https://doi.org/10.2352/ISSN.2470-1173.2017.12.IQSP-249

Towards the Development of the IEEE P1858 CPIQ Standard – A validation study

Elaine W. Jin [a], Jonathan B. Phillips [a], Susan Farnand [b], Margaret Belska [c], Vinh Tran [c], Ed Chang [a], Yixuan Wang [c], Benjamin Tseng [d]
a. Google Inc, Mountain View, CA/USA; b. Rochester Institute of Technology, Rochester, NY/USA; c. Nvidia, Santa Clara, CA/USA; d. Apkudo, Baltimore, MD/USA.

Abstract

The IEEE P1858 CPIQ Standard is a new industry standard for assessing camera image quality on mobile devices. The CPIQ standard provides test methodologies for evaluating seven image attributes: spatial frequency response, texture blur, visual noise, color uniformity, chroma level, lateral chromatic displacement, and local geometric distortion. In addition, the CPIQ standard provides mathematical transforms between objective metric values and perceived image quality quantifiable in just noticeable differences, and a framework to combine individual attributes into prediction of overall image quality. This study aims at validating the CPIQ set of image quality metrics and the CPIQ prediction of overall image quality. The two key components of the study are objective measurements of image quality in the lab and subjective evaluation of real-world images by human observers. Nine smartphones were used in the study, with the expected camera quality ranging from low to high. The CPIQ methodology was implemented and practiced in an industrial lab, and measurements of the CPIQ metrics were obtained at varying lighting conditions. The subjective evaluation study was performed in a university lab, using paired comparison and softcopy quality ruler as test methods. The results from this study revealed that objective measurements defined in the CPIQ standard are highly correlated with perceived image quality.

Introduction

The IEEE P1858 CPIQ (Camera Phone Image Quality) Standard has recently been published after a decade of development. This standard attempts to establish a uniform means of evaluating the quality of cameras in mobile devices, allowing objective comparison between device models and manufacturers, using a variety of metrics that are relevant to consumer photography. More than 30 companies have participated in the development of the standard since the inception of CPIQ in 2006 under I3A (International Imaging Industry Association). In 2012, the working group transitioned to IEEE standards development.

The fundamental building blocks of the standard include a set of objective metrics (OM) with psychophysically obtained formulations for calculating predicted quality loss (QL) values for individual metrics [1-8] and a total quality loss using a Minkowski summation of the individual QL values [9]. These QL values are in units of secondary Standard Quality Scale Just Noticeable Differences (SQS2 JNDs) as defined in the ISO 20462-3 standard [10].

For the spatially dependent metrics [11], the impact of the objective metric level on image quality varies depending on the intended usage. Therefore, the standard provides guidance for selection from a set of given use cases such as viewing on a computer monitor at 100% magnification, on a smartphone display, or via a print. For each of these use cases, the critical component is knowing the spatial frequency distribution of the observed image, as specified in units of cycles per degree on the human retina. This information enables modeling of perception by incorporating achromatic and chromatic contrast sensitivity functions (CSFs) into the calculation process.

The CPIQ set of image quality metrics includes seven metrics, providing a comprehensive coverage of camera image quality for spatial attributes, color attributes, and imaging artifacts. In early 2016, the CPIQ Conformity Assessment Steering Committee (CASC) commissioned a validation study to examine the validity of the CPIQ set of image quality metrics. In this study, nine smartphones were used to perform rigorous lab test using the CPIQ metrics. Overall quality loss was obtained by combining the quality loss values predicted for individual attributes. In parallel, a set of real-world images was captured using the same nine smartphones. These images were subsequently evaluated by a group of human observers for overall image quality. The validity of the CPIQ set of metrics was examined using the subjective evaluation results as ground truth.

The main purpose of the CPIQ validation study is to validate the CPIQ set of image quality metrics. A secondary purpose of the study is to practice the softcopy quality ruler method [12-13] as a camera-benchmarking tool. The paired comparison method is used as a reference method.

This paper is organized as follows. In Methods we will discuss the experimental design, device selection, procedures for objective measurements, real-world scene capture, and the two subjective evaluation methods. In Results we will present the results from objective and subjective studies, followed by checking on the correlation between the two. In Conclusion we will summarize the learning from this study and discuss future works.

