https://doi.org/10.2352/ISSN.2470-1173.2017.15.DPMI-252

Cell Phone Camera "Rankings"!

Dietmar Wueller; Image Engineering; Frechen, Germany

Abstract

Right now there are at least three publicly known ranking systems for cell phones (CPIQ [IEEE P1858, in preparation [1]], DxOmark [2], VCX [3]) that try to tell us which camera phone provides the best image quality. Now that IEEE is about to publish the P1858 standard with currently only 6 Image quality parameters the question arises how many parameters are needed to characterize a camera in a current cell phone and how important is each factor for the perceived quality.

For testing the importance of a factor the IEEE cellphone image quality group (CPIQ) has created psychophysical studies for all 6 image quality factors that are described in the first version of IEEE P1858. That way a connection between the physical measurement of the image quality aspect and the perceived quality can be made.

How to determine overall image quality

There are two key aspects for creating a ranking system that are described in the Handbook Image Quality by Brain W. Keelan [4]. Number one is:
"The performance of an image quality system is quantified by the full distribution of image quality it produces under representative manufacturing and customer usage conditions."

The second one is:
"To model the image quality distribution resulting from a photographic system, one identifies a minimal set of independent factors called primitives that, if known for a particular image, are sufficient to predict its quality through a series of deterministic relationships."

Image Capture Conditions

Looking at the first key aspect the manufacturing conditions in our case are the image capture conditions of images. To get the full picture we have to look into typical capture conditions. It is important under which lighting conditions the images are captured because the light level as well as the spectral distribution has a significant impact on the image quality the camera produces.

On the other hand it is also important to know what is captured (landscapes, people, documents…) because that defines the focus distance, the scale of the object. It also triggers the auto focus position and the image processing in general. Moving objects may also produce challenges for the camera.

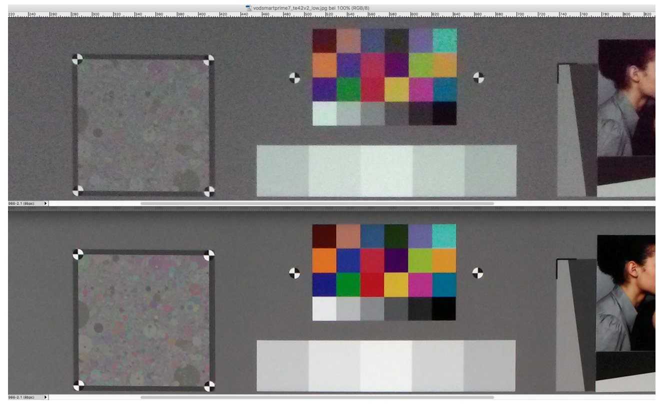


Figure 1: The upper image under low light (64 lx) is noisier, shows less detail and has a stronger a color shading than the lower image that was captured under bright light (1000 lx).

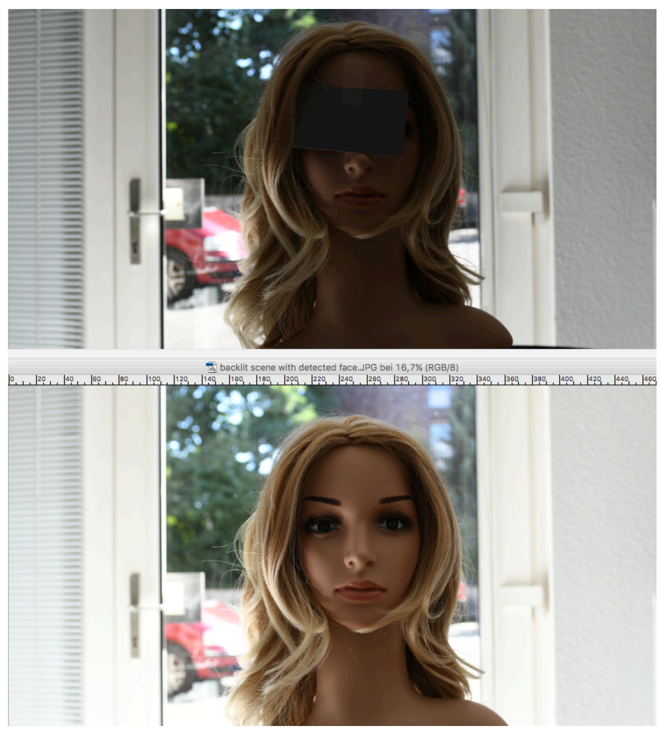


Figure 2: The impact of face detection on the exposure and image processing in a camera. In the upper image of this backlit scene the face is covered and in the lower image the face was detected.

Under bright light conditions most of the current phones deliver images that users do not complain about. The separator in image quality of cell phone cameras often times is the performance under low light conditions. Some of these images are captured with and some without flash.

There has been a study on the capture and capture conditions presented at EI 2008 [5] that analyzed millions on images for their capture conditions and 10,000 images for their content.

