https://doi.org/10.2352/ISSN.2470-1173.2020.9.IQSP-166

DXOMARK Objective Video Quality Measurements

***Emilie Baudin*[1]*, François-Xavier Bucher*[2]*, Laurent Chanas*[1] *and Frédéric Guichard*[1]**
1: DXOMARK, Boulogne-Billancourt, France
2: Apple, Cupertino, CA, USA

Abstract

Video capture is becoming more and more widespread. The technical advances of consumer devices have led to improved video quality and to a variety of new use cases presented by social media and artificial intelligence applications. Device manufacturers and users alike need to be able to compare different cameras. These devices may be smartphones, automotive components, surveillance equipment, DSLRs, drones, action cameras, etc. *While quality standards and measurement protocols exist for still images, there is still a need of measurement protocols for video quality. These need to include parts that are non-trivially adapted from photo protocols, particularly concerning the temporal aspects. This article presents a comprehensive hardware and software measurement protocol for the objective evaluation of the whole video acquisition and encoding pipeline, as well as its experimental validation.*

Introduction

Still image quality measurement is a well covered subject in the literature. There are several ISO standards on photography, as well as the IEEE-P1858 CPIQ standard [3]. The CPIQ metrics with JND mapping/perceptual comparison are described by Baxter *et al.* [1] and Jin *et al.* [5]. Other objective quality measurements for still images with correlation to perceptual measurements or JND mapping are also published. For instance, an objective exposure quality measurement calibrated on the human perception has been proposed by He *et al.* [2].

As for motion pictures, all still image quality metrics can be applied to individual video frames to provide a first estimation of the video quality from static scenes. However, light changes are very common in video use-cases and an evaluation of the temporal behavior is required as well. Light changes can occur for example when walking from indoor to outdoor, when filming a show where the lighting constantly changes, when a car goes out of a tunnel, when panning with different light sources in the scene, *etc*. In this article we consider end-user cameras that are used in auto mode. The camera must automatically adjust all parameters to the environment. The most visible quality attributes for still images are exposure, white balance, and color; the same applies to video, and we need to assess the convergence of these attributes during light changes. Furthermore, some artifacts that exist in photo also exist in video but their impact on the overall perception is different. For instance, the temporal aspect of the noise can bring a different perception from the spatial aspect and must therefore be tested separately.

Degradation introduced by video encoding and transmission is widely studied by VQEG, with correlations to Mean Opinion Score (MOS) [10][11]. These measurements do not evaluate any step before encoding and they do not include shooting conditions, thus they cannot be considered comprehensive video quality measurement protocols. Temporal aspects of video quality begin to be studied; for instance, a perceptual study of auto-exposure adaptation has been presented by Oh *et al.* [8]. To the best of our knowledge, there is not as yet a measurement protocol with correlation to subjective perception for video image quality.

The goal of this work is to provide a measurement protocol for the objective evaluation of video quality that reliably allows to compare different cameras. It should allow to automatically rank cameras in the same order a final user would do it. In this article, we focus on three temporal metrics: exposure convergence, color convergence, and temporal visual noise.

The novelty of this work includes a comprehensive approach that jointly designs hardware (closed-loop controlled lighting, custom-printed charts, *etc*) and software, to provide a video quality measurement protocol. These allow to test different lighting conditions representative of real use-cases (indoor, outdoor, low-light). Also, the measurements are done in auto mode and in the final user perspective, thus the measurement protocol includes the whole video acquisition and encoding pipeline (optics, sensor, ISP, codec).

Proposed Metrics

Most still image quality metrics can be adapted to video by processing each frame independently and then extracting quality statistics. For instance, we can compute a metric on each frame of a video and assign its average value to the whole sequence. Thus we can measure the target exposure, the white balance, the static color fidelity, the texture preservation, the spatial visual noise, the modulation transfer function (MTF), the vignetting, the ringing, the distortion, and the lateral chromatic aberrations. These metrics are provided in the Analyzer solution and in others commercial distributions.