Methods

Experimental Design

This study involves both objective measurements in the lab and subjective evaluations using real-world images. The steps of the experiment are:

1. A set of smartphones will be selected to use in the test, with the range of image quality from low to high.
2. A set of camera capture modes will be selected to use in the camera test, which includes settings on 3As (auto focus, auto white balance, auto exposure), HDR (high dynamic range), flash, resolution, and timer.
3. An industrial camera test lab will be selected to perform the CPIQ measurements.
4. A set of real-world scenes will be defined and captured with the same set of cameras.
5. Subjective evaluation studies will be carried out to evaluate the real-world scenes.
6. The relationship of subjective evaluation and CPIQ prediction of overall image quality will be examined.

Device selection

The main consideration in device selection is the range of image quality performance for the rear-facing cameras. Table 1 shows the nine smartphones selected for this study. These devices were first launched as new products between 2010 and 2016, and they represented the flagship devices from seven of the world's leading smartphone manufacturers, including Apple, Samsung, Nokia, Google, HTC, LG, and Sony. The DxOMark Mobile Photo scores for the rear-facing cameras on these devices cover a range of 50 (the lowest score to date) to 88 (the highest score as of June 2016).

All smartphone cameras can be configured into different modes with various camera settings. With input from the CPIQ member companies, a decision was made to turn off the HDR setting of a camera if available in the menu in order to maximize the similarity between the capture of charts in lab conditions and those conditions encountered in the capture of real-world images. Settings with HDR ON tend to apply localized image processing, which can result in diverging performance assessment between lab and the field. All cameras were used in auto mode for auto focus, auto white balance, and auto exposure. The flash was turned off, and a tripod was used in all captures. The timer was used in all captures when present in the smartphones.

Table 1: List of smartphones and camera settings

Phone ID	OEM	DxOMark Photo	Pixel Count (MP)	Aspect Ratio
1	Apple	50	5	4:3
2	HTC	70	4	16:9
3	LG	77	13	4:3
4	Apple	78	8	4:3
5	Nokia	79	5	16:9
6	Apple	84	8	4:3
7	Google	86	12.3	4:3
8	Sony	88	8	16:9
9	Samsung	88	12	4:3

As shown in Table 1, the resolution of the cameras ranges from 4 to 13 megapixels (MP). For two of the phones (#5 and #8), lower resolution settings were used, which made them fitting in the resolution range of the other smartphones. This step is needed in order to facilitate the comparison for the 100% magnification use case in the objective and subjective evaluations. As stated, the purpose of the validation is to validate the CPIQ set of metrics, not to characterize any particular phones. Thus, resolution choices were made from a design-of-experiment point of view.

Objective Measurements

The validation study quantifies the seven metrics included in the version 24 of the P1858/D1 Draft Standard for Camera Phone Image Quality [14]. These metrics are indicated in the list below.

1. Spatial frequency response (SFR)
2. Lateral chromatic displacement (LCD)
3. Chroma level (CL)
4. Color uniformity (CU)
5. Local geometric distortion (LGD)
6. Visual noise (VN)
7. Texture blur (TB)

Various charts are needed for testing the seven metrics. A slanted edge with 4:1 contrast level is used for SFR (Fig. 1(a)). A chart with black dots on white background is used for LCD and LGD (Fig. 1(b)). A ColorChecker SG is used for chroma level (Fig.1 (c)). A reflective white card is used for color uniformity (not shown in Fig.1). An OECF (opto-electronic conversion function) chart is used for visual noise (Fig. 1(a)). A monochrome dead leaves pattern is used for texture blur (Fig. 1 (d)).