Table 1: ISO speed distribution extracted from statistical image analysis [5]

How many images were shot at which ISO level?	Images	Percent
Total number of images with ISO tag	2.386.274	100
50	269.753	11,3
64	152.198	6,4
70	11.217	0,5
80	252.732	10,6
100	827.059	34,7
125	69.983	2,9
128	12.953	0,5
140	51.193	2,1
141	9.815	0,4
160	95.797	4,0
200	225.671	9,5
250	25.332	1,1
320	42.836	1,8
400	152.043	6,4
>400	41.012	1,7
others	146.680	6,1

Table 2: scene content analysis extracted from statistical analysis [5]

Images used to categorize	10.000	Percent
Group portraits	3.628	36,3
Children	1.688	16,9
Single Portraits	1.510	15,1
Landscape	553	5,5
Architecture	541	5,4
Urban areas	340	3,4
Animals	328	3,3
Plants	309	3,1
Sports	186	1,9
Indoor	151	1,5
Food	81	0,8
Night images	28	0,3
Others thereof: 156 signs, 142 boats, 107 cars	657	6,6
Overexposed (visual impression)	552	5,5
Underexposed (visual impression)	98	1,0

Taking these numbers and additional indicators from that study on fired flash and exposure level analysis together with estimated changes of using cameras because of a cell phone being always with people (e.g. taking food images in restaurants…) an approximate estimation on the illumination conditions together with flash usage can be made.

Approximately 40% of the images are captured under bright daylight conditions, app. 30% are captured under lighting conditions of below 70 lux with artificial warm white illumination (tungsten fluorescent and increasingly LEDs) in case the flash is not used and app. 30% are distributed over the light levels in between.

The statistical analysis also shows that app. 70% of the images contain people.

Viewing conditions

The only source we have on how cell phone camera users view their images is a customer survey of Vodafone. Based on that most users view their images on the cell phone but at the same time a lot of users zoom into their images on the phone to view and show details in the captured images. This is especially because the phones usually provide a wide angle lens and the important part only covers a small portion of the image. Some users zoom into their images prior to the image capture (if the phone allows) some zoom or crop images after capture. There are no reliable surveys (according to the authors knowledge) on how often this occurs and how it is done. But the fact that it is done should also be reflected in a scoring system.
This means that a good scoring system should know the users behavior and expectations to weight the different shooting conditions accordingly. How far the scene specific image processing requires different types of tests depends on the individual camera but should be kept in mind as well.

For CPIQ the scoring was not final at the time this paper was written.

The VCX score has the following weighting for the illumination conditions:

Hi_Lux	34%
Mid_Lux	23%
Lo_Lux	17%
Flash	11%
Zoom	14%

Flash and zoom have been separated because several image quality aspects (like e.g. color shading, noise, texture etc.) may change if the camera switches on the flash or uses an optical zoom. Zoom feeds into all lighting conditions and flash mostly feeds into the low light condition.
Within the individual lighting conditions two viewing conditions are used for the performance evaluation. One is the viewing in 4 x 6 Inch format (small print, a little bigger than the display of most phones) and the 100% viewing (zooming in) on the display.

Image Quality aspects (primitives)

Select the specific conference and download the Authors Kit. The template may vary from one conference to another.

This template is set up for MS Word, but you may recreate it in other text programs, including LaTEX. After keying the text into a word file, apply the style tags. Please check the paper carefully to confirm that the styles have been applied correctly, then print it out and double check to ensure that the paper appears as intended.
A variety of image quality aspects have been identified as primitives in the times before cell phones had cameras integrated. These aspects are based on the performance of the optical system as well as on the performance of the sensor and the connected circuitry.

These primitives are:

- Exposure (ISO 12232 [6])
- Dynamic range and noise (ISO 15739 [7])
- Detail Reproduction (resolution) (ISO 12233 [8])
- Sharpness / Acutance (lens performance) (ISO 12233 [8])
- Optical aberrations like distortion (ISO 17850 [9]), chromatic displacement (ISO 19084 [10]), shading (ISO 17957 [11]), flare (ISO 18844 [12]) etc.

A few aspects can be added to these based on the image processing in the camera that is necessary to produce a good image from the captured raw data:

- Color (color reproduction) including chroma level and preferred color rendering
- White balance
- Sharpness /Acutance (added sharpening)
- Tonal Rendering

To compensate for the limitations in cell phones on size and cost the images require a much higher level of image processing. Especially dealing with low signal levels and the related amplification of noise has lead to new ways of image quality enhancement in these miniature cameras. Thus we can find additional aspects in cell phone cameras that need to be addressed and have not been issues with larger cameras.

These are:

- The reproduction of small randomly oriented and often times low contrast structures, publicly known as texture (ISO 19567 [13])
- Color shading (ISO 17957 [11])

In addition there are variety of image quality aspects that lead to scene related image processing like HDR imaging and tonal enhancement (with a single and multiple capture), scene related color enhancement, etc. For these aspects it will be difficult to define a measurement procedure because some of this has also a "preference" aspect to it.

