https://doi.org/10.2352/EI.2023.35.8.IQSP-317

VCX Version 2023 The latest transparent and objective mobile phone test scheme

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

Abstract

VCX or Valued Camera eXperience is a nonprofit organisation dedicated to the objective and transparent evaluation of imaging devices like mobile phone cameras and webcams. 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. This paper presents that latest development with the newly released version 2023 and the process behind it. New metric included are extended tests on video dynamics, AE and AWB, dedicated tests on ultra wide modules and adjustments to the metric system based on a large scale subjective study.

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.

In this paper we describe the progress to develop the next mayor release (Version 2023), which will modify some parts of the current version and will add some important new categories and metrics.

A complete new test protocol for cameras used in context of web meetings (webcams) is presented in a separate paper by Orchard et. al.[3]

Version 2020

The current test protocol is the Version 2020. The whole procedure is described in greater extend compared to this publication in a public available white paper [2]. Previous paper describe details of the development. [4] [5]

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 using a test target (see Figure 1) that is described in the low light performance standard ISO19093[6]. This target features structures that allow for the measurement of many different KPIs.

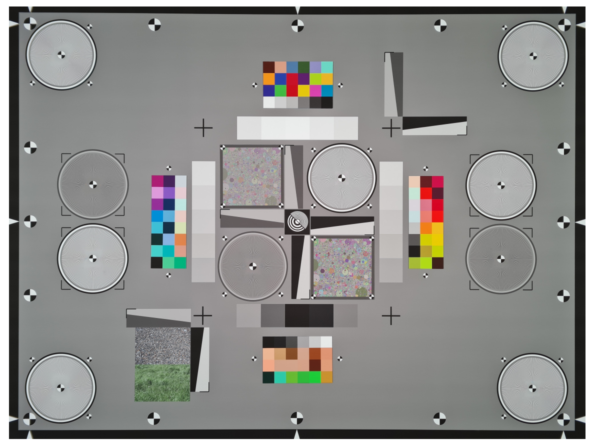

Figure 1. *The multipurpose chart as described in ISO19093*

Visual Noise Based on the gray patches, noise is measured as Visual Noise [7] 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 [8] is measured. This method has shown to be less influenced by image enhancement algorithms like sharpening [11] 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 [8] 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[9].

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.

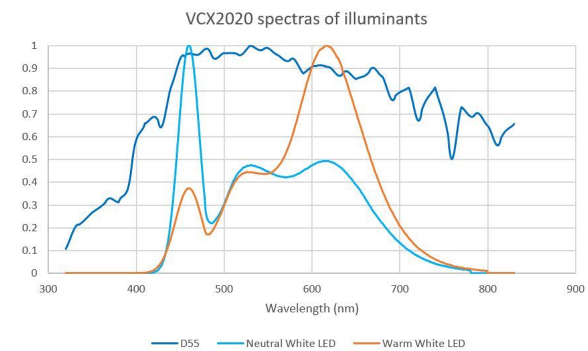

Figure 2. The relative spectral distribution of the used spectra.

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.

Motion Test

Motion test addresses the performance of optical image stabilisation system to compensate handshake. The same target and light conditions are used and the device is shaken with a natural artificial handshake[10] (see Fig. 5).

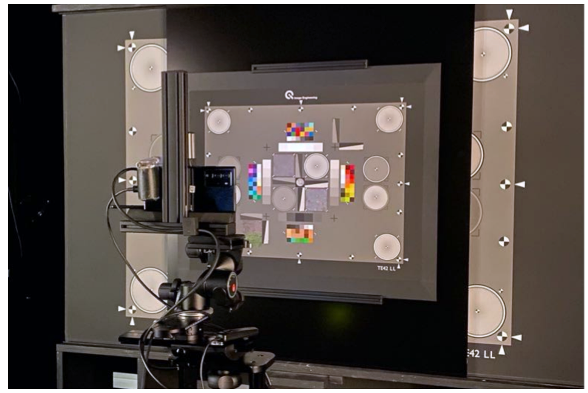

Figure 3. Setup for evaluation of selfie camera.