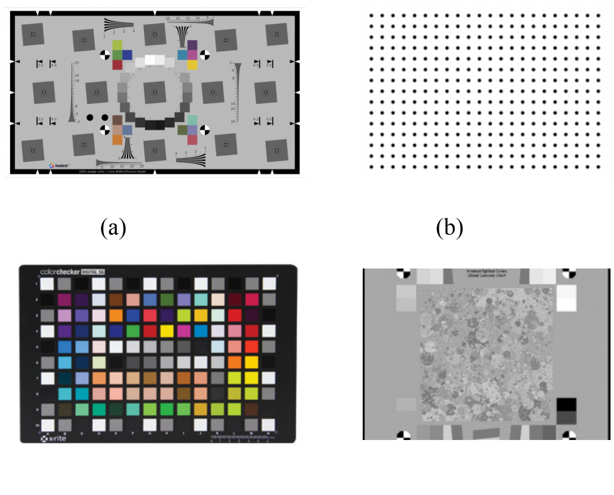

Figure 1. Test charts used in the CPIQ measurements. (a) Slanted edge and OECF; (b) dot chart; (c) ColorChecker SG; (d) Dead leaves.

Quantification includes the respective objective metric and subsequent QL (quality loss) in JND (just noticeable difference) units. For spatially dependent metrics, i.e., spatial frequency response, visual noise, and texture blur, the objective metric assumes a viewing condition of a 100 ppi (0.254 ppm) monitor and 100% magnification viewed at 34 inches (86 cm). This corresponds to a cutoff spatial frequency of 29.6 CPD (cycles per degree) and a k_{disp} of 0.0243 for the display device MTF. Note that the current process of calculating the lateral chromatic displacement has an objective metric in pixel units and that the image evaluation to obtain the QL conversion was calculated from the subjective data with images viewed as above. Thus, for this experimental condition, the pixel units correspond directly to arc-minutes.

Once the seven individual QL values have been obtained for a given capture condition, they are combined to generate a predicted overall QL. The Minkowski metric used to combine the QL values is shown in eq. (1).

$$QL = (\sum_i (QL_i)^{n_{max}})^{(1/n_{max})} \quad (1)$$

where $n_{max} = 1 + 2 \cdot tanh(QL_{max}/16.9)$ and QL_{max} is the maximum QL for a given test condition for a given camera. The *tanh* function accounts for the dominating QL in the summation, as is typical for image quality evaluation where the maximum quality degradation dominates the overall perception.

The lab capture conditions are described in Table 2. Note that three categories can be obtained from this regarding simulations of night/dim, indoor, and outdoor simulations, that is 25 lux, 100 lux, and 500 lux, respectively.

Table 2: Lab lighting conditions using X-Rite Spectralight QC

Category	Lux	CCT (K)	Type
Night/Dim	25	2700	U30
Indoor	100	3700	TL84
Daylight	500	6500	D65

Subjective Evaluation

The subjective evaluation study involved generating the image quality ratings associated with real-world test images using both Paired Comparison and Softcopy Quality Ruler protocols. Paired comparison is a well-established psychophysical research method for assessing image quality differences [15-16]. Softcopy quality ruler [12-13] was developed in parallel to the CPIQ developmental work and was extensively used in calibrating the CPIQ set of image quality metrics. This is the first study to compare the Softcopy quality ruler method with the classical paired comparison method for use in camera benchmarking. The images created were used in both studies and judgments for both studies were made under essentially the same viewing conditions, allowing for comparison of results. The experiments were conducted in the Munsell Color Science Lab at the Rochester Institute of Technology.

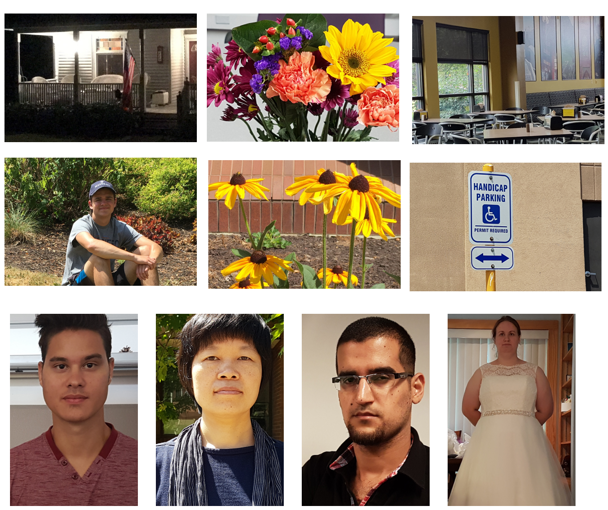

Figure 2. Ten test scenes used in the validation study. Scene names from upper left: House at Night, Flowers-Pink LED, Empty Restaurant, Person in Garden, Yellow Flowers, Handicapped Sign, Portrait – Low Light, Portrait – Outdoors, Portrait – Pink LED, Wedding.