Illumination dependency of image quality aspects

Optical aberrations

The optical aberrations like distortion, chromatic displacement, luminance shading and flare are more or less illumination level independent. Therefore it would be sufficient to measure these under bright light conditions where a measurement is most accurate due to the high signal level. But of course if one of these aspects shows a high aberration level under bright light it will limit the performance under the other light levels as well.
With the latest generation of phones distortion, chromatic displacement and luminance shading have not been issues anymore. None of the recently measured phones showed degradation of image quality based on these aspects. However for completeness they still need to be tested in case a manufacturer has not done his homework.
Flare can still be and issue depending on the lighting situation so it should be measured for completeness. Especially because of the cost and space restrictions it is difficult to minimize flare in small camera modules. On the other hand it is also difficult to characterize flare in a single number because of its dependency on the scene and its illumination conditions. For cameras with an extremely high flare level one will never measure a high dynamic range when using the chart based camera OECF measurement because the white patches will always ruin the level in the dark areas.

Figure 3: The impact of flare on image quality. The upper images shows a large amount of flare and the lower image does not. The variation between the two is a slightly different angle to the sun.

Exposure and Tonal Rendering

Generally exposure can be adjusted by using the three variables aperture, exposure time and amplification (exposure index). But we find that certain cameras tend to limit the exposure and intentionally underexpose images at low light situations. The maximum aperture is limited by size and lens design. The maximum exposure time is limited by handshake if there is not image stabilization built in and the maximum amplification is limited by noise performance. Therefore it can happen that a manufacturer decides to underexpose images rather than increasing exposure time or amplification.

The problem with judging exposure is that we only get access to the final rendered image out of which it is difficult to differentiate between exposure and the scene dependent tonal rendering. Even looking at a mid grey or digital levels in the highlights will not give us a definite answer if an image is underexposed or simply rendered too dark. In general however looking at the luminance reproduction of a x-rite Color Cecker SG will provide a general indication on exposure and for image quality it does not matter if an image is underexposed or rendered too dark. At this time there is no measurement procedure defined that allows judging the quality of the tonal rendering of an image especially because this is scene dependent.

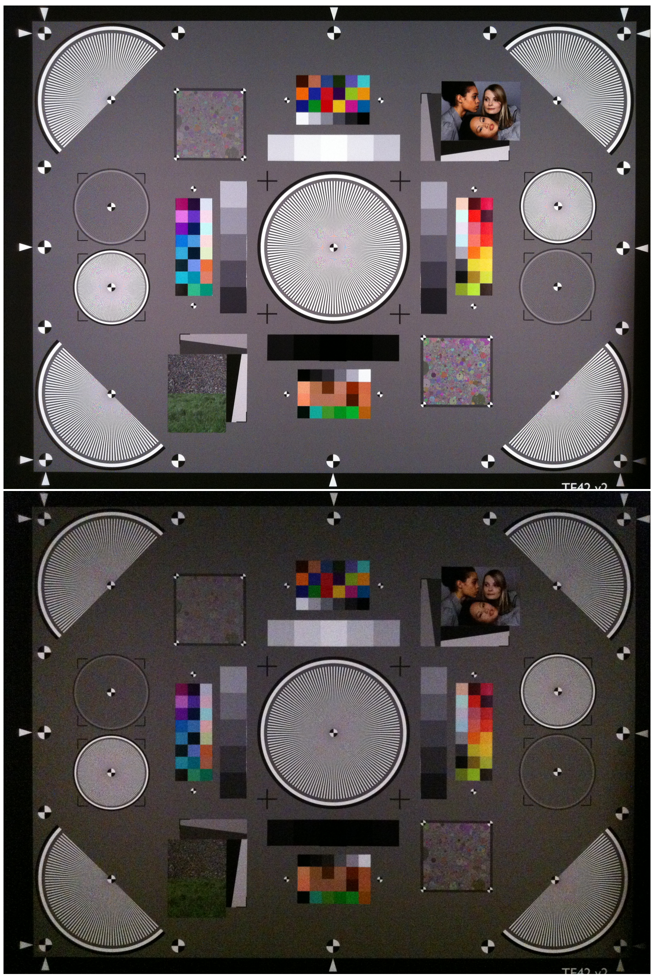


Figure 4: The upper image of the iphone 4 as captured with an illumination level of 100lx and the lower with one of 12 lx. The exposure time is limited to 1/15 sec. and making the image brighter would amplify the noise even more.

