VCX Version 2020 - Further development of a transparent and objective evaluation scheme for mobile phone cameras

Uwe Artmann, Image Engineering GmbH & Co KG; Kerpen, Germany

Abstract

VCX or Valued Camera eXperience is a nonprofit organization dedicated to the objective and transparent evaluation of mobile phone cameras. The members continuously work on the development of a test scheme that can provide an objective score for the camera performance. Every device is tested for a variety of image quality factors while these typically based on existing standards. Tests include texture loss, resolution, low light performance, shooting time lag, image stabilisation performance and more, all for a variety of different capture conditions. This paper presents that latest development with the newly released version 2020 and the process behind it.

Introduction

VCX-Forum e.V. is the non-profit organisation that develops the VCX test protocol. The aim is to create a transparent and objective way to describe the customer experience with cameras in mobile devices. It is formed by a large group of mobile phone manufacturers, chipset and module manufacturers, test labs, and mobile phone carrier companies. In contrast to other commercial services, the published score[1] is created by independent labs with a fixed test plan which was developed in a joined effort of all members. VCX functions as a standardisation group and test organisation at the same time.

Version 2020

The test procedure is now defined in its Version 2020. The previous version is already described in publicly available documents [2], so this document will mainly focus on the differences and improvements. The whole procedure is described in greater extend compared to this publication in a public available white paper [3].

Image Quality

A key component of the VCX test procedure is the usage of multi-purpose test targets. So these are test targets that contain a large variety of test patterns for different aspects of image quality. These charts allow for the measurement of many different key performance indicator (KPI) from a single image. Beside that this is time effective, we can make sure that all KPIs are measured under the exact same condition. In version 2020, VCX is now using a test target (see Figure 1) that is described in the low light performance standard ISO19093[4]. This target features structures that allow for the measurement of many different KPIs.

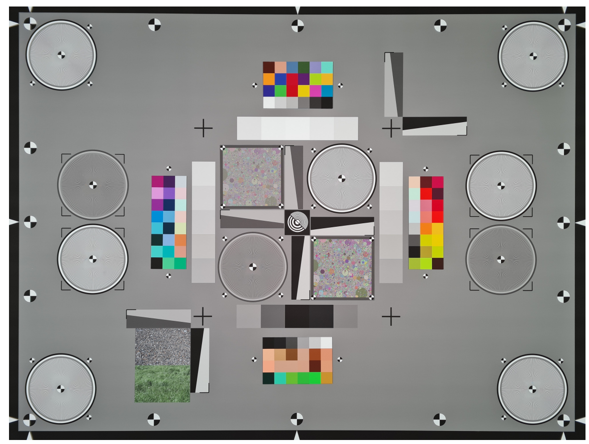

Figure 1. *The multipurpose chart as described in ISO19093*

Visual Noise Based on the gray patches, noise is measured as Visual Noise [5] for three different viewing conditions.

Color Reproduction Based on the color patches (individually measured) the color reproduction is measured and expressed in ΔE. The measurement is performed for different subsets of colours, so that e.g. the color reproduction of skin tones can be evaluated and assessed separately.

s-SFR Based on the harmonic Siemens stars, the s-SFR [6] is measured. This method has shown to be less influenced by image enhancement algorithms like sharpening [9] and be able to be useful to evaluate limiting resolution in fully processed images.

e-SFR Based on slanted edges with two different edge modulation, the e-SFR [6] is measured. This method is mainly utilised to describe the sharpening applied to the image.

Texture loss Based on two different dead leaves pattern, the texture loss is measured[7].

Shading The gray background is used to evaluate intensity and color shading.

TV-Distortion Marker in the image allow for a measurement of the TV-Distortion.

For each test, four images are captured, all are analysed and the image with the highest score is used for the report. All these KPI are measured under different light and capture conditions.

Within the VCX Version 2020 test procedure, three main light conditions are defined, see Figure 2 for details on their spectrum.

Bright This is defined with an intensity of 2000lux and a spectral distribution matching D55.

Medium This is defined with an intensity of 250lux and a spectral distribution matching a neutral white LED.

Low This is defined with an intensity of 10lux and a spectral distribution matching a warm white LED.