The nine phones shown in Table 1 were used in this study. To align with the reference viewing condition used in the objective measurements, the captured images were to be displayed at 100% magnification, i.e., 1 camera pixel corresponding to 1 display pixel. The nine cameras have different resolutions, aspect ratios, and fields of view. In order to capture a variety of scenes with all phones and displaying similar content at 100% magnification on the monitor, the camera-to-subject distance would need to vary among cameras. In particular, the cameras with higher pixel heights needed to be moved farther from the target. A look-up table (LUT) was generated for each phone in order to give an idea of the distances required between the target and the phone. Several images were taken at around the distance calculated from the LUT. The pixel height of an object in the field of view was measured in each image; the images where the pixel height of the object was consistent for each phone were then chosen.

Ten real-world scenes were selected to represent the variety of illumination levels and subject matters in today's consumer photos. Figure 2 shows the thumbnails of the 10 test scenes. Table 3 further provides information on the illumination level and light source types for these real-world scenes. When selecting scene contents, consideration was given to scene contents that match closely to the softcopy ruler image set. Scene similarity would help when these images were to be evaluated in the softcopy quality ruler study.

Table 3. Capture lighting conditions for the real-world scenes

Scene Name	Illuminance (lux)	Lighting
House at Night	15.9	Incandescent
Portrait - Low Light	15.4	Daylight
Portrait - Pink LED	140.1	LED
Empty Restaurant	257.8	Daylight & Fluorescent
Wedding	309.5	Daylight & Fluorescent
Flowers - Pink LED	422.8	LED
Handicapped Sign	> 99,999	Daylight
Person in Garden	> 99,999	Daylight
Yellow Flowers	> 99,999	Daylight
Portrait - Outdoors	> 99,999	Daylight

Once all the images had been captured, the images were cropped in order to fit properly in the GUI window used to present the images to the observers. The landscape-oriented images were cropped to 1253x834 pixels and the portrait-oriented images were cropped to 834x1253 pixels. Both the paired comparison study and the softcopy ruler study used the same set of cropped images.

Twenty observers participated in each of these two experiments. The observers were screened for normal or corrected-to-normal acuity using a Snellen chart at a distance of 20 feet or at the observation distance of 34 inches with a scaled Snellen chart. The observers' color vision was also tested using the Ishihara Plate Test of pseudo-isochromatic plates. A chin rest was used for the quality ruler test in order to keep the participants at exactly the same viewing distance and angle for the entire test session. Since the paired comparison test was not as dependent on acuity, it was decided that the chin rest was not necessary for that portion of the experiment.

In the paired comparison study, the participants were presented a pair of images of the same scene, taken using two different smartphone cameras (see Fig. 3). They were told to choose which image of the pair they preferred. The preference could be a function of several factors, including sharpness, color, and noise. They were also verbally instructed to ignore the facial expressions in images that had people as the target; this was because the facial expressions were not always consistent through each scene and it was not desired to bring in an additional factor of the psychology behind, for example, preferring a smiling face to a neutral face. The images presented in one pair were always from the same scene but the order of presentation of the images was randomized for each observer.

Figure 3. An example pair of test images used in the Paired Comparison Study.

In the softcopy quality ruler study, a pair of images was presented to the observer, a ruler image and a test image. The observers were asked to use a slider bar to adjust the sharpness of the ruler image until they felt that the quality of the ruler image was equal to that of the quality of the test image. The ruler images were supplied with the ruler package, and the images from the nine smartphone cameras served as the test images. The ruler images were selected carefully to match the contents of the test scenes. For example, a Restaurant image from the ruler set was matched with the Empty Restaurant test scene (see Fig. 4). The GUI ran through all of the images of a particular set before moving on to the next set. The order in which the images within a set were presented was randomized. Additionally, a ruler image was inserted into each set of test images to serve as a null image for determining how accurately the participants were completing their task.