Dynamic range and noise

The higher the amplification level of the signal the lower the difference between the signal that produces the maximum digital output level and the signal level that equals the dark signal non uniformity (DSNU) which is called the dynamic range. This means the lower the light level the smaller the dynamic range. With increasing the amplification the noise increases as well. This means that dynamic range and noise are both illumination level dependent. With the given spectral sensitivity of a camera the individual color channels also need to be amplified to different levels when the spectral distribution of the light source changes. E.g. tungsten light sources have a low blue content and most sensors have a low blue sensitivity anyways this means that using tungsten light at low light levels increases the problem of amplification and therefore reduces image quality even further.

Denoising algorithms are used to overcome this problem. These algorithms lead to higher dynamic range and lower noise level measurements, which is the intention. But at the same time they have an impact on the reproduction of fine detail especially detail with low contrast because the camera can not differentiate between noise and the real scene content. That is the reason why the texture analysis needs to be part of the measurement as well.

Detail Reproduction (resolution)

Detail reproduction of fine detail depends on a variety of aspects that all feed into the measurement. This includes the quality of the lens, the accuracy and repeatability of the auto focus system, the sample rate (number of pixels), and some image processing aspects like demosaicing. To a small extend even sharpening has an impact on the measured resolution level. At low light conditions the smoothing effect of denoising algorithms leads to degradation in resolution. This degradation depends on the algorithms applied and affects structures in scenes based on the contrast and the orientation of the structures in different ways. Without the denoising applied resolution has proven to be almost light level independent as long as the signal is not completely covered by noise or the autofocus system stops working accurately.

The resolution measurement also needs to be performed over the imaging field. Many cameras in phones show a reasonable performance in the center but fail in the corners.

Texture

The term Texture defines structures that are often times randomly oriented and of low contrast. Noise reduction algorithms affect these structures because they cannot be differentiated from noise. A lot of research has gone into accurate ways to measure the degradation of texture [14] [15].

This image quality aspect highly depends on the illumination level and is essential for a quality rating of a camera in a phone.

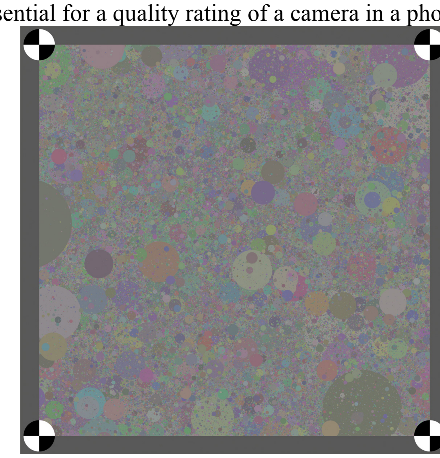


Figure 5: The original dead leaves structure for texture analysis.

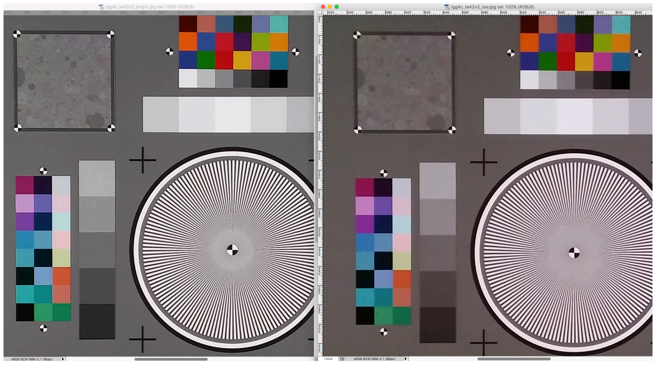


Figure 6: The LG4C shows a very high level of texture loss even under bright light conditions (image on the left side). Under low light (right image) the texture content is reduced even more and also the Siemens star shows some loss in the center.

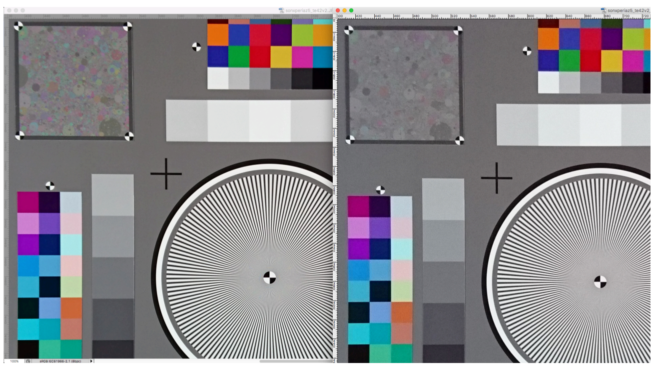


Figure 7: The Sony xperia Z5 shows a high texture loss from bright light condisitons (left) to low light conditions (on the right). But even though it is difficult to see in this image the loss in the high contrast Siemens star structure is by far not as high.