In this article, we focus solely on the temporal aspects of the video quality evaluation: exposure and color convergence on light transitions, and temporal visual noise on stable light. All measurements are performed on custom-printed charts under controlled lighting conditions.

Exposure and Color on Lighting Transitions

Lighting transitions are a challenging occurrence for video quality that happen when either the lighting in the scene or the framing changes. They can occur when filming a show where the light changes constantly, when walking from outdoor to indoor or from indoor to outdoor, when driving in or out of a tunnel, when panning, *etc*. In auto mode, devices need to adapt the exposure and the white-balance to the change of lighting and the resulting transition can be unpleasant to the viewer. In this section we propose metrics to evaluate the performance of auto-exposure and

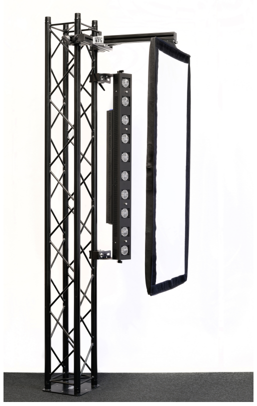


Figure 1. DXOMARK video lighting system

auto-white-balance on lighting transitions.

Test Environment

For still image quality measurement, DXOMARK uses the in-house designed Automatic Lighting System. This system has been designed to produce a very stable, precise and continuous lighting at several luminance values and color spectrum. However it does not allow to simulate fast lighting changes. For instance, the fluorescent tubes take several seconds to switch from 1000 lux to 10 lux. This is why we have designed a dedicated video light source, which is visible in Figure 1. This light source can produce from 1 lux to 2000 lux with color temperatures ranging from 2300 K to 2800 K. It also allows to perform up or down steps in luminance alone in less than 100 ms, and variation of both luminance and color temperature in ramps that converge in several seconds.

We perform the measurements on an X-rite ColorChecker Classic chart with a DXOMARK frame, reproduced in Figure 2. The metrics are computed in the CIELAB 1976 color space [6] with the reference white point in D65. For each patch of the chart, we compute the average L^*, Δa^*b^* and ΔH^*. We measure the target exposure as the L^* value on a 18 % gray patch. The white-balance and the hue error are respectively measured as the Δa^*b^* and ΔH^* values on the gray patches, and the color rendering as the Δa^*b^* value on all or tone-specific (blue, green, skin tone, ...) colored patches.

Luminance Steps

Luminance steps occur when the light changes abruptly in the scene, for example when filming a show in a dark room and the lights suddenly turns on. The device is presented with a very sudden change, and in auto-mode it has to respond to it. For each measured value (for instance L^*), we want to evaluate how fast, smoothly and accurately it converges. For that we define:

Convergence time The total time it takes to converge to a stable value.

Convergence error Convergence value error, compared to the value before the transition.

First convergence time The first time where the convergence value is reached.

Figure 2. X-rite ColorChecker Classic chart with DXOMARK frame

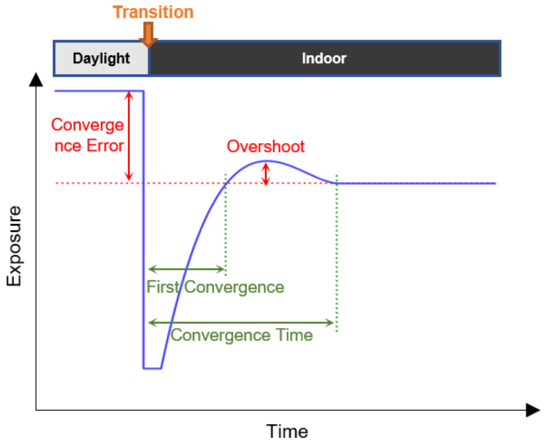


Figure 3. Luminance step convergence: terms definition

Oscillation time The time it takes for the value to converge to a stable value after reaching the value at convergence for the first time.

Overshoot The maximum overshoot during oscillation.

These concepts are illustrated in Figure 3.