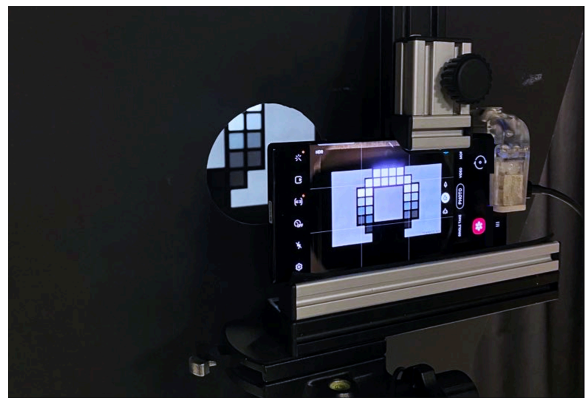

Figure 4. Setup for additional measurement of the main camera with a high dynamic range test target.

Timing

A near target and a far target are used 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 procedure allows 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

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). The metric to score calculation is based on 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.

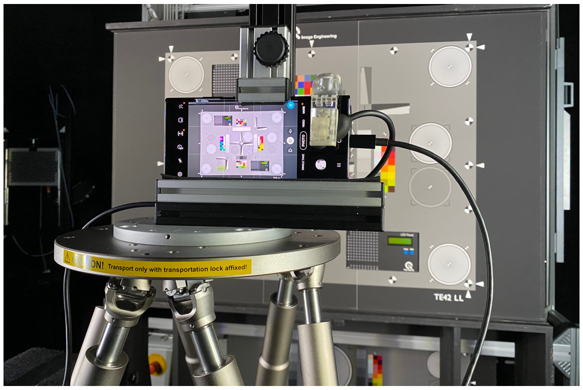

Figure 5. *Setup for evaluation image stabilisation.*

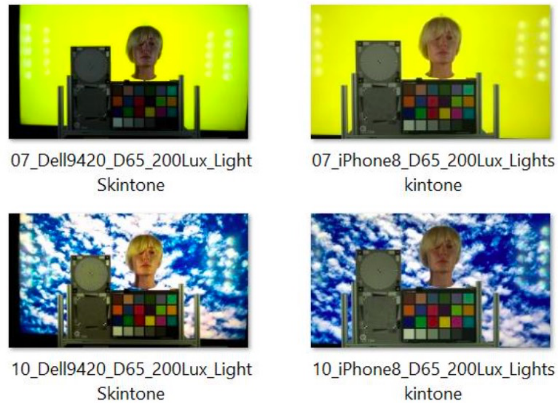

Figure 6. *Example of different backgrounds to create a challenging scene for AE and AWB.*

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 (Figure 11).

The total score and all sub-score are summarised with a certain weight as shown in Figure 12.

Version 2023

Version 2023 is the next mayor release. In the past years of using Version 2020, no mayor issue has been detected beside the wish of the group to extend the test protocol. So the test protocol of Version 2023 will remain mostly the same, while it will be extended by additional measurements and the metric to score algorithms including its weights is reviewed.

Video Performance

While Version 2020 covers the video image quality, Version 2023 covers also dynamic video parameters like and AE/AWB convergence. The protocol will be synchronised withe the WebCam protocol[3]. A scene is illuminated with different illuminants (intensity and spectrum) and the time is measured the device needs to adjust the auto exposure and the automatic while balance to the new scene. The main focus is to check how long it take to converge to a stable setting, not so much if the new setting is correct for the scene.

Challenging AE und AWB

Also from the WebCam protocol a scenario will be used to have a more challenging scene for auto exposure (AE) and auto white balance (AWB). This scene is created by using a dynamic background to a mannequin head. The foreground is illuminated with different intensity and spectra, while the background shows different images on a calibrated high quality display. The display shows bright background images to create high dynamic scenes with difficult exposure setting and scenes with dominant colours, which can influence poor AWB algorithms (see Figure 7).

Extended Zoom

The Version 2020 protocol tests all devices with a zoom factor of 4, regardless if the device features an optical zoom, additional modules or relies entirely on digital zoom. This procedure is extended to 10x zoom to reflect the existence of devices featuring optical zoom or other techniques supposed to create decent images at high zoom ratios. As high zoom ratios like 10x have the disadvantage that handshake is very critical, therefore the motion test is also extended to larger zoom ratios. These tests are only applied if the UI of the device allows to set high zoom ratios of 10x. It is decided not to penalise devices for not providing such high zoom ratio as it is currentely not a ”must have” feature for mobile phones.