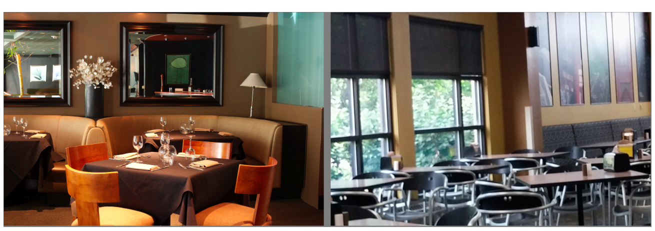

Figure 4. An example pair of ruler image (left) and test image (right) in the Softcopy Quality Ruler experiment.

Analysis was performed on the null image responses of the 20 observers. Based on the established criteria [13], data from 2 observers were identified as deviating from the group means. As a result, data from these two observers were excluded in the following presentation of the study results for both paired comparison and softcopy quality ruler.

Results

Objective Measurement Results

Objective measurement results were obtained from seven CPIQ metrics for each of the nine smartphones. Imatest Master 4.4.12 was utilized to obtain the objective metrics (OMs) and quality losses (QLs) for each of the captures conditions. QL results are shown in Table 4.

Table 4. CPIQ QL results in JNDs for the 9 phones

U30/25lux	1	2	3	4	5	6	7	8	9
VN	11.10	2.40	4.50	7.70	3.70	8.30	4.60	3.80	1.80
SFR	5.46	0.09	5.91	1.74	0.65	0.00	0.00	2.20	0.00
TB	15.95	4.59	18.36	12.04	3.02	10.18	9.12	10.76	2.18
CL	0.53	1.95	0.96	0.13	4.59	0.56	1.64	1.44	0.06
CU	6.93	1.72	0.68	0.33	1.17	0.97	0.77	1.38	0.64
LGD	0.03	0.15	0.45	0.47	0.10	0.26	0.55	0.21	0.33
LCD	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
TL84/100lux									
VN	8.10	2.10	3.50	5.70	4.30	3.30	3.60	3.60	1.80
SFR	1.11	0.00	0.61	1.69	0.32	2.57	0.00	0.00	0.00
TB	6.12	3.38	9.37	10.53	2.06	9.72	2.60	3.83	0.00
CL	1.93	0.75	0.65	0.35	4.20	0.19	0.31	0.35	0.39
CU	1.96	0.48	0.14	0.98	1.27	0.40	0.17	1.35	0.28
LGD	0.54	0.01	0.46	0.01	0.33	0.20	0.38	0.01	0.20
LCD	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
D65/500lux									
VN	5.90	2.40	3.00	5.00	3.10	4.10	3.10	2.40	2.50
SFR	0.00	0.01	0.06	0.15	0.14	0.00	0.00	0.09	0.00
TB	4.00	2.08	0.00	5.47	0.37	4.59	0.96	5.32	0.00
CL	0.06	2.00	0.72	1.92	0.20	0.17	0.02	1.62	0.32
CU	1.13	0.75	0.44	0.97	0.65	0.26	0.04	0.62	0.00
LGD	0.07	0.51	0.39	0.68	0.19	0.41	0.32	0.06	0.65
LCD	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

It can be seen from this table that the main driver for camera performance is texture blur and visual noise, especially under low light conditions. Spatial frequency responses can also go down significantly in low light conditions. Another measure worth noting is the color uniformity. Phone #1 suffers from severe color shading, and the measured quality loss due to color uniformity is significant for U30, 25 lux (6.93 JNDs).

Subjective Evaluation Results

The Paired Comparison experiment yields results in the form of probabilities that each test image was chosen over the other images of the same scene in the experiment. These probabilities were transformed into Z-scores. Figure 5 shows the Z-score values from the 9 smartphones and the 10 test scenes. Here positive Z-score values are associated with above average performance, and negative Z-score values are for below average performance. It can be seen that there is a spread in Z-score values for all phone models, indicating that the camera quality level changes with scene content and light level.