We consider the convergence of a value v (for instance the L^*, ΔH^*, or Δa^*b^* value), in a video sequence that contains a transition. We first measure v on the ColorChecker chart patches on all the video frames as described above. Then, we detect the transition start time (with a threshold on the derivative of v) and the convergence time (with a threshold on the standard deviation and the slope of v over a moving window). We also compute the median value of v on a stable period before and after the transition.

Luminance and Color Temperature Ramps

In this section, we consider ramps where both the luminance and the color temperature vary over several seconds. These can occur when panning in a scene that contains several light sources, for example when walking in a room with both windows and artificial light sources. The difference with the steps described in the previous section is that the changes between consecutive frames are relatively small. Some devices react smoothly and adapt to the transition, others adapt with a lag (which creates abrupt changes or oscillations later on), or do not adapt at all. To estimate these

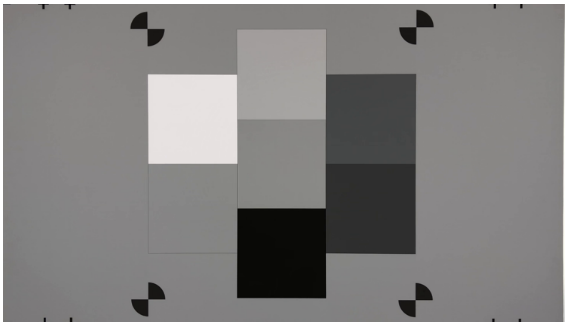


Figure 4. DXOMARK Visual Noise chart

parameters we define the following metrics:

Amplitude Amplitude of the variation during the transition.

Oscillations intensity Robust estimator of signal variation over time. A constant signal has an oscillations intensity of 0; that intensity increases with the amount of oscillations in the signal. The oscillations intensity is the difference of moving averages on a large and a small integration window

$$\frac{1}{T}\int_0^T |E_{T_S}[t,v]-E_{T_L}[t,v]|\,\mathrm{d}t, \tag{1}$$

where

$$E_T[t,v]=\frac{1}{T}\int_{t-T}^{t} v(\tau)\,\mathrm{d}\tau. \tag{2}$$

And v is the L^* or ΔH^* or Δa^*b^* value, T is the duration of the video, and T_S and T_L are the small and large integration windows durations respectively.

Temporal Visual Noise

Visual Noise for still images has been standardized in IEEE 1858 [3] and ISO 15739 [4]. Although these standards are supposed to provide good perceptual correlations, it has been argued by Wüller *et al.* [12], that while the general formulae are valid, there is still an effort in determining the parameters for the metrics. With an appropriate set of parameters, these metrics can be applied to video frames independently and the average value or the standard deviation can be assigned to the whole video sequence.

However, since photonic noise on the sensor varies with time, temporal noise adds up to the spatial noise in videos. This creates blinking that can be annoying and distracting to the viewer. To the best of our knowledge, regarding video visual noise, there is no proposed measurement protocol and no study on perceptual correlation. The goal of this section is to propose an adaptation of existing visual noise measurements to measure video temporal noise.

Test Environment

We use the DXOMARK Visual Noise chart, which can be seen in Figure 4. This chart is an ISO 14524:2009 compliant OECF test chart and it has large patches to be able to work with video resolutions that are usually smaller than photo resolutions. We consider constant lighting at different luminances from 1 lux to 1000 lux, and different color temperatures from 2300 K to 6500 K.

