https://doi.org/10.2352/EI.2022.34.9.IQSP-313

New visual noise measurement on a versatile laboratory setup in HDR conditions for smartphone camera testing

Thomas Bourbon , Coraline Hillairet , Benoit Pochon , Frederic Guichard ; DXOMARK Image Labs, Paris, France

Abstract

Cameras, especially cameraphones, are using a large diversity of technologies, such as multi-frame stacking and local tone mapping to capture and render scenes with high dynamic range. ISO defined charts for OECF estimation and visual noise measurement are not really designed for these specific use cases, especially when no manual control of the camera is available. Moreover, these charts are limited to few measurements. We developed a versatile laboratory setup to evaluate image quality attributes, such as exposure, dynamic range, details preservation and noise, as well as autofocus performance. It is tested in various lighting conditions, with several dynamic ranges up to 7EV difference within the scene, under different illuminants. Latest visual noise measurements proposed by IEEE P1858 or ISO-15739 are not giving fully satisfactory results on our laboratory scene, due to differences in the chart, framing and lighting conditions used. We performed subjective visual experiments to build a quality ruler of noisy gray patches, and use it as a dataset to develop and validate an improved version of a visual noise measurement. In the experiments we also studied the impact of different environment conditions of the grey patches to assess their relevance to our algorithm. Our new visual noise measurement uses a luminance sensitivity function multiplied by the square root of the weighted sum of the variances of the Lab coordinates of the patches. A non-linear JND scaling is applied afterwards to get a visual noise measurement in units of JND of noisiness.

Introduction

Smartphones are nowadays one of the most popular ways to take photos, at least among non professionals. Being able to objectively test and compare smartphone cameras carries a lot of value to advise customers in their choice of smartphone. But these tests must be relevant to the customer's use cases. At DXOMARK Image Labs, we design repeatable laboratory setups to fairly test smartphone cameras in a wide range of conditions. Since in digital photography many quality attributes are interdependent, versatile laboratory setups are necessary to measure several quality attributes on the same scene and evaluate the compromise between them. For noise evaluation, existing ISO standard charts, like the ISO-14524 [1] 12-patch test chart for OECF estimation and visual noise measurement (ISO-15739 [2]), are not relevant in our case as they are limited in terms of number of measurement[1]. Adding to that, the existing visual noise measurements proposed by IEEE P1858 (CPIQ) [3] or ISO-15739 standards do not give entirely satisfactory results when used with a different framing, different chart or in more challenging HDR conditions. Moreover, these standards generally apply for images with perfect white balance, which is not always the case on real images taken with smartphones.

[1]The ISO 19093 standard proposes a chart intended for multiple measurements, however it does not always trigger HDR pipelines in all automatic cameras such as smartphones.

Existing setups and their limitations

To test imaging systems, some versatile setups are available to measure their capacity to render details, control noise and so on.

Deadleaves setup

In 2012, DXOMARK introduced the Deadleaves chart in its testing protocol. This chart is described in the IEEE P1858 standard [3]. The presence of the Deadleaves pattern and the 12 grey patches allows to measure acutance as well as target exposure and noise on this chart. In 2017, DXOMARK developped the Deadleaves setup around this chart to add measurements on autofocus performances and to complexify the content of the test scene. The Deadleaves setup can be shot in two holding conditions: tripod and handheld. The tripod case is static, to simulate a perfect conditions photography use case. The handheld case uses a moving platform to simulate the small hand shake occurring when a person is taking a picture. Movement is also added in the scene with a small test chart (similar to the Dead leaves chart), which is moving up and down in a controlled constant velocity.

The shooting procedure of the Deadleaves setup is automated thanks to a small defocus chart mounted on an automated moving object (AMO) placed in front of the camera and linked to a triggering system. This movement is detected by lasers and triggers the capture of an image by the smartphone.

The Deadleaves setup is shot in a laboratory with a controlled lighting system, featuring several standard illuminants: Daylight D65, Fluorescent (TL84), Tungsten (A) and Horizon (H). The intensity of the illumination can be configured according to different lux levels: 1000 lux, 300 lux, 100 lux, 20 lux, 5 lux and 1 lux.

Thirty-five pictures are taken for each shooting condition, being the combination of a holder (tripod or handheld), an illuminant and a lux level. This allows to evaluate the capacity of the device to deliver a consistent image quality over consecutive shots.

