device's preservation of contrast, however they generally assume that the imaging system is well-behaved, in the sense that it presents a single and well-defined transfer function. This makes them ill-suited to handle advanced ISP pipeline such as those of mobile cameras, where local tone mapping is the norm rather than the exception.

Also, most contrast measurements are based on gray scales, such as the Dynamic Range and Contrast Performance Indicator measurements in the P2020 standard [1] rely on the exact value of luminance emitted by each uniform patch. This requires a precise measurement to be provided each time, which is a time-consuming and error-prone endeavor. By estimating the OOTF with a fitting model, our technique can be robust to uncertainty in the ground truth and does not require recalibration each time.

Furthermore, the model we propose here for the OOTF is based on the well-known Naka-Rushton mode [3], which we augmented with terms to accurately describe the saturation of the highlights and clipping and inversions of the dark parts of the image.

Bibliography

[1] IEEE Standards Association P2020. *IEEE Approved Draft Standard for Automotive System Image Quality*. URL: https://sagroups.ieee.org/2020/.

[2] Anna Tigranyan et al. “Objective color characterization of HDR videos captured by smartphones: laboratory setups and analysis framework”. In: *Electronic Imaging* (2023).

[3] K. I. Naka and W. A. H. Rushton. “S-potentials from colour units in the retina of fish (Cyprinidae)”. In: *The Journal of Physiology* 185.3 (Aug. 1966), pp. 536–555. ISSN: 1469-7793. DOI: 10.1113/jphysiol.1966.sp008001.

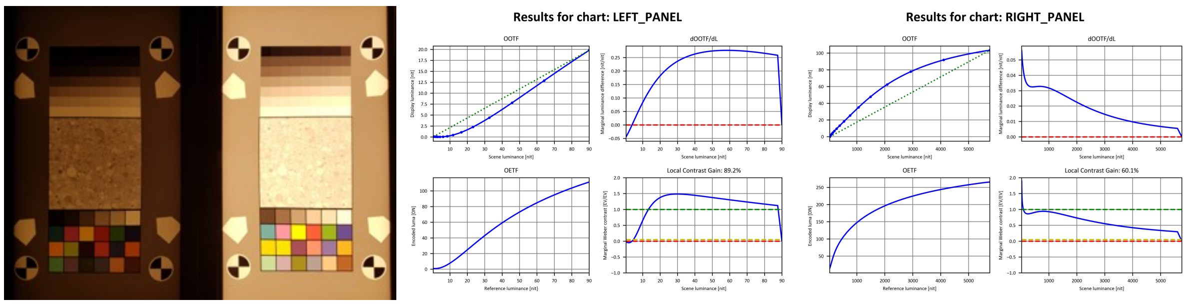

Figure 10. *Left: HDR Composite setup. Center: Results measured on the left Composite chart. Right: Results measured on the right Composite chart.*

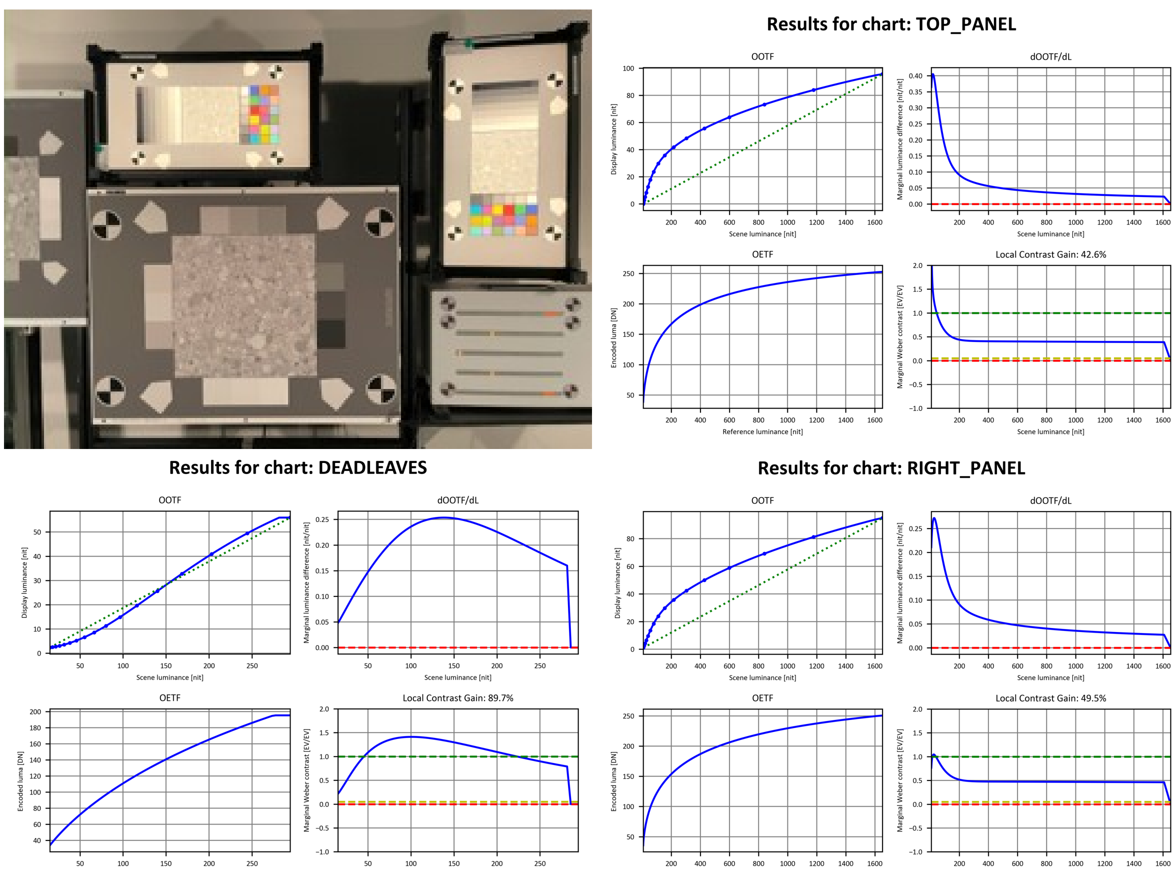

Figure 11. *Top Left: Autofocus HDR setup. Top Right: Results measured on the top Composite chart. Bottom Left: Results measured on the DeadLeaves chart. Bottom Right: Results measured on the right Composite chart.*