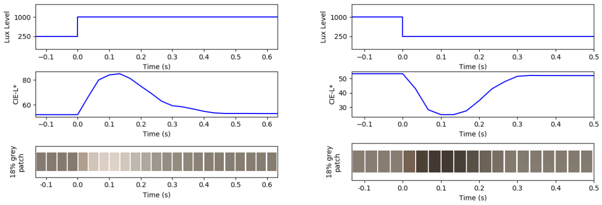


Figure 5. 18 % *gray patch on successive frames during an increasing transition from* 250 *lux to* 1000 *lux and a decreasing transition from* 1000 *lux to* 250 *lux, for the OnePlus 6T*

Temporal Visual Noise Metric

The temporal visual noise metric is computed in the CIELAB color space [6]. Exposure and white balance are compensated to isolate the temporal noise from other distinct effects (exposure drift, white-balance drift, *etc*). We measure temporal noise variances σ_{L^*}, σ_{a^*}, and σ_{b^*} as spatial averages (over the pixels of an entire patch) of the temporal variances (over all frames for a single pixel). For computational reasons, the variances are computed using Welford's online algorithm [7]. Since the noise depends on the luminance, these variances are computed on seven ROI with different reflectance. A normalization is required to compare different devices with different exposures: we interpolate on the measured CIE-L* values for all patches and compute the noise variances for $L^* = 50$, which is the average value of correctly exposed frames.

A recent perceptual study on still images [12] shows a good correlation of perceptual measurements to the square root of a weighted sum of the noise variances. Until a similar study is performed on videos, we set all weights to 1, and we define the Temporal Visual Noise (TVN) as:

$$\mathrm{TVN}=\sqrt{\sigma_{L^*}^2+\sigma_{a^*}^2+\sigma_{b^*}^2}. \tag{3}$$

Notice that this is the average euclidean distance of each pixel of a correctly exposed patch from its average.

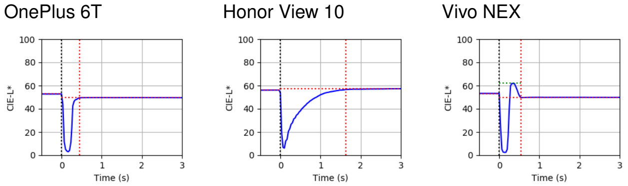
The perception of luminance and chromatic noise is different, users tend to be more sensitive to chromatic noise. Denoising algorithms for luminance and chroma noise are different as well. To evaluate how colored the noise is, we define the Temporal Noise Chromaticity (TNC) as:

$$\mathrm{TNC}=\frac{\sigma_{a^*}^2+\sigma_{b^*}^2}{\sigma_{L^*}^2+\sigma_{a^*}^2+\sigma_{b^*}^2}. \tag{4}$$

Results

Luminance Steps

Several devices were tested on comprehensive test scenario containing increasing and decreasing lighting transitions of different amplitudes (from 20 lux to 1000 lux) and in different lighting conditions (low-light, indoor, outdoor). For each transition, we measured the exposure convergence metrics defined in the previous section on a 18 % gray patch. Figure 5 shows the response of a device to increasing and decreasing transitions between 250 lux and 1000 lux. Figure 6 shows the exposure convergence for three devices: short convergence time, long convergence time, and overshoot. These are representative of the convergence shape and time of our data-set.



Convergence in 0.5 s
No Overshoot

Convergence in 1.6 s
No Overshoot

Convergence in 0.6 s
Overshoot 12 (L^*)

Figure 6. *Different devices reactions to a decreasing transition from* 1000 *lux to* 25 *lux*

	Luminance Transitions			
Transition size	**Small**		**Large**	
Transition direction	**Inc.**	**Dec.**	**Inc.**	**Dec.**
Apple iPhone 8 Plus	0.7 s	0.5 s	1.3 s	0.8 s
Apple iPhone XS Max	0.4 s	0.5 s	0.4 s	0.5 s
Google Pixel 3	0.6 s	0.7 s	1.2 s	0.9 s
Honor View 10	1.0 s	0.7 s	2.0 s	1.8 s
Huawei P20	0.5 s	0.6 s	1.0 s	1.6 s
Nokia N8	0.2 s	0.2 s	0.5 s	0.4 s
OnePlus 6T	0.5 s	0.4 s	0.5 s	0.4 s
Samsung Galaxy A8	0.2 s	0.2 s	1.9 s	0.8 s
Samsung Galaxy J5	0.2 s	0.2 s	1.7 s	0.9 s
Sams. Gal. S10 Plus	0.6 s	0.6 s	0.7 s	0.8 s
Sony Xperia 1	0.8 s	0.7 s	1.4 s	1.4 s
Sony Xperia L2	0.2 s	0.3 s	0.5 s	0.4 s
Vivo NEX	0.4 s	0.3 s	0.4 s	0.4 s
Xiaomi Mi 9	0.4 s	0.4 s	0.6 s	0.5 s