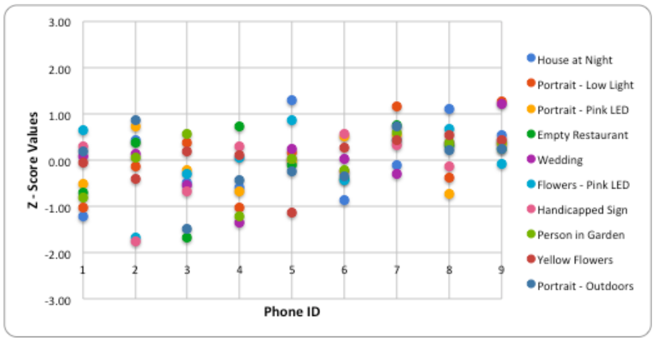

Figure 5. Study results from the paired comparison study, reported as Z-score values.

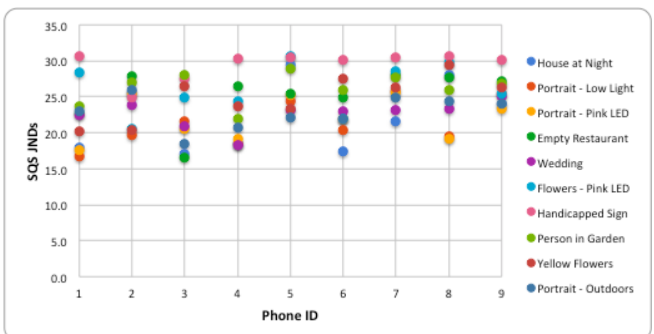

Figure 6. Study results from the softcopy quality ruler study, reported as SQS JNDs.

The results from the softcopy quality ruler experiment are shown in Figure 6. SQS is an absolute quality scale, with SQS > 30 at DSLR quality. Again it can be seen that for the same phone model image quality can vary significantly with scene content and light level. The data suggests that majority of the phones can reach high quality (SQS > 30) under ideal lighting conditions (e.g., daylight). However, the quality level drops when light level is low (e.g., at night). Low light performance is definitely a differentiating factor between high and low quality phones.

In order to validate the softcopy quality ruler method, correlation coefficients were calculated for all test scenes and reported in Table 5. A linear relationship was expected between the SQS values and the Z-score values because both measurements would yield results on a perceptually uniform scale. A high correlation between the two methods would be an indication of the validity using softcopy quality ruler method in camera benchmarking tasks.

Table 5. Comparison of paired comparison and softcopy quality ruler study results

Scene	Correlation Coefficient
Portrait - Low Light	0.936
House at Night	0.962
Flowers - Pink LED	0.927
Empty Restaurant	0.947
Portrait - Pink LED	0.831
Wedding	0.931
Person in Garden	0.855
Portrait - Outdoors	0.974
Handicapped Sign	0.929
Yellow Flowers	0.639
Mean correlation	**0.893**

Table 5 shows that for 9 out of 10 test scenes the correlation between the two sets of results is very strong (r > 0.83). Overall the mean correlation reaches 0.89. There is only one outlier in the set, i.e., Yellow Flowers. A closer examination revealed that the outlier in the set is Phone #5. In paired comparison, the Z-score for this scene is -1.12, the lowest Z-score among the 10 scenes for this phone. In the softcopy ruler, however, the same scene was rated as SQS = 23.3, in the middle of its SQS range. This scene has a greenish colorcast, making the yellow flower and surrounding areas slightly off on color. It is possible that the color error was more obvious when comparing with other images of the same scene content than when it was compared to a ruler image with a different scene content.

Correlation between Objective and Subjective Results

In order to test the correlation of objective results with subjective results, the 10 real-world scenes were grouped by lighting conditions. Night/dim scenes include House at Night and Portrait – Low light. Indoor scenes include Portrait – Pink LED, Empty Restaurant, Wedding, and Flowers – Pink LED. Daylight scenes include Person in Garden, Portrait – Outdoors, Handicapped Sign, and Yellow Flowers.