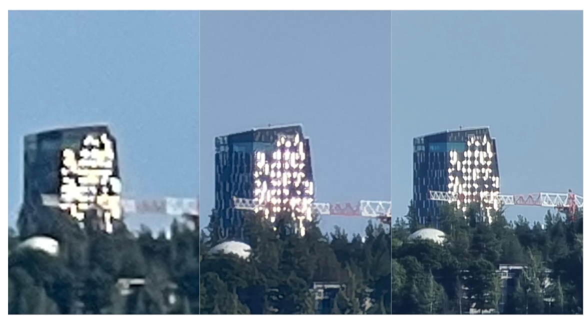

Figure 7. *Example of different devices capturing the same scene with 10x zoom*

Virtual Boukeh

The limitations in physical size of mobile phone cameras leads to a large depth of field (DOF) for these type of devices, while system cameras with e.g. full frame sensor can have a shallow DOF and make use of this a creative element in photography, particularly portraits. To compensate for this optical limitation, many mobile phone feature a technique that allows to apply virtual boukeh, so to virtually blur the background to give the impression of much larger cameras with shallow DOF. These algorithm have to overcome some challenges depending on the foreground object. Small features belonging to the foreground (hair, glasses, etc.) might be blurred while enclosed background (gap between headphones, area between fingers) might not be blurred. These artefacts can render the algorithms useless to the user, therefore the performance shall be tested objectively.

The process uses a mask with multiple holes (rectangles, circles, triangle, polygons) in front of a feature rich background (see Figure 8). The test is preformed with maximum blur effect (in some UI defined by large aperture and small f#) and minimum or no blur effect (small aperture, large f#).

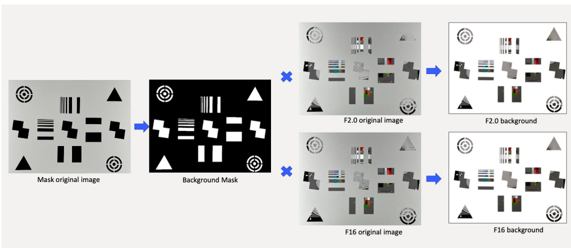

Figure 8. *Different masks used to evaluate performance of virtual boukeh*

Wide Module

While the new Version covers extended zoom, it also covers the test of wide modules. So these tests apply to the user setting that is typically applied by a zoom factor of smaller than one, mainly realised by an additional camera module with wide field of view.

While many metrics can be obtained int he normal process of capturing a test target with this setting, the problem remains for the measurement of spatial frequency response in the very corner of an image. Most devices showed issues on areas larger than 80% field and the loss of details is not so much purely optical, but a combination of optical design, distortion correction and other field depending image enhancements.

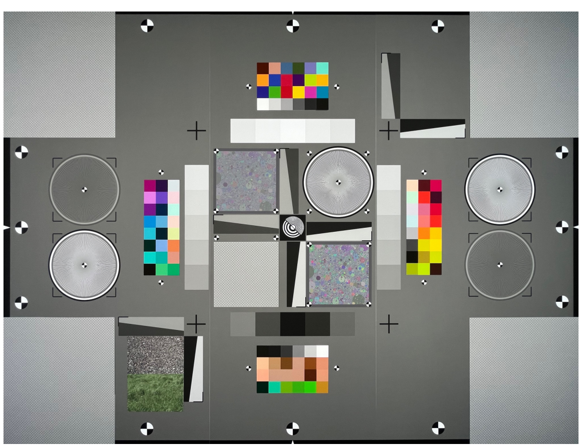

Figure 9. *Target used for wide module evaluation.*

Tests showed, that Siemens stars can not cover the very corner of the image and slanted edges do not reflect the texture loss like effects of image enhancement in the corner. So a new method is developed that allows for measurement in the corner. The target is shown in Figure 9. The main components remain the same compared to the standard target (see Fig.1), but has replaces the low contrast Siemens star close to the center and the corner region with a measurement pattern. This pattern consists of only a few discrete spatial frequencies in two directions. By detecting these frequencies in the power spectrum of the image (Fig.10) we can obtain a spatial frequency response.

As this measurement is influenced by image noise, we use the center patch for reference, so the spatial frequency response in the corner relative to the center is used as a metric.