Figure 7. *Convergence time for different increasing and decreasing luminance transitions, both small (amplitude* < 200 *lux) and large (amplitude* > 600 *lux)*

Figure presents the convergence time for several devices for increasing and decreasing transitions both large (more than 600 lux difference) and small (less than 200 lux difference). These two categories cover the full range of tested transitions. Approximately half of the devices have convergence times under 0.6 s. For example, the Vivo NEX is around 0.4 s for all transitions. For other devices the performance changes significantly between large and small transitions. For instance, the convergence time for the Honor View 10 is twice as large for big transitions than for small. Generally speaking, convergence times are larger for increasing transitions than for decreasing ones. For some devices this difference is very noticeable, in particular for the Samsung Galaxy J5 and A8 on big transitions. We can see this difference in Figure 8.

Notice that convergence time is not enough to measure the performance of a device, specifically a device should converge to a suitable value of exposure. Figure 9 shows that the Samsung Galaxy A8 has a very fast convergence but it does not convergence to a suitable luminance value.

Luminance and Color Temperature Ramps

We have tested different devices ramps with intensity and color temperature variations. In this section, we focus on the

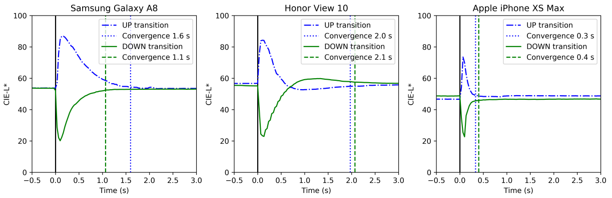


Figure 8. *CIE-L* value during luminance transitions for 3 devices: Comparison of Increasing (*250 *lux to* 1000 *lux) and Decreasing (*1000 *lux to* 250 *lux) transitions*

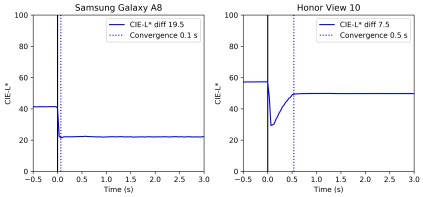


Figure 9. *CIE-L* value during a* 30 *lux to* 10 *lux luminance transition for 2 devices*

white-balance Δa^*b^* behavior on two transitions, which are representative of tested transitions:

Transition T1 15 s ramp from Daylight to Daylight + Tungsten,
Transition T2 15 s ramp from Tungsten to Horizon.

with

Daylight 1000 lux, 6500 K,
Tungsten 2000 lux, 2800 K,
Horizon 100 lux, 2300 K.

Figure 10 shows all different behaviors that have been observed on transitions T1 and T2. All devices have a bigger reaction to T2 than to T1. The OnePlus 6T and the Apple iPhone XS Max fully compensate the ramp on transition T1. On transition T2, the OnePlus 6T compensates most of the ramp though with more oscillations than on T1. On the other hand, the Apple iPhone XS Max does not compensate the ramp and has a large amplitude variation on transition T2. Large amplitude variations is seen on the Apple iPhone 8 Plus on both transitions. We can see that on both transitions the Sony Xperia 1 corrects the white balance with a lag, which creates abrupt variations. These can be perceived as oscillations.

These observations together with results of other tested devices can be seen in terms on amplitude variation and oscillations in Figure 11. Generally speaking devices have a bigger reaction to the transition T2, both regarding amplitude variation and oscillations.