Limitations

One of the main limitations of the Deadleaves setup is that the dynamic range is rather low in the scene. There are no particularly bright parts, nor dark areas. However, a lot of photographed scenes in real life have content with higher dynamic range, be it

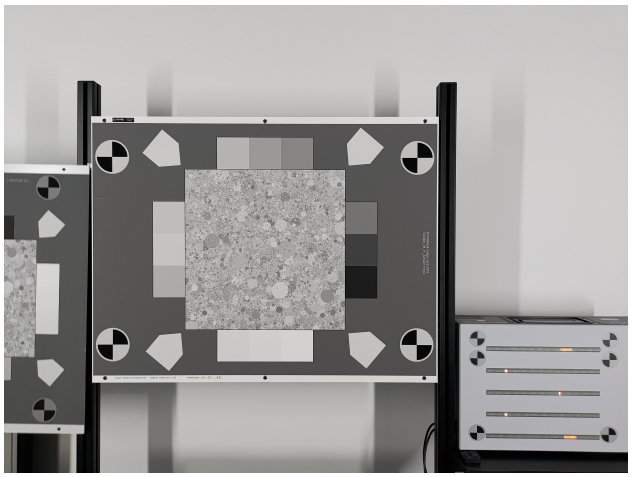

Figure 1. *Example of picture taken with a smartphone on the Deadleaves setup.*

in an outdoor setting with the sun in the background, or indoors facing a window. In case of a high dynamic range scene, given the sensor limitation in a smartphone, a single shot with a single exposure time is often not enough to capture details in both bright and dark areas of the scene. In these cases, smartphone may activate image multi-frame fusion techniques to capture the whole dynamic of the scene. These consist in taking several captures of the same scene, with potential different exposure times, to then merge them together, with the general idea to keep the parts with the most details for each image. This can bring image artifacts such as halo or tone compression, for instance.

As it is a rather common use case, we needed a laboratory setup with a higher dynamic range to be able to test fairly, in controlled conditions, the image quality given those kind of high dynamic range technologies.

A new versatile laboratory setup

In the light of the limitations of the Deadleaves setup, and to challenge even more the ever improving image fusion algorithms, we decided to create a new laboratory setup called the AF-HDR setup. The aim of this setup is to be challenging in terms of lighting conditions and dynamic range, but also to have enough elements in it for us to measure as many image quality attributes as possible on this one setup.

The AF-HDR setup

The AF-HDR setup was introduced into DXOMARK's camera testing protocol in 2020. An example of picture taken with a smartphone is given in Figure 2.

It inherits all the elements of the Deadleaves setup presented previously (Figure 1). Added to it are 2 light boxes placed on top and on the right of the center chart. These light boxes have Plexiglas slides on them, printed with a Deadleaves pattern, a color checker chart and a "grey gradient pattern". The light boxes can be controlled in intensity and color temperature via software. The moving object on the left of the chart is present in the tripod condition for the AF-HDR setups, however it is not moving in this condition.

Like for the Deadleaves setup, the AF-HDR setup is shot with different illuminants and lux levels. The addition of the light boxes allows to create controllable challenging dynamic range conditions. These conditions are identified by their Exposure Value differences between the Deadleaves chart and the light boxes. The EV differences are computed as follows:

$$\Delta EV = \log_2\left(\frac{L \cdot \pi}{I}\right) \tag{1}$$

where L is the luminance emitted by the light boxes, measured in cd/m² facing the light box, and I is the illuminance incident on the Deadleaves chart, measured in lux, facing the lighting system, away from the chart.

The shooting of this setup is also automated, the same way as it is for the Deadleaves one.

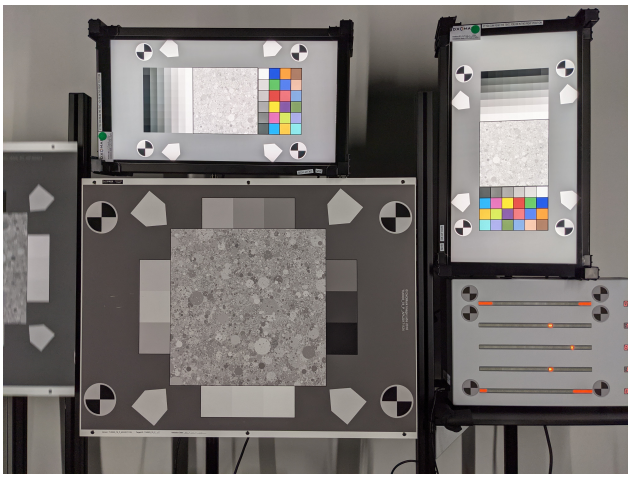

Figure 2. *Example of picture taken with a smartphone on the AFHDR setup.*

Usage and measurements

The AF-HDR setup has been designed as a multi purpose setup, allowing to measure, for now, three image quality attributes (exposure, texture and noise) and autofocus performances.