Metric to Score

It is obvious, that the new metric need to be included into the score system. So for each metric, the look-up-table for the metric to score conversion needs to be defined and then we will have new groups, so the weight of each group needs to be re-adjusted. To put this into an objective process, this is done by a dedicated group of experts that also conducted a large scale user study.

For this study, 1500 users have been interviewed in an online survey. They provided data via a questionnaire and also had to review images captured with ten different devices from different scenes. Each image was judged on a scale of 1 (bad) to 7 (best) plus an additional question of what was the main driver for the decision. We can see that all scenes have a similar mean value and a good distribution, so the can see that the scenes do not have an impact on the score of the devices.

The collected data is very complex and shows multiple dependencies. For now, it seems that the high importance of spatial frequency response and texture loss in the Version 2020 weight system is backed by this study. The questionnaire focused on user expectations and typical device usage, that will also drive the score system.

Conclusion

The Version 2023 will extend the Version 2020 and reflects the work of a large, multi-company, international work group. This paper reflects the current status of discussion, a final release is expected for Summer 2023 and will be released via the VCX-Forum e.V. website.

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

[3] Orchard, Artmann "VCX-WebCam 2023: A Transparent and objective test scheme for webcams", in Proc. IS&T Intl. Symp. on Electronic Imaging: Image Quality and System Performance XIX [TBD], 2023

[4] 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)

[5] Uwe Artmann, "VCX Version 2020 - Further development of a transparent and objective evaluation scheme for mobile phone cameras" in Proc. IS&T Intl. Symp. on Electronic Imaging: Image Quality and System Performance XVIII, 2021, pp 204-1 - 204-6, https://doi.org/10.2352/ISSN.2470-1173.2021.9.IQSP-204

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

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

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

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

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

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

[12] 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

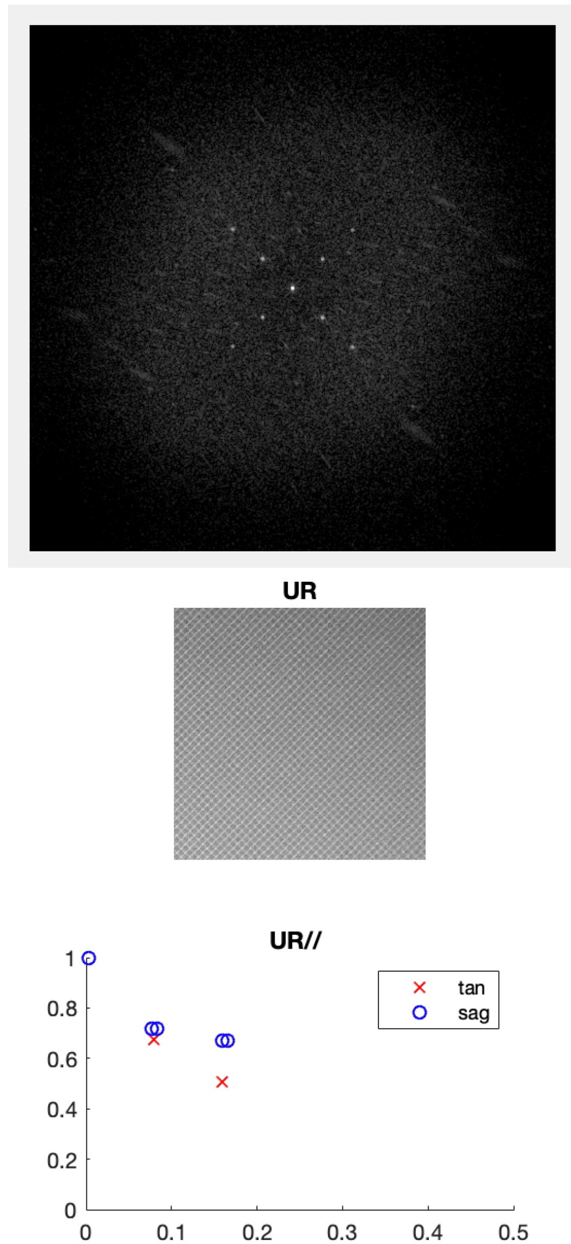

Figure 10. *Top: Discrete spatial frequencies in the power spectrum of the image, showing the pattern. Down: Sample SFR obtained for one ROI (UR = Upper Right)*