Noise

Figure 12 shows the Temporal Noise Variance defined in Equation 3 and the Temporal Chromaticity for the Huawei P20 at different luminance values. We can see that the temporal noise increases when the luminance decreases, and that the noise chromaticity is significantly lower in bright light.

In Figure 13 the intensity of each pixel is proportional to the logarithm of the temporal noise variance of L^* on a 18% re-

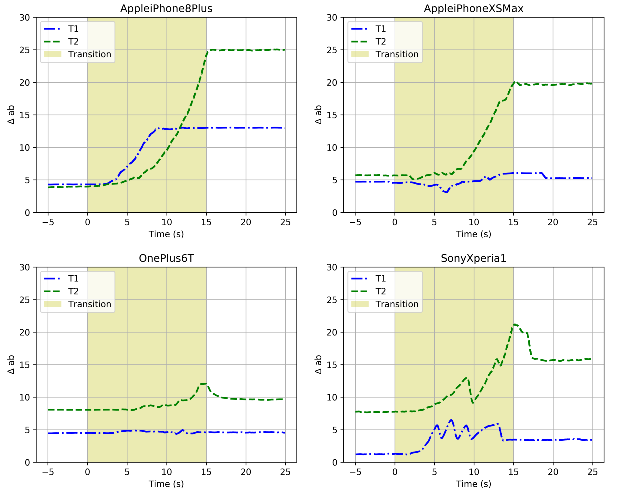


Figure 10. Δa^*b^* *value on ColorChecker gray patches during transitions T1 (blue curve) and T2 (green curve).*

	Amplitude		Oscillations	
	T1	T2	T1	T2
Apple iPhone 8Plus	8.8	21.1	0.6	0.8
Apple iPhone XS Max	3.1	15.1	0.8	1.2
Google Pixel 3	1.8	4.4	0.5	0.5
Honor View 10	6.9	14.0	0.5	1.2
Huawei P20	4.7	17.1	0.8	1.4
Nokia N8	3.5	13.0	0.8	1.6
OnePlus 6T	0.7	4.2	0.4	0.7
Samsung Galaxy A8	4.6	13.6	0.5	0.9
Samsung Galaxy J5	3.3	9.4	0.5	1.0
Galaxy S10 Plus	17.6	10.0	0.7	0.9
Sony Xperia 1	5.3	13.7	2.0	2.2
Sony Xperia L2	1.3	9.6	1.1	1.2
Vivo NEX	3.1	10.7	0.6	1.0
Xiaomi Mi 9	10.5	13.2	0.8	1.4

Figure 11. *White-balance* Δa^*b^* *amplitude variation and oscillation on transitions T1 and T2 for different devices.*

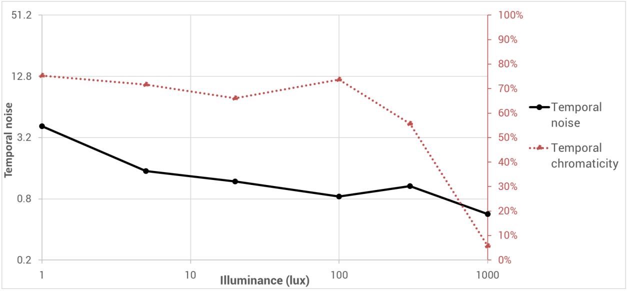


Figure 12. *Temporal noise per lighting condition for the Huawei P20*

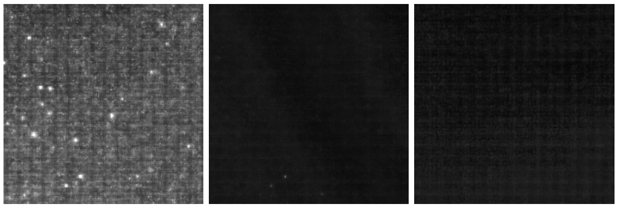


Figure 13. *Luminance noise for each pixel of a* $18\,\%$ *gray patch for Huawei P20 at* $1, 100,$ *and* $1000\,$*lux*