In comparison with the previous version, the intensity has been increased for bright to reach "base ISO" for all devices under test. The low light condition is now significantly darker, as it changed from 63lux to 10lux. To further evaluate the performance in even darker conditions, a separate "extended low light" test has been introduced which measured down to 1 lux.

In the previous version the light spectrum was constant for all intensities. In Version 2020, the spectral distribution changes with the intensities to reflect more likely scenarios. It is done by intention to mix white light with (D55) and without (neutral LED) IR content, as this is a potential difficulty for devices under test and their color shading correction.

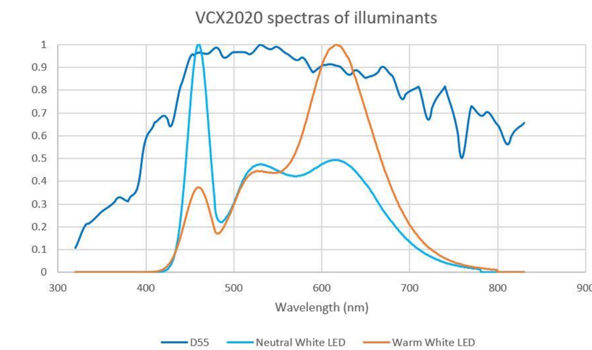

Figure 2. *The relative spectral distribution of the used spectra.*

Image Quality measurements are performed for different measurement conditions. For all tests the device under test is set to its default mode and the default, pre-installed camera app is used.

Main The main camera of the device is used to capture images. The camera is set to default zoom, typically shown as "1x" in the user interface.

Zoom The performance of any kind of zoom (optical or digital) is measured at 4x zoom. In contrast to the previous version, zoom is now measured for all light conditions, which reveals significant differences in the image quality in low light as multi-module devices will switch back to digital zoom using a more sensitive module rather than using the less sensitive module with a longer focal length.

Video The video image quality is measure by extracting frames from a video sequence captured under the three different light conditions. The frames are extracted from a sequence that is at least 10s long and shows the test target.

Selfie The selfie camera is measured in the same way as the main camera, with the difference that a smaller version of the test target is used. This shall make sure that devices with fixed focus lenses in the selfie camera are capable to have the chart in focus. (see Fig. 3)

For the main camera, an additional measurement using a high contrast test target has been introduced (see Figure 4). This target is also back-illuminated with the specified lighting conditions.

In the previous test version, the dynamic range was extracted from the multi-purpose test target and it was assumed that the dynamic range of the target would be the dynamic range we measure

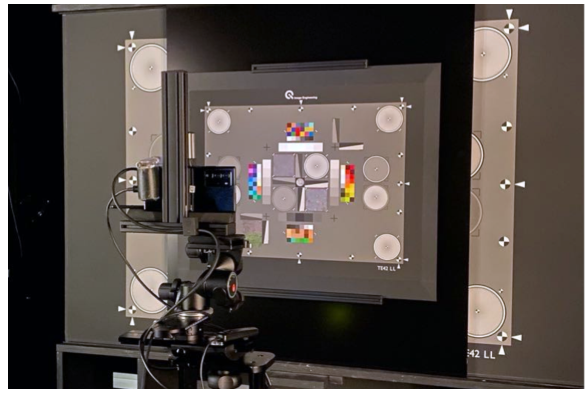

Figure 3. *Setup for evaluation of selfie camera.*

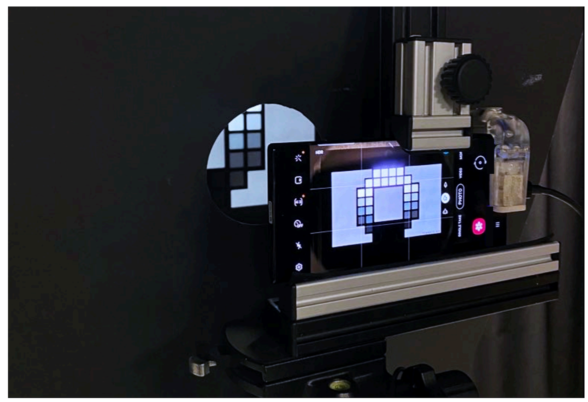

Figure 4. *Setup for additional measurement of the main camera with a high dynamic range test target.*

if a device has a better dynamic range than that. With the modern mobile phone and their possibilities in high dynamic range imaging and local tone mapping applied, this is no longer true.