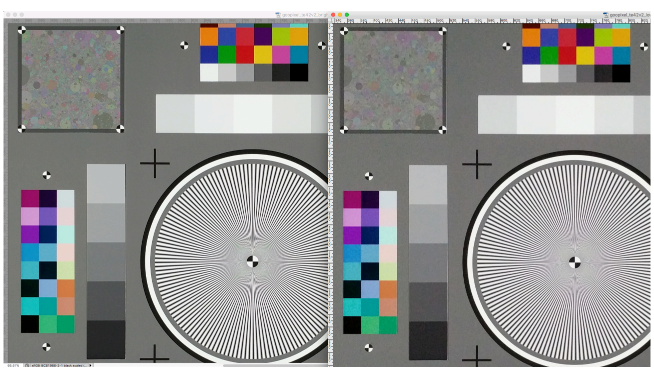


Figure 8: For the Google pixel we can see a slight low in texture and almost no loss in the Siemens star from bright (left) to low light (right).

Sharpness

The perceived image quality aspect called sharpness can be measured as the area under the spacial frequency response (SFR). It depends on the viewing condition and therefore needs to be weighted with the contrast sensitivity function (csf) of the human eye resulting in the acutance. The shape of the SFR depends on the transfer function of the optical system and also partly on the image processing in the camera (demosaic). It therefore also depends on the image height. The SFR and with it the image quality can be enhanced by applying sharpening. However adding too much of it or using the wrong algorithm can lead to artifacts in form of overshoot and undershoot around edges.

So either a measurement of overshoot and undershoot needs to be added to limit the sharpening or the maximum SFR values needs to be cut at a level of 1 and significant areas above the SFR need to lead to negative rating components.

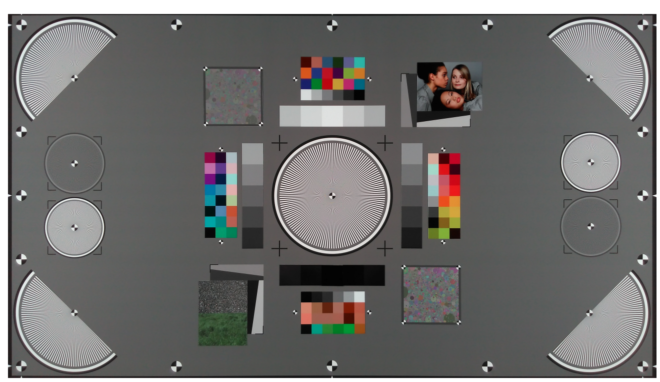
Sharpening in general is illuminance level independent. But when applied to noisy images shot under low light conditions it not only enhances the structures from the captured scene. It also increases the noise and its visibility.

Figure 9: The bright and dark stripes along the edge and the artifacts around the girl's hair are cause by strong sharpening.

Color Shading

Color shading mainly occurs in systems with wide fields of views and small sensors. It is a color and angle dependent change in transmission of the optical system. Often times the IR cut filter contributes to this.

Color shading does not depend on the light level but it depends on the spectral distribution of the illumination. Especially the IR content of the light is important. Therefore the color shading needs to be calibrated and also tested at daylight, tungsten (with IR) and also at fluorescent or LED spectra that almost contents no IR.



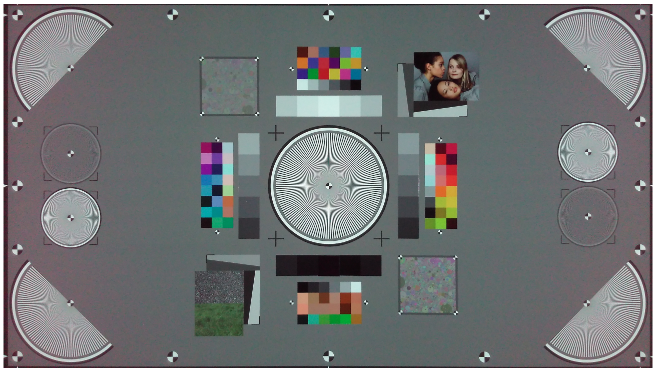


Figure 10: These two images are captured with the same cellphone and HTC desire 626. The upper image is captured under 1000 lux daylight (including IR) and the lower image was captured using the LED flash that does not contain IR. So the shift turns from red in the center with cyan surrounding to a cyan center with red surrounding.

Color

The color evaluation is one of the most difficult aspects. When capturing known colors the reproduction of color can easily be determined but on the one hand a colored test chart does not represent real world colors very well and on the other hand nobody wants a photographic camera to be a colorimeter. Colors are modified towards preferred colors with respect to saturation, optimized contrast, and enhanced colors like skin tones, blue sky etc.

For the saturation a psychophysical study has been performed by the CPIQ group. But it turned out that the optimum value is also slightly scene / original dependent. In general people prefer slightly (about 10%) higher saturated colors. Therefore saturation can be measured and rated but in general a color difference analysis can only be an indicator if something goes completely wrong.