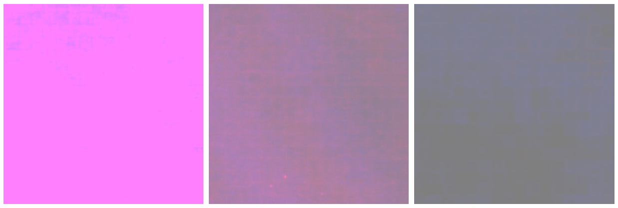


Figure 14. *Color noise for each pixel of a* $18\,\%$ *gray patch for Huawei P20 at* $1, 100,$ *and* $1000\,$*lux*

flectance patch at $1, 100,$ and $1000\,$lux. It illustrates the observation we made above that the temporal noise is higher in low-light. We can also observe temporal artifacts. First we can see there is less temporal noise on the grid where the deblocking filter [9] has been applied by the video decoder. We can also see that at $1\,$lux some pixels are very bright, which means they are very sensitive to light and vary a lot during time. In Figure 14, the color saturation of each pixel is proportional to the logarithm of the temporal variance of a^* and b^* for the same patch and the same lighting conditions as in Figure 13. We can see that the noise chromaticity increases when the luminance decreases.

The Temporal chromaticity in Figure 12 is the color noise ratio in the total amount of noise, and the temporal noise variance is the total noise on all channels. We can relate these to the images in Figure 13 and Figure 14 as follows. At $1000\,$lux, the image of the luminance noise image is dark and the image of the color noise is not saturated a lot, therefore both the temporal noise and the temporal chromaticity are low. At $100\,$lux, the luminance noise is almost the same but the image of chroma noise is more saturated than the ones at $1000\,$lux. Therefore the temporal chromaticity is larger than the one at $1000\,$lux. At $1\,$lux we can see that both the color noise and the luminance noise are larger than the ones in brighter light. This is why the temporal noise is larger at $1\,$lux, but the temporal chromaticity is the same at $1\,$lux and at $100\,$lux.

Conclusion

This work has led to the development of a comprehensive measurement environment with reproducible hardware, software, testing protocol and technical reports generation. It has been tested on many smartphones including high-end, mid-range and entry-level devices. It allows to better understand the limitations of the devices in terms of video quality. The technical reports are used for consulting, which allows camera manufacturers to improve video quality, and for publications on the DXOMARK website. They provide objective measurements that allow consumers and manufacturers to compare cameras.

Though the design of the metrics is not specific to smartphone video, it needs to be tested on other kinds of devices in future work, including DSLRs, automotive cameras, action cameras. On DSLRs, we expect good noise performance due to larger pixel pitches, but poor exposure and color behavior on light transitions, because DSLRs are not optimized for this purpose. For automotive cameras, performance on light transitions is crucial for the applications and it is important for manufacturers to be able to test it.

More effort is needed to correlate these measurements with perceptual analysis, despite the difficulty to perform perceptual analysis on video.

Disclaimer

This work was conducted solely by DXOMARK employees. Former employees listed as co-authors only contributed while working at DXOMARK/DxOLabs.

References

[1] D. Baxter, F. Cao, H. Eliasson, and J. Phillips. Development of the I3A CPIQ spatial metrics. In *Image Quality and System Performance IX*, volume 8293, page 829302. International Society for Optics and Photonics, 2012.

[2] Z. He, E. W. Jin, and Y. Ni. Development of a perceptually calibrated objective metric for exposure quality. *Electronic Imaging*, 2016(13):1–4, 2016.

[3] IEEE. IEEE 1858-2016: IEEE Standard for Camera Phone Image Quality. Standard, 2016.

[4] ISO. ISO 15739:2017: Photography – Electronic still-picture imaging – Noise measurements. Standard, 2017.

[5] E. W. Jin et al. Towards the development of the IEEE P1858 CPIQ standard – A validation study. *Electronic Imaging*, 2017(12):88–94, 2017.