Motion Test

Optical image stabilisation became a common feature of most mobile phone cameras. The test procedure has been updated based on new research [8] regarding the applied motion and the used test target has been updated. So now the same target and light conditions are used and the device is shaken with a more natural artificial handshake (see Fig. 5).

Timing

The previous version featured timing measurements that turned out to have a significant "human factor" in the measurement. So the alignment between different labs and also different team members took some time and was something that needed constant tracking. With the new version, the procedure has been updated to be fully automated. The biggest challenge is to create a controlled environment for timing measurement as mobile phones tend to have a continuous focus (no time to change the scene manually) and a shutter priority (devices can be very fast but did not focus).

In the new procedure, a near target and a far target are de-

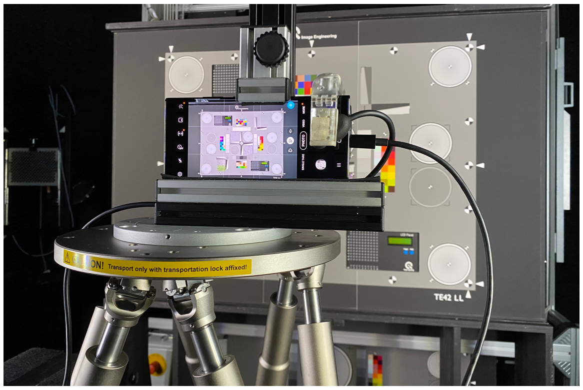

Figure 5. *Setup for evaluation image stabilisation.*

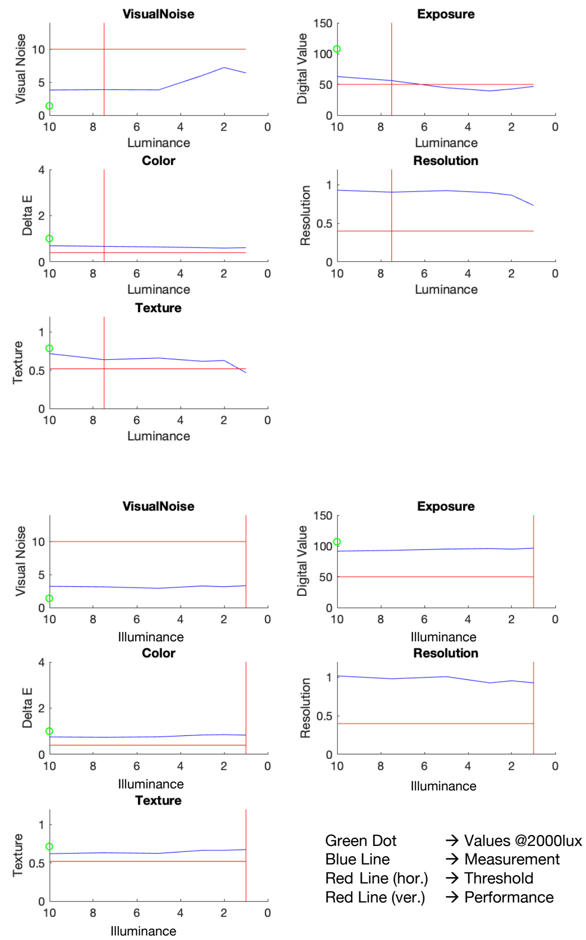

Figure 6. *Sample of extended low light performance measurement of two devices. The upper shows a low light performance of 7.5lux (as exposure falls below threshold at 5lux), the lower one of 1lux.*

signed with the possibility to quickly remove the near target out of the field of view of the device under test and trigger a touch on the release button with an automated and synchronised solution. The captured image shows an LED based timing device and the multi purpose chart, so from that image we can evaluate the time it was captured and the resolution for a focus check.

The new procedure will allow for an evaluation of negative shooting time lag, so the case when the captured image was exposed before the release button was captured.