White balance

In case of daylight images the images need to be neutral meaning that a neutral gray needs to be rendered to equal RGB values in the final sRGB encoded image. This can easily be measured on the reproduction of a neutral grey test target. For images captured under tungsten conditions or scenes like a sunset it is way more difficult. In most cases the warm atmosphere under tungsten illumination shall be preserved. But what is the optimum white balance in these cases? So far no consensus could be reached. So we have a way to measure the white balance but no aim values for tungsten condition.

Figure 11: A white balanced sunset (upper image) and a "natural" one.

Generating scores from the measurements

Basic concept

In case a cell phone camera fails completely in any one of the criteria mentioned above the image quality will be low. That means that all mentioned criteria have to be measured and analyzed to not miss any significant failure.

A typical way to create a rating system now is that each criterion will be analyzed separately and depending on the importance for the individual category (usage condition like illumination level) it gets a certain amount of points. The sum of points then leads to a score for the usage condition and the overall score is a weighted mean of all conditions based on the relative amount of images from a statistical analysis taken under the individual conditions.

Another potential way to get to a score is the one that the CPIQ group chose and that is described in the next paragraph.

The CPIQ process

The cell phone image quality group (CPIQ) has started off with those aspects that were easy to address and to measure. Those are usually the optical aberrations. That way the first 3 metrics that were created and that have by now also made it into ISO standards are local geometric distortion (LGD) [9], chromatic displacement [10] (mainly chromatic aberration) and color shading [11]. At this time the group also adopted the edge SFR measurement from ISO 12233 [8] but it was already clear that it could not be used to measure the reproduction of fine detail (resolution) because these

cameras use relatively high sharpening levels and a sharpened edge cannot be used to describe the detail reproduction. However sharpness in form of the acutance measure was introduced by using the edge SFR and even though the method is specified for different locations in the image the current document only uses the measurement in the center.
In this phase the ISO 15739 visual noise evaluation [6] was adopted and modified. A method for measuring the texture content [10] currently still translated as being low contrast fine detail on a so-called dead leaves target was developed and introduced. It is based on the acutance value of the power spectrum derived from the dead leaves structure. This means that cameras with a high sharpness level but a lower detail level at higher frequencies get a good rating even though a lot of the texture is gone. The latest method that was added to the catalogue of measurements is the chroma level. Additional characteristics are in preparation.
Given the current set of image quality parameters used by CPIQ it is easy to find examples that will have a good rating but have poor image quality. The LG4C is one of these.

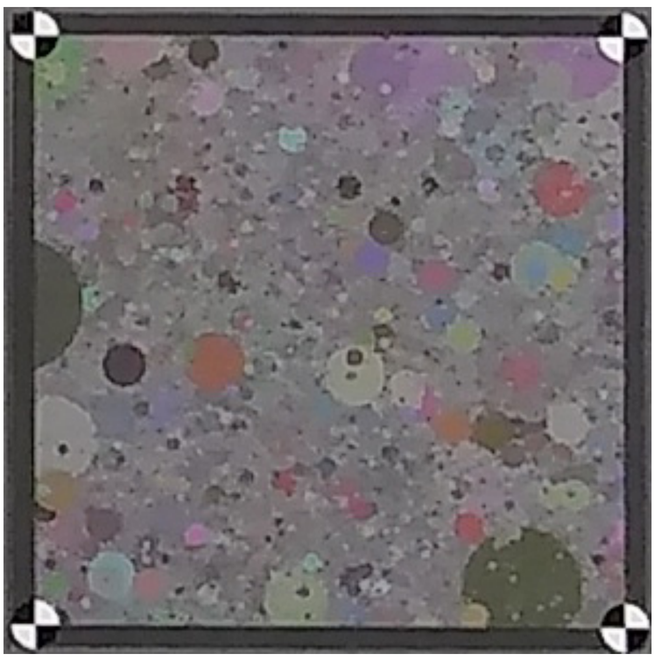


Figure 12: Texture acutance does not work for the LG4c. Because of the strong sharpening the cellphone gets an acutance value of around 1 (which should be the highest rating) but it misses a lost of high frequency detail.

For the CPIQ ranking three standard light levels and spectral distributions are discussed and used for the initial tests. These are 1000 lux at D55 daylight, 100 lux at TL84 Fluorescent and 25 lux or maybe 10 lux at 3050 K tungsten.

In contrast to the system using points for every quality criterion the group performed psychophysical measurements with test images viewed under standardized conditions. ISO 20462-3 has been used as the basis for this test. The evaluation assumes a perfect phone with ideal quality. All the imperfections measured with the image quality parameters add "quality losses" to it.
To find out about the losses images with known degradation were presented to observers and rated against a set of ruler images. That way it was possible to transfer the degradation of images in one parameter into just noticeable differences (JNDs) and therefore into a quality loss scale.