For exposure, we measure target exposure and dynamic range capabilities thanks to the 12 gray patches and the light boxes.

The performances of the auto-focus system of the device can also be evaluated thanks to the automatic system to trigger the capture of an image and the LED box. We can measure the speed, the accuracy and the repeatability of the autofocus system.

With the Deadleaves chart in the setup, we can perform as well the texture and noise measurements that were performed with the Deadleaves setup.

Existing visual noise measurement

Several visual noise measurements exist in the literature. The most common are the one described in ISO-15739 [2] standard and the one described by CPIQ [3]. In 2017, DXOMARK derived from them another visual noise measurement (henceforth referred as DXOMARK 2017 algorithm). The algorithms details and differences are visible in Figure 3.

Figure 3. *Details of the four visual noise algorithms discussed in this article. CPIQ algorithm is described in [3]. ISO one is described in [2], [4] and [5].*

The Deadleaves setup's conditions were close enough to what was summarized in the ISO-15739 standard to give accurate results with the DXOMARK 2017 visual noise algorithm. However once transferred on the AF-HDR setup, the results of the measurements, be it from DXOMARK 2017 algorithm or from the standards, were sometimes quite unexpected and not correlated with perceptual observations. This led us to study the current model, understand its limitations causing the unexpected results, and to finally develop a new visual noise measurement compatible with high dynamic range laboratory setups.

New visual noise measurement

Requirements

Our AF-HDR setup diverges quite considerably from shooting conditions recommended by CPIQ [3] and ISO-15739 [2] standards for visual noise measurements. For one, we are not using an OECF chart, just the gray patches extracted from it and rearranged on our own chart. The framing used is also different. Our chart is not centered in the scene. Finally, the content of the overall scene is different. As two light boxes are added in the field, this changes significantly the scene content, it impacts the lighting of the scene and also the reaction of the camera.

Moreover, the visual noise measurements available have been designed for patches with lightness value not deviating much from $L^* = 50$. We wish to extend this, to be able to measure noise in darker and brighter parts of an image.

With all these departures from the standards, we understand that the results obtained from existing visual noise algorithms cannot give satisfying results. We needed to adapt these algorithms to our testing conditions. The new measure needed to:

- Be robust to the scene exposure, i.e. take into account the impact of patch environment conditions (the lightness of the patch's surrounding) compared to the mean lightness of the patch.
- Be robust to different noise types and shapes. Noise patterns in digital photography, with automated image processing, can be very different from photon shot noise or read-out noise, but its still need to be assessed properly by the measurement.
- Be robust to HDR condition side effects. In HDR conditions, image fusion or local processing can be applied to the images, resulting sometimes in artifacts like halo. These phenomena might not be considered as noise by some observers. Although discussions are ongoing on the subject in the community, we decided that these artifacts would not be taken into account by the measurement.
- Provide a visual noise metric in Just Noticeable Difference (JND) unit to be easily understandable and more reusable. It can be either JND of quality loss (quality JND) or JND of perceived noise in the image, what we call noisiness (attribute JND) [6].

Perceptual validation Dataset

Independently from the solution that we propose, we needed to be able to validate that our new visual noise algorithm would give results correlated to human perception. To do so, we needed a dataset of grey patches, annotated relatively to their noisiness. The dataset, whose construction is described in this section, is publicly available at the following link: https://share.dxomark.com/s/yRqAdGAbyx7m2Ac

CPIQ visual noise toolbox (see [3], annex D) proposes a dataset of patches annotated in quality JND. This unit refers to a defined set of images, and refers to a virtual degradation of the image quality. If we were to use new real images taken with smartphone cameras, we would part from the SQS2 base of quality JND (described in [7]). On the other hand, redoing an entirely new base of quality JND on a new set of images would have been to costly.

We thought more adapted to work with a JND of noisiness scale for our dataset with real images taken with smartphone cameras. Thus we set to create our own ruler of JND of noisiness based on the ISO-20462 triplet experiment [8]. We extracted grey patches from images taken with several smartphones of our database. We tried to have the most variety of noise types and intensities as possible, while having a mean L^* close to 50 and avoiding images with strong white balance casts. We ended up with 35 patches of size 143 x 143 pixels. We also took in our dataset all the CPIQ patches and a perfect grey patch.