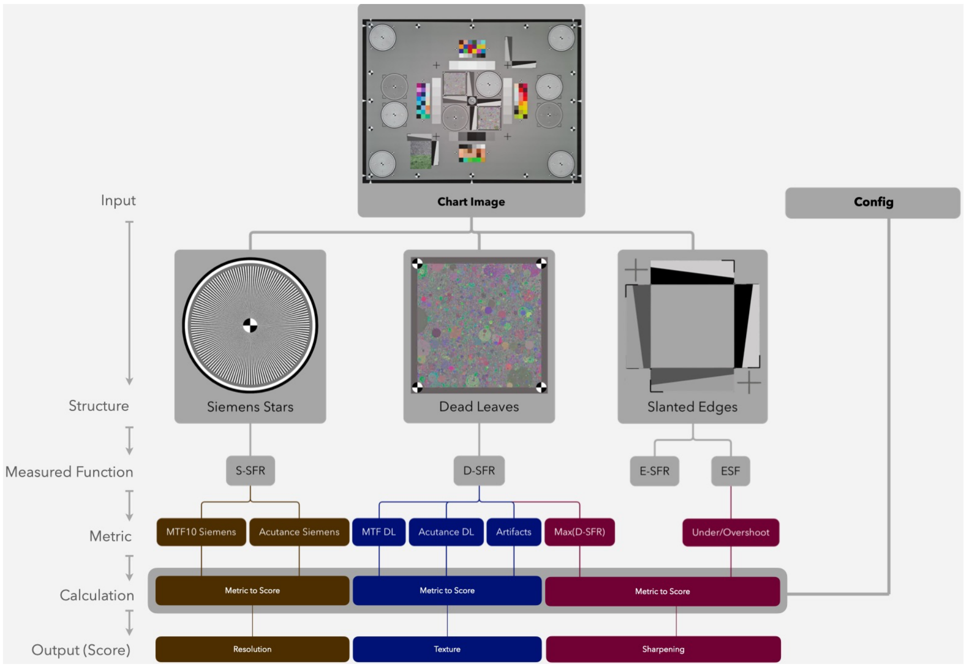

Figure 11. *Different metrics form a group, these contribute to the total score depending on the weight per metric and per group. All measurements are performed per light condition.*

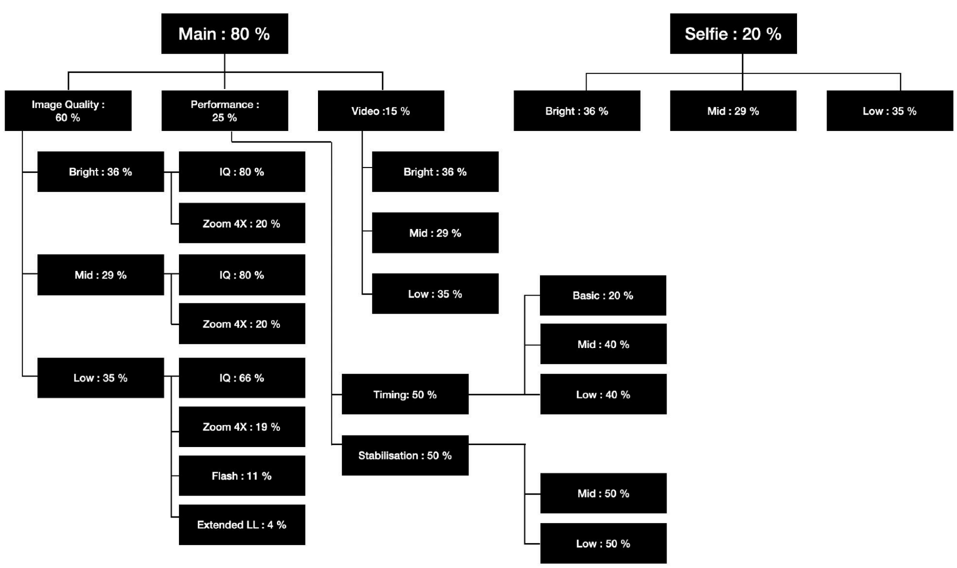

Figure 12. *The weights per group and category for version 2020*