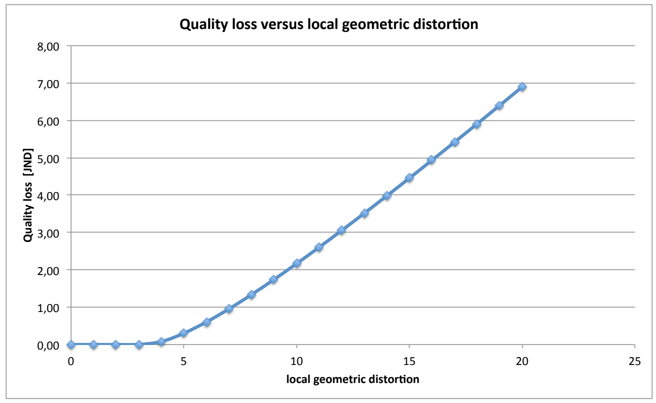


Figure 13: Quality loss for distorted images. The curve was originated from a psychophysical study.

Of course these tests were done under one viewing condition (looking at a 100 Ppi Monitor from 864 cm distance and the images being displayed in 100% mode meaning 1 pixel in the image corresponding to 1 monitor pixel) and it is a lot of work to perform these tests with sufficient observers for each image quality aspect. This has been a bottleneck in the work of CPIQ.
These quality ratings exist for all published CPIQ criteria and the theory from Keelan [4] is that the quality losses can be added up to give an overall quality loss for a specific usage condition. If a camera completely fails in one criterion it will get a huge quality loss from this criterion alone meaning that this approach should work better - especially for such extreme cases - than the method to provide a certain number points out of a max number for each criterion.

For a first run we evaluated the data of 9 phones using the CPIQ metrics with a slightly different (updated) approach for texture and visual noise measurements. So the current absolute numbers may not be exactly according to the CPIQ specs.

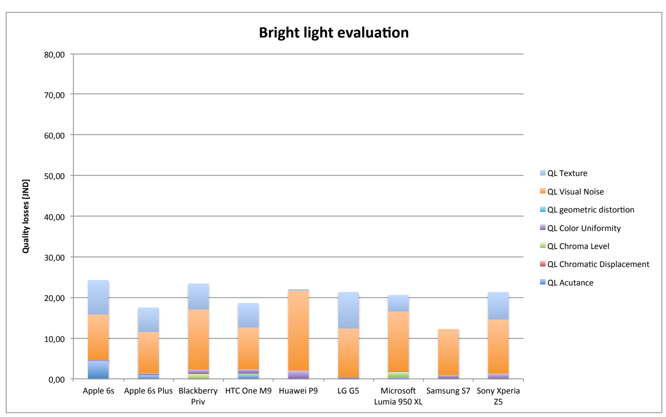


Figure 14: CPIQ aligned Quality losses for bright light conditions (the smaller the bars the better).

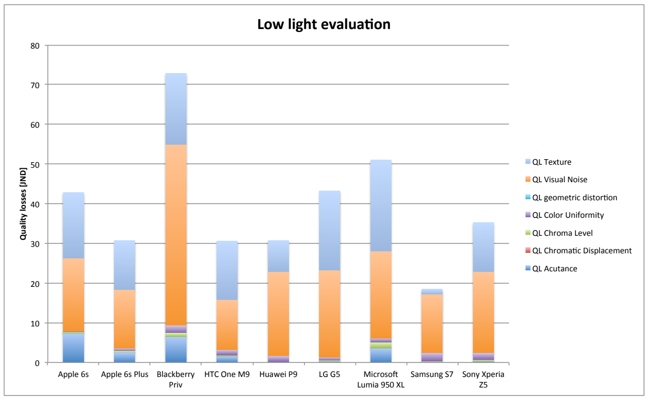


Figure 15: CPIQ aligned Quality losses for low light conditions.

Figure 14 and 15 show the difference in quality losses for bright and low light conditions. But they also show that there is no quality loss at all for 2 of the 6 image quality parameters and that is distortion and chromatic displacement. None of the current phones has a problem with these aspects. Also the chroma level is not a big issue for the phones. The differentiation happens in visual noise and texture. For low light the noise cleaning sometimes also results in a lower acutance level.

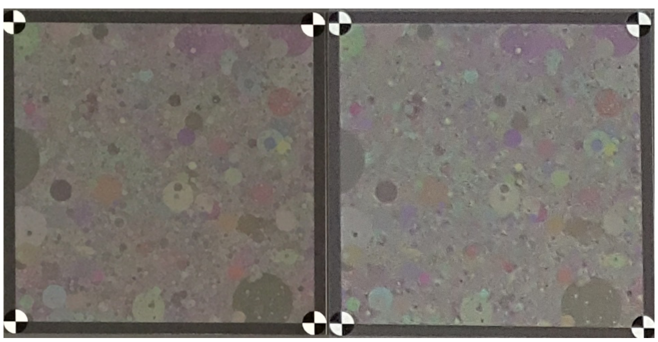


Figure 16: On the left hand side is the dead leaves structure captured with the iPhone 6s Plus and on the right hand side the same structure captured with the Samsung S7.