The first step was to do a fast categorical sort, splitting the patches into several groups that do not have too much variation in it. Indeed, the triplet comparison must be done only between images that do not have too much differences in JND (no more than 6JND of range, but ideally less than 4JND of range). We supposed that our final ruler range will be of about 0 to 15 and thus we split the data in 4 groups of 9 patches, chosen such that:

- The patches in the group look well distributed in level of noise
- No patch is obviously better or worst that all others
- A patch (or two for the middle groups) is in common with the next and/or precedent group.
- The first group must contain the perfect grey patch

Triplet comparison

We then applied the triplet comparison method on each group. To do so we developed a hardware and a software tool, similarly to what has been done in [9]. For the hardware, we set up a viewing lab with the following characteristics:

- 933 cm between the display and the observer. The distance is guaranteed by a bar on which the analyst has to rest their forehead during the whole experiment
- An Eizo CS2420 computer display, color calibrated with a D65 whitepoint with a luminance at 75 cd/m²
- A background illumination done with a light box set to D65, to reduce eye stress during the experiment

For the software, we developed a tool to perform the experiment, following the ISO-20462 triplet comparison process [8], on which images are displayed without magnification. For each group, patches are presented by subgroups of 3, on a $L^* = 50$ background. Twelve comparison per groups are made, as it covers each possible paired comparison exactly once, without redundancy. For each patch, there are five evaluation propositions: Good, Acceptable, Just acceptable, Unacceptable and Poor. Analysts are asked to evaluate the noisiness of a patch relatively to the other two, while ignoring variation of exposure and white balance coming from the real images. The levels are not absolute, they must reflect the relative ranking, in terms of noisiness, between the shown patches. A group of 31 persons did the experiment, composed of 25 image quality experts and 6 non experts, all from DXOMARK Image Lab company.

The output is 4 relative JND rulers with their uncertainties. We then needed to combine them to get one absolute ruler. We have one reference image, the perfect grey patch, for which we know the relative JND value in the 1rst group and the absolute JND of noisiness value, which is 0. Per group, for the rest of the patches, we start by identifying the reference patch. It is the one that is both known within the absolute ruler and relative group ruler (the patch that was in common between this group and the previous one). Its JND values and uncertainties in both the absolute and relative scales are $r_{abs} \pm \sigma_{r_{abs}}$ and $r_{rel} \pm \sigma_{r_{rel}}$ respectively. We then build the absolute value and uncertainty of the rest of the patches of the group :

$$x_{abs,i} = x_{rel,i} + r_{abs} - r_{rel} \tag{2}$$

$$\sigma_{abs,i} = \sqrt{\sigma^2_{rel,i} + \sigma^2_{r_{abs}} + \sigma^2_{r_{rel}}} \tag{3}$$

where, for patch i, $x_{abs,i}$ and $x_{rel,i}$ are respectively the absolute and relative JND values, and $\sigma_{abs,i}$ and $\sigma_{rel,i}$ are respectively the absolute and relative JND uncertainties.

The generated ruler can be consulted in figure 6. This full ruler is comprised of 33 patches. For the rest of our experiment we need a lighter version of the ruler with less patches. We decided to remove patches with too similar JND of noisiness values in order to increase the analysis speed and reduce the analyst stress (it would be more difficult for them to take a decision if images are too close in the ruler). We then created what we called the decimated ruler which is composed of 19 patches selected among the full ruler patches.

Patch environment study

With this first reference ruler created, we conducted an extra perceptual experiment to study the impact of the environment of the patches on our perception of the level of noise, to determine if it needed to be taken into account in future perceptual experiments as well as our visual noise metric. We conducted a softcopy ruler quality based analysis [7] on an homemade software on 8 patches (7 from our ruler and 1 validation patch from CPIQ ones) for which we changed the surrounding grey level to 7 different levels. This makes a total of 50 evaluations. The ruler used for this experiment is the decimated ruler mentioned previously.

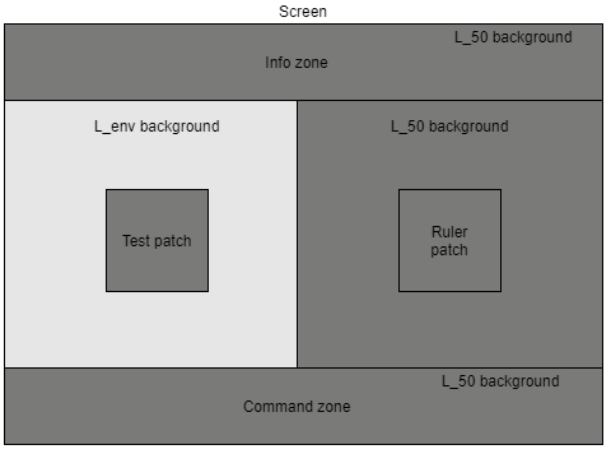

Figure 4. *Layout of the software used for the softcopy ruler quality based analysis on the impact of the patches environment, with* $L_{env} \in \{0, 17, 33, 50, 67, 83, 100\}$

The results of this experiment show that modifying the environment contrast in which we look at the patches has, overall, only a small impact on our perception of the noise (see Figure 5 the blue curve, less than 1 JND delta even with a $\Delta L = 50$). As a consequence we decided not to take it into account in our following experiments and visual noise metric design.