[6] G. A. Klein and T. Meyrath. *Industrial color physics*, volume 154, chapter 3, pages 141–148. Springer, 2010.

[7] D. E. Knuth. *The Art of Computer Programming, volume 2: Seminumerical Algorithms*. Addison-Wesley, Boston, 3 edition, 1998. p.232.

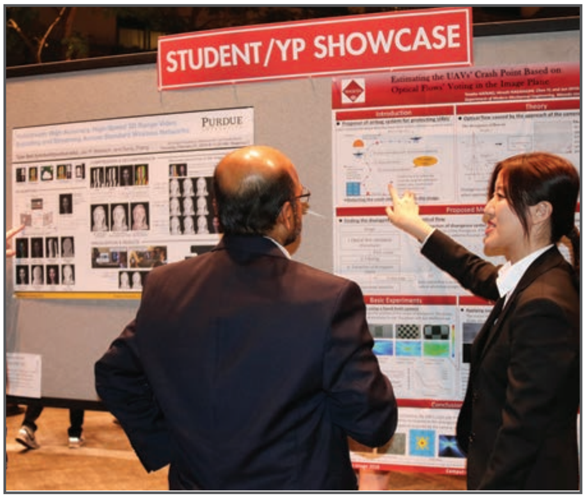
[8] S. Oh, C. Passmore, B. Gold, T. Skilling, S. Pieper, T. Kim, and M. Belska. A framework for auto-exposure subjective comparison. *Electronic Imaging*, 2017(12):202–208, 2017.

[9] I. Rec. H. 264: Advanced video coding for generic audiovisual services. *ITU-T Rec. H. 264-ISO/IEC 14496-10 AVC*, 2005.

[10] A. M. Rohaly et al. Video Quality Experts Group: Current results and future directions. In *Visual Communications and Image Processing 2000*, volume 4067, pages 742–753. International Society for Optics and Photonics, 2000.

[11] M. Shahid, A. Rossholm, B. Lövström, and H.-J. Zepernick. No-reference image and video quality assessment: a classification and review of recent approaches. *EURASIP Journal on image and Video Processing*, 2014(1):40, 2014.

[12] D. Wueller, A. Matsui, and N. Katoh. Visual noise revision for ISO 15739. *Electronic Imaging*, 2019(10):315–1, 2019.


STUDENT/YP SHOWCASE
PURDUE

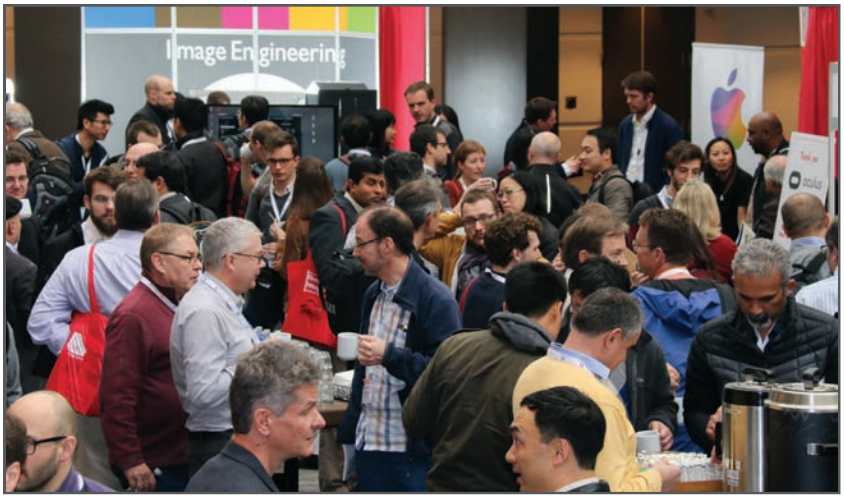


Image Engineering


AR is a fundamental and transformative interface and platform.
20+ years from my initial working demo to today's AR displays
Optimistic that it won't take another 20 to reach ubiquitous consumer AR!
My first working AR prototype
Today's platforms
Ubiquitous consumer AR