Figure 16 shows that the iPhone 6s Plus has as many details if not more than the S7 left in the image at low light. But the S7 gets a much better rating because the analysis uses texture acutance instead of the real texture level and the sharpening applied to the image of the S7 provides a higher acutance value and therefore a lower quality loss.

Unfortunately the two major aspects for differentiating the image quality are based on imperfect measurements. As seen in figure 12 and 16 the current texture analysis as described in the current CPIQ document does not work for all cameras [14] and the current CPIQ visual noise measurement method has issues as well with its weighted standard deviations. These problems will be addressed in future versions of the P1858 document. Also aspects like dynamic range, exposure and resolution have not been addressed yet. For the CPIQ procedure to become a reliable method for the characterization of cameras in mobile devices we will have to wait for the next generation of the CPIQ tests.

DxOMark and VCX

DxOMark and VCX use a huge number of parameters that are measured at different light levels to get to a final score. The downside of these two approaches is the lag of psychophysical studies for the individual parameters. Experts do the rating of the individual parameters. This may appear to be problematic but if the experts are trained well they can do a pretty good job on the ranking of the individual image quality measurements in case the measurements really reflect what can be seen in the images. This seems to be a problem with the power spectrum dead leaves analysis that is used by the DxO and CPIQ approach at the time this documents was written. To overcome this DxO openly speaks about a subjective evaluation, which they add to each group of aspects to overcome differences between the measurement results and the appearance of real images.

Ideally these differences should not exist if the measurements perfectly reflect the appearance in the image. Therefore the VCX approach uses the latest technologies including a texture approach that is not based on the noise corrected power spectrum [15] but on the intrinsic approach described in [14]. The VCX team hopes that this will work in all cases but given the continuous development in the area of cameras in mobile devices it is clear that every system used will require to be updated every few years to address new features and algorithms in the cameras.

Conclusions

- The quality loss based on just noticeable differences (JNDs) seems to be the most promising approach when it comes to a ranking system for camera quality because it addresses the fact that a failure in a single factor can ruin an image. It also accounts for the visibility of a factor in the image.
- The psychophysical studies are time consuming and therefore expensive and they only address one specific viewing condition.
- The current CPIQ measurement that will be published early 2017 is not sufficient to characterize a camera and the current methods need an update to work with latest cameras.
- VCX is the ranking system that includes the latest technical developments in image quality measurements. Therefore the measurements reflect the image quality as best as possible.
- The VCX ranking derived from the correct measurements may fail under certain conditions especially if a camera fails in just a single image quality aspect.
- All procedures will require constant work and updates because of the developments in camera technology.

References

[1] IEEE P1858 Standard under development, http://grouper.ieee.org/groups/1858/.

[2] DxOmark, https://www.dxomark.com.

[3] VCX, Valued Camera eXperience, http://camtest.eu/vcx-md/mobile-devices

[4] Handbook of Image Quality, Brian W. Keelan, CRC Press, ISBN 0-8247-0770-2

[5] Dietmar Wueller, Reiner Fageth, Statistic Analysis of Millions of Digital Photos, Proc. SPIE 6817, Digital Photography IV, March 2008

[6] ISO 12232 Photography — Digital still cameras — Determination of exposure index, ISO speed ratings, standard output sensitivity and recommended exposure index

[7] ISO 15739 Photography - Electronic still-picture imaging - Noise measurements

[8] ISO 12233 Photography - Electronic still-picture cameras - Resolution measurements

[9] ISO 17850 Photography - Digital camera - Geometric distortion (GD) measurements

[10] [ISO 19084 Photography - Digital cameras - Chromatic displacement

[11] ISO 17957 Photography - Digital cameras - Shading measurements

[12] ISO 18844 Photography — Digital cameras — Image flare measurement

[13] ISO 19567-2 Photography - Digital cameras - Part 2: Texture analysis on non cyclic pattern

[14] Kirk, Herzer, Artmann, Kunz " Description of texture loss using the dead leaves target: Current issues and a new intrinsic approach", Proc. SPIE 9023, Digital Photography X, 90230C (EI2014)

[15] John McElvain, Scott P. Campbell, Jonathan Miller and Elaine W. Jin, "Texture-based measurement of spatial frequency response using the dead leaves target: extensions, and application to real camera systems", SPIE EIC 2010

[16] [16] ISO 20462-3 Photography - Psychophysical experimental methods for estimating image quality - Part 3: Quality ruler method

Author Biography

Dietmar Wueller studied photographic technology at the Cologne University of applied sciences. He is the founder of Image Engineering, an independent test lab that tests cameras for several photographic and computer magazines as well as for manufacturers. Over the past 20 years the company has also developed to one of the world's leading suppliers of test equipment. Dietmar Wueller is the German chair of the DIN standardization committee for photographic equipment and also active in ISO, the IEEE CPIQ (Cellphone Image Quality) group, and other standardization activities.