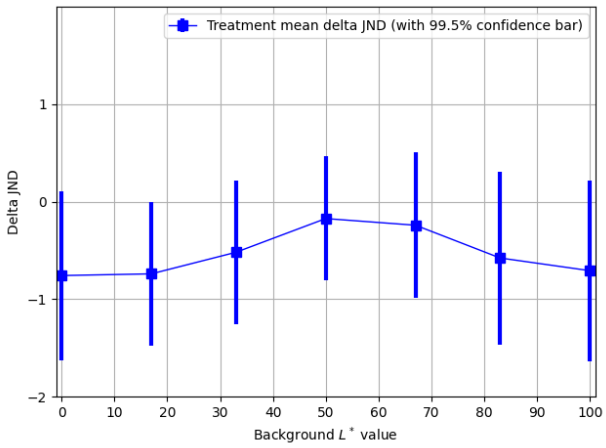

Figure 5. *Results of the environment experiment. JND difference between the value given by analysts and the JND value of the patch in the ruler.*

Extending the ruler

We then decided to extend our ruler, with patches extracted from more varied conditions of our AF-HDR laboratory setup. We took 60 new patches, with various types of noise, in shape and intensity, and with different lightnesses (not only $L^* = 50$) and chroma. We also took 15 patches with specific HDR artifacts, like halo, and 2 validation patches. A softcopy ruler quality based analysis is done by 40 analysts from DXOMARK, 33 image quality experts and 7 non-experts. We can see on the figure 6 that the 99.5% error interval are greater in the analyzed scenes than in the reference ruler dataset. It is explained by the fact that analysts have more difficulties to reduce the notion of visual noise to only one dimension and place it on the ruler. In that way the softcopy ruler quality based method is less precise (while faster) than the triplet comparison method used to build the first reference ruler dataset.

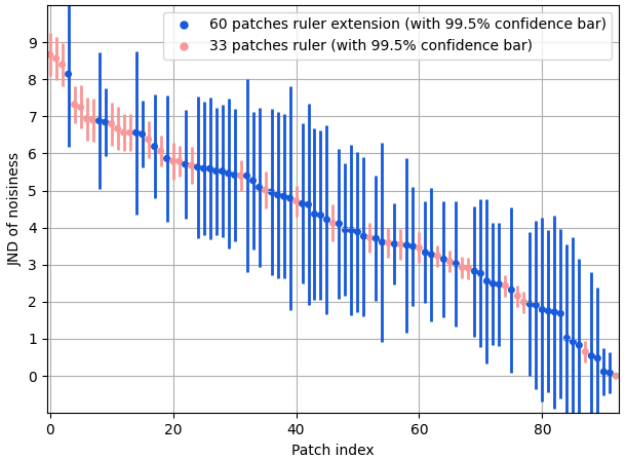

Figure 6. *DXOMARK full setup dataset, composed of 93 noisy patches annotated in JND of noisiness.*

We decided to keep this extended ruler and to take into account the confidence that we have on each patch to design and test our new visual noise measurement.

Defining a new visual noise measurement

Linearization

The first step of the visual noise measurement is to linearize the input image in order to be able to use linear color spaces such as CIELAB. In DXOMARK 2017 algorithm, visual noise was considered scene referred. An OECF estimation then inversion was used. During the process, a factor was applied such as the mid-range patch ends up at 50% of the dynamic after linearization. ISO-15739 recommends to use visual noise for output referred noise evaluation. Since OECF estimation can come with measurement difficulties (especially with smartphones doing some local processing) and following ISO recommendation, we use a simple gamma linearization for our new visual noise measurement.

Frequency filtering

The next step is to apply the contrast sensitivity function (CSF). The CSF are 3 frequency filters, which describe the human perceptual sensitivity to frequencies. Several definitions of these curves exist. In our previous measurement, we used the Johnson & Fairchild CSF, as used in CPIQ. For the new measurement, we decided to follow a recent ISO-15739 revision study [4] [5] showing that it is preferable to use the Johnson & Fairchild CSF with the luminance normalized to 1.