Score Calculation

The score calculation process has been updated to reflect the new test items and also reflect experience from the previous version. The process to generate the total score for a device remains the same: Every metric that was decided to contribute to the final score is converted into a score range from -1 to 1. In this conversion it is described which result is considered as an excellent result (score = 1), as a poor result (0) and which result has such impact on the overall performance, that it gets a negative score to compensate for a possible win in other metrics (score = -1). In the previous version, a certain type of function for this conversion was defined, this has been removed by defining a look-up-table which allows for interpolation. Depending on the metric we can have a simple linear conversion between metric and score. In other cases, we can have a complex-shaped conversion allowing to define ”sweet spots” or logarithmic relationships.

The ”per metric” score is then multiplied by a weight which reflects the importance of this metric within one group. A group can be for example texture loss, where multiple metrics contribute. The group will then also get a weight that reflect the importance of this group for the total image quality in the given condition.

The total score and all sub-score are summarised with a certain weight as shown in Figure 8.

Spotlights

The spotlight shall highlight some important findings in the development process of Version 2020.

Low Light Performance

A separate extended low light performance measurement has been established with Version 2020. Beside the image quality measurement at 10lux, a measurement will provide the illumination that leads to acceptable image quality. This procedure is close to the ISO19093 [4] standard procedure by capturing images under different illumination levels and then checking at which light level one of multiple KPIs is below a certain threshold. In the VCX procedure illumination steps of 10lux, 7.5lux 5 lux, 3lux, 2 lux, and 1lux are used and at each step a metric for exposure, visual noise, resolution, texture loss, and color reproduction is checked. The thresholds for these KPIs have been defined based on a psychophysical study among the members based on artificially treated RAW image and a software-based image pipeline.

A sample result is shown in Figure 6 where one device shows a low light performance of 7.5lux and the other one did not hit any threshold for the given illuminances. Our finding was that many devices could handle a low light situation at 1 lux quite well. We decided not to measure below 1 lux because the light intensity measurement in that range becomes very critical and a lab to lab comparability will be a huge challenge. Another surprise was, that the acceptable noise level for many observers is very high, so a Visual Noie value of 10 was considered acceptable, which indicates clearly visible noise.

Subjective Studies

For the development work, the VCX workgroup performed multiple studies. Images were shared and participants were asked to rate these images. Figure 7 shows the outcome for one dataset. Something that was not expected was the huge variation between the observers. As the constraints were very low (just guideline on viewing conditions but not controlled), a variation was expected, but to that extend. The data needed to be double checked as well, as we found that we had some outliers in the data which we tracked back to different rating systems the observers use as default in their companies so that they used the one they were more familiar with even though it had a different meaning int he VCX study.

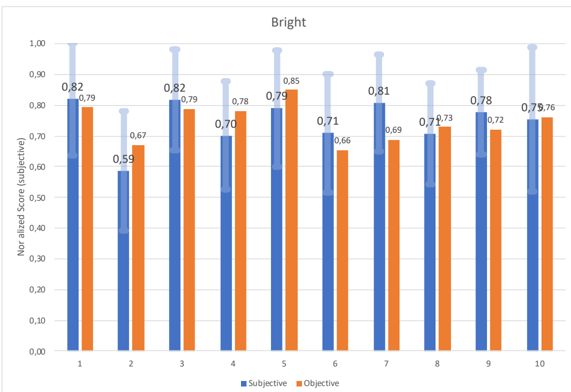

Figure 7. *Comparison of normalised score from a study with objective score (not final score). Note the high standard deviation on the subjective data, shown as error bars.*

Author Biography

Uwe Artmann studied Photo Technology at the University of Applied Sciences in Cologne following an apprenticeship as a photographer, and finished with the German 'Diploma Engineer'. He is now CTO at Image Engineering, an independent test lab for imaging devices and manufacturer of all kinds of test equipment for these devices. His special interest is the influence of noise reduction on image quality and MTF measurement in general. He is also the head of the standards department within VCX-Forum e.V.