In addition to CSF and MTF frequency filters, a third one is considered to remove image artifacts that are not considered as noise by a human observer, but which can have a huge impact on the L^*, a^*, b^* channel variances. Our previous measurement used the high-pass filter designed by CPIQ standard to remove shading effect. On the other hand, ISO-15739 standard does not include this kind of filtering in the visual noise calculation. We want to extend our previous high-pass filter to also deal with halo effects which are more likely to appear in HDR conditions of our new setup. The proposed filter (defined in cycle-per-degree) is as follow:

$$HPF : f \mapsto \frac{1}{1 + e^{-w_f(f - f_c)}} \tag{4}$$

with f_c the cutoff frequency and w_f the slope parameter.

We observed that the error on the JND estimation was reduced by about 30% on patches affected by halo and shading.

Visual noise formula

For the visual noise metric in itself, our previous measurement, as well as CPIQ and ISO-15739 propositions, are all based on a weighted sum of the L^*, a^*, b^* channel variances. CPIQ's formula, using the base-10 logarithm of the weighted sum of the variances and covariances, has been found to sometimes lead to math domain errors because of the negative coefficient in front of σ_b^2 and because it can happen that $\sigma_{La} \leq 0$. We decided to go with a formula similar to the one used in the ISO-15739 standard revision proposition from 2019 and 2021 [4] [5]:

$$V = \sqrt{\sigma_L^2 + w_a \sigma_a^2 + w_b \sigma_b^2} \tag{5}$$

with w_a and w_b weighting coefficients for a and b component noise variances.

However we wanted our algorithm to take into account that an observer is not perceiving the noise the same way for component a and b depending on the signal. In our case the signal is simply the mean values of the image pixels across the color channels: $\mathbb{E}(L^*)$, $\mathbb{E}(a^*)$ and $\mathbb{E}(b^*)$. Looking at the results from the formula above, we noticed a clear dependency between $\mathbb{E}(L^*)$ and the generalized error, while on chroma channels the dependency is quite limited. Thus we concentrated our study on L channel, to introduce a new luminance sensitivity function. Neither CPIQ nor ISO-15739 standards use this kind of function.

To be able to properly study the sensitivity of the noise observation with respect to the patch signal $\mathbb{E}(L^*)$, we conducted a new perceptual experiment based on a softcopy ruler quality analysis. We needed a smaller ruler than our full dataset, so we selected 25 patches from our full setup dataset, providing an appropriate JND values sampling and range. To generate the set of patches to be evaluated, we first selected a set of 8 patches with a known JND values for a particular mean L^* value (in our case it is for $L^* = 52$ for all patches). Then for each patch we applied an offset on the L^* channel, paying attention not to clip some pixels, so

that we end up with new patches having the same variance σ_{L^*} as the original one, but their signal $\mathbb{E}(L^*)$ is in: 17, 22, 27, 32, 42, 62, 72, 82. Our experiment is then composed of 7 patches with 9 treatments and a validation patch, for a total of 64 patches. This dataset will further be called the luminance dataset.

We asked 11 analysts to evaluate the perceptual noise of these 64 patches by comparing them to the ruler patches. For each of the 7 patches, we analyzed the evolution of JND values assigned by the analysts with respect to the patch signal $\mathbb{E}(L^*)$. We then averaged, for each single L^*-mean value, the normalized JND values over all the patches to get a representation of the noise perceptual evaluation sensitivity with respect to the L^*-mean value. (see Figure 7)

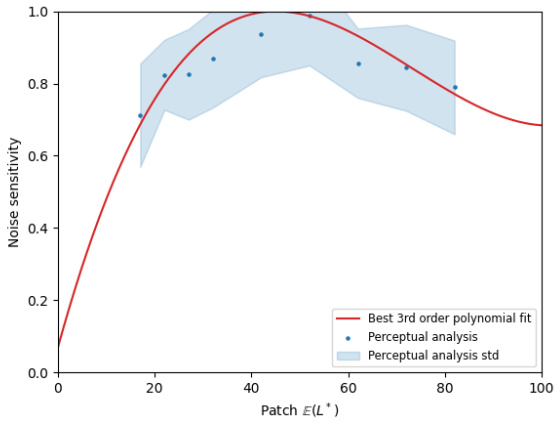

Figure 7. *Results of the L^*-mean influence study averaged over the 7 patches of the study, with uncertainties.*

It shows that the sensitivity is less important for low or high L^*-mean values and reaches a maximum at $L^* = 50$. The sensitivity varies by about 20% for $L^* = 20$ or $L^* = 80$. To take into account this sensitivity to the L^*-mean value and extrapolate the missing values at the extremities of the L^* range, we fitted a 3rd order polynomial function on our luminance dataset. Applied to our full setup dataset, we noticed that the algorithm residual error was about 15% lower than without this luminance sensitivity function, which confirmed the improvement due to the sensitivity function.

The polynomial sensitivity function is defined as follow:

$$S_L : L \mapsto a + bL + cL^2 + dL^3 \tag{6}$$

with $a = 0.068641535$, $b = 0.048546862$, $c = -7.7856422e-4$ and $d = 3.5483275e-6$

The new proposed visual noise metric is as follow:

$$V = S_L(\mathbb{E}(L^*))\sqrt{\sigma_{L^*}^2 + w_a\sigma_a^2 + w_b\sigma_b^2} \tag{7}$$

with $\mathbb{E}(L^*)$ the mean of lightness L^* on the patch.

JND mapping

The output of this metric is not in a unit easily understandable in terms of human perception, so we need to map it to a JND of noisiness scale. While ISO-15739 does not provide a JND mapping, CPIQ proposes an integrated hyperbolic incremental function (IHIF) for their JND scale transformation. Looking at our metric defined above (equation 7), we notice that its growth was quite correlated with the JND values growth of our full setup dataset. However the relation between them did not seem linear. Looking at the data, we observed that a non-linear mapping function F_{JND} should have the following properties:

$$F_{JND}(V) \underset{V\mapsto 0}{\sim} V^3 \text{ and } F_{JND}(V) \underset{V\mapsto inf}{\sim} V^{\frac{1}{3}} \tag{8}$$

Therefore we propose to use a JND mapping as follows:

$$F_{JND} : V \mapsto \frac{w_1 V^3}{1 + w_2 V^{\frac{8}{3}}} \tag{9}$$

with w_1 and w_2 are two shaping parameters.

Results, discussion and further improvements

Numerical results

To evaluate the performances of each algorithm, we computed what we called heteroscedastic residual sum of squares (*HRSS*) on our full setup dataset in order to take into account the variable standard deviation σ_k of the perceptual evaluation for each image in our full setup dataset.

$$HRSS = \sum_k \left(\frac{e_k}{\sigma_k}\right)^2 \tag{10}$$

with e_k the difference between the visual noise computed each algorithm (with JND mapping) and the JND value of the patch in our dataset.

We optimized the parameters of the new visual noise algorithm on our full setup dataset through a least square method, minimizing the *HRSS*. The optimal numerical values found for the parameters are given in table 1. All algorithms performances are then given in table 2. We can observe that our new proposed algorithm is giving significantly better results than the others on the full setup dataset.

Parameter	Value
w_f	13.5
f_c	1.4
w_a	0.04977
w_b	0.2790
w_1	323
w_2	46

Table 1. *DXOMARK new algorithm optimal parameters when optimized on the full setup dataset. Parameters are the ones described in equations 4, 7 and 9.*

Discussion and further improvements

The numerical results were obtained through a joint optimization of several parameters.

Algorithm name	HRSS	RMSE
DXOMARK 2017 algorithm	7861	2.19
CPIQ	4456	2.10
ISO 15739 (2021 tentative) with linear JND mapping	3102	1.99
DXOMARK new algorithm	786	0.88

Table 2. *Algorithm performances on DXOMARK full setup dataset. DXOMARK new algorithm performance is given when using the parameters defined in table 1*

Values for w_a and w_b deviates quite significantly from the ones found in [5]. This may be explained by the fact that the patches of the dataset come from images taken with camera equiped with automatic white balance and no additionnal white balance correction were applied. Therefore the remaining cast for some illuminants might influence the perception of noise. Just as we included in our algorithm a luminance sensitivity function, the same could be done for a^* and b^* channels as our sensitivity to noise (especially chroma noise) is probably different depending on the white balance cast of the patch (pointed out by [10] and [11]). However it should be noted here that the influence of the proposed values for w_a and w_b is actually quite small on the residual error. Incorporating w_a and w_b values from [5] actually increase marginally the heteroscedastic residual error while it decreases slightly the residual error on the dataset.

We studied the influence of the patch luminance on our noise sensitivity. If this is valid for most of signals $\mathbb{E}(L^*)$, some concerns could be raised for very low or very high $\mathbb{E}(L^*)$ values as the screen performance (quantization or imprecisions for instance) could be a non-negligible part of what people see during the ground-truth experiments.