References

[1] The VCX website, `https://vcx-forum.org`

[2] Wueller, Rao, Reif, Kramer, Knauf, "VCX: An industry initiative to create an objective camera module evaluation for mobile devices", Electronic Imaging, Photography, Mobile, and Immersive Imaging 2018, pp. 172-1-172-5(5)

[3] Whitepaper VCX - Version 2020, `https://vcx-forum.org/standard/white-paper`

[4] International Organization of Standardization, "ISO19093:2018 Photography Digital cameras Measuring low-light performance"

[5] International Organization of Standardization, "ISO15739:2013 Photography - Electronic still picture imaging - Noise measurements"

[6] International Organization of Standardization, "ISO12233:2017 Photography - Electronic still picture imaging - Resolution and spatial frequency responses"

[7] International Organization of Standardization, "ISO/TS 19567-2:2019 Photography Digital cameras Part 2: Texture analysis using stochastic pattern"

[8] Bucher et. al., "Issues reproducing handshake on mobile phone cameras", https://doi.org/10.2352/ISSN.2470-1173.2019.4.PMII-586

[9] Artmann, "Image quality assessment using the dead leaves target",Proceedings Volume 9404, Digital Photography XI; 94040J (2015)

[10] Artmann, Wueller,"Improving texture loss measurement: Spatial frequency response based on a colored target", January 2012, Proceedings of SPIE - The International Society for Optical Engineering 8293:4-DOI: 10.1117/12.907303

Annex

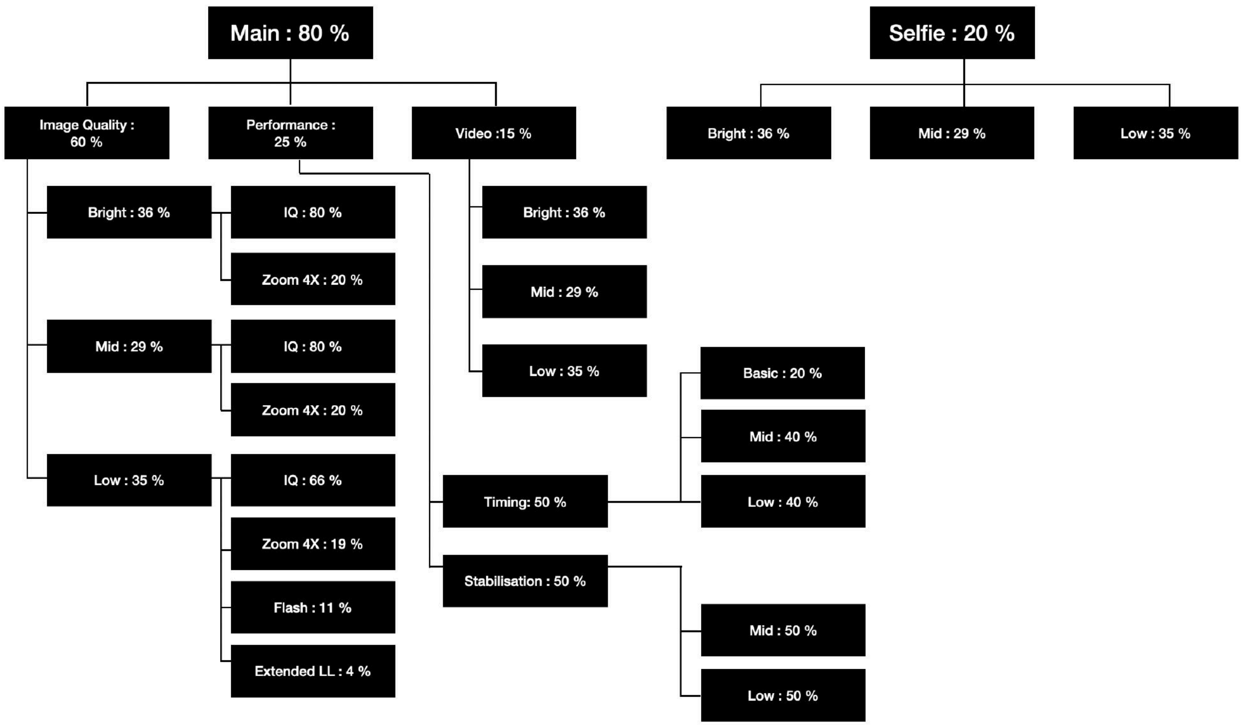

Figure 8. *The weights used to calculate the total score and subscores.*