It could be also interesting to include into the model the spatial correlation of the noise. With the current model an image with a white noise and an image with a pattern noise could be ranked the same as long as the variances in our filtered $L^*a^*b^*$ spaces, and the signal $\mathbb{E}(L^*)$, are the same. However for an observer the correlated noise is more likely to be considered more "noisy" that the white noise one. In our study incorporating spatial correlation indicators like Moran's I or band-pass frequency filter were tested but not included in the formula as we did not find any significant correlation between the indicator and the generalized error.

Finally, with the proposed algorithm, the results are within the ruler dataset uncertainties. This means that very few improvement can be made from this point without updating our dataset, because we might only try to improve the dataset "noise" (which generally leads to overfitting) instead of really improving the model.

Conclusion

Image processing pipelines in smartphones are becoming more and more complex in order to achieve the best image quality possible. Depending on the content of the scene being shot, compromises have to be made between different image quality attributes like texture, noise, dynamic range and artifacts. That is why we need versatile laboratory setups on which we can measure these different attributes at the same time and under a large variety of light conditions. In this article we propose a new visual noise algorithm suited for noise evaluation on images of both standard and high dynamic range scenes. To validate our algorithm, we built a dataset of images of our laboratory setup captured with smartphone cameras. The dataset was annotated on a JND of noisiness scale and is provided to the community. The new visual noise algorithm derived from our studies is dependent of the lightness of the patch measured, and comprises a high-pass filter to remove low frequency content not perceived as noise. Finally a JND mapping is proposed.

Thanks to this new measurement, we are now able to have a perceptually accurate measurement of visual noise in HDR scenes conditions. It raises up to four the number of image quality attributes that we can objectively evaluate on our HDR setup (exposure, focus, texture and now noise).

The results of these study were shared, discussed within the ISO/TC42/WG18 Electronic still picture imaging working group in the context of the AWI 15739 standard.

References

[1] I. 42, "Iso 14524:2009 photography — electronic still-picture cameras — methods for measuring opto-electronic conversion functions (oecfs)," tech. rep. (2009).

[2] I. 42, "Iso 15739:2013 photography — electronic still-picture imaging — noise measurements," tech. rep. (2013).

[3] "Ieee standard for camera phone image quality," *IEEE Std 1858-2016 (Incorporating IEEE Std 1858-2016/Cor 1-2017)* , 1–146 (2017).

[4] D. Wueller, A. Matsui, and N. Katoh, "Visual noise revision for iso 15739," *Electronic Imaging* **2019** (2019).

[5] A. Matsui, N. Katoh, and D. Wueller, "Experimental study for revising visual noise measurement of iso 15739," (2021).

[6] I. 42, "Iso 20462-1:2005 photography — psychophysical experimental methods for estimating image quality — part 1: Overview of psychophysical elements," tech. rep. (2005).

[7] I. 42, "Iso 20462-3:2012 psychophysical experimental methods for estimating image quality — part 3: Quality ruler method," tech. rep. (2012).

[8] I. 42, "Iso 20462-2:2005 photography — psychophysical experimental methods for estimating image quality — part 2: Triplet comparison method," tech. rep. (2005).

[9] E. W. Jin, B. W. Keelan, J. Chen, *et al.*, "Softcopy quality ruler method: implementation and validation," in *Image Quality and System Performance VI*, S. P. Farnand and F. Gaykema, Eds., **7242**, 53 – 66, International Society for Optics and Photonics, SPIE (2009).

[10] J. Kuang, X. Jiang, S. Quan, *et al.*, "Perceptual color noise formulation," in *Image Quality and System Performance II*, R. Rasmussen and Y. Miyake, Eds., **5668**, 90 – 97, International Society for Optics and Photonics, SPIE (2005).

[11] D. Baxter, J. Phillips, and H. Denman, "The subjective importance of noise spectral content," in *Image Quality and System Performance XI*, S. Triantaphillidou and M.-C. Larabi, Eds., **9016**, 12 – 23, International Society for Optics and Photonics, SPIE (2014).

Author Biography

Thomas Bourbon received his Master's degree in engineer-

ing from Institut d'Optique Graduate School (2015). He joined DXOMARK Image Labs in 2019, as an image quality engineer. Coraline Hillairet received her Master's degree in engineering from Institut d'Optique Graduate School (2018). She joined DXOMARK Image Labs in 2018, as an image quality engineer. Benoit Pochon received his Master's degree in engineering from Centrale Supelec (2001) and his Master's degree in Electrical Engineering from GeorgiaTech University (2001). After several years working in the signal processing domain, he joined DXOMARK Image labs in 2017, as image science director.