For each lighting condition, the seven QL values obtained for each of the phones as shown in Table 4 were combined in a total QL using the Minkowski summation as shown in Eq. 1. Recall that the *tanh* function accounts for the dominating QL in the summation, as is typical for image quality where the maximum quality degradation dominates the overall perception. This is an important factor as often the texture blur QL is the dominant component, particularly in lower lux conditions. Recall also that each metric's QL was obtained via a pre-established CPIQ model that utilized evaluation with the ISO 20462-3 quality ruler such that the conversions from the OM values to QLs are into the same SQS JND scale. This means that each of the metrics is "leveled" before being combined with the Minkowski summation.

In order to make a direct comparison between the objective results and the subjective results, the SQS values obtained from the softcopy quality ruler study were converted into quality loss values by applying Eq (2):

$$SQS\ QL = 32.10 - SQS \quad (2)$$

Where 32.10 represents the max SQS value achievable in the softcopy quality ruler method.

The results for the total QL from the objective prediction and subject evaluation for each of the three light level categories are tabulated in Table 6. Note that the QL predictions from both objective metrics and the subjective results have the same units (JNDs) and can be compared directly.

Table 6. Summary of objective vs. subjective results

	CPIQ QL Prediction (JND)			Subjective QL Results (JND)		
Phone ID	U30 25lux	TL84 100lux	D65 500lux	SQS Night	SQS Indoor	SQS Daylight
1	19	11	8	14.216	9.33	7.7
2	6.9	5	5.6	9.41	7.71	7.55
3	18.9	10	3.6	12.72	11.37	6.93
4	13.9	11.9	8.7	13.58	10.04	7.95
5	8.2	7.9	3.6	5.26	6	5.89
6	13	10.5	6.9	13.25	9.17	5.77
7	10.4	5.3	3.7	8.4	5.77	4.72
8	11.6	6.5	6.7	8.34	7.13	4.53
9	3.9	2.3	3	7.35	6.84	5.31

Table 7 shows a summary of the relationship between objective and subjective results, including the mean QL error, mean absolute QL error, and correlation coefficient for the 3 lighting conditions. The mean QL errors indicate the potential system bias in the CPIQ model prediction. For all 3 lighting conditions this error is less than 1.5 JNDs. More specifically, the data shows that the mean QL is somewhat overestimated for the Night category and mildly underestimated for the Indoor and Daylight categories. The mean absolute QL errors show the quality of model fit to the subjective data. For indoor and daylight conditions the error is less than 2 JNDs. For night condition this number is a bit higher (2.85 JNDs). An explanation for this deviation is shown below in the discussion for Fig. 7. The correlation coefficients between objective and subjective are moderately high (>0.7) for night and indoor conditions, and slightly lower for the daylight condition (0.49).

Table 7. Summary of objective vs. subjective results

	Mean QL Error	Mean abs QL Error	Correlation Coefficient
U30/25lux	1.47	2.85	0.79
TL84/100lux	-0.33	1.83	0.73
D65/500lux	-0.73	1.69	0.49

Fig.7 is a visual illustration of the study results from all 3 lighting conditions and all 9 phones. Note that the majority of comparisons fall along the line of equality, as is theoretically expected. The correlation between the objective and subjective results for the entire dataset is high at R = 0.83. Some data points in the night condition category show strong over-prediction compared to the subjective QL (e.g., the two blue diamonds to the right of the diagonal line). One could hypothesize that the image quality degraded to a saturation point at about 14 SQS JNDs in the real-world images. However, that value of 14 on the scale is only in the middle of the quality ruler, so further degradation could have been readily selected by the observers in the subjective experiment. Therefore, a more plausible cause would relate to the QL values from the objective metrics. In fact, closer inspection of the OM values and the conversions to QL reveals that the texture acutance values went out of range for the pre-established QL loss functions in the CPIQ standard. More specifically, any OM value less than 63% texture acutance is out of range. The 2 outliers with > 18 total QL predictions have OM values of 36.1% and 36.6%, well below the minimum of the range. Thus, the total QL predictions from these out-of-range OM values are extrapolated and have high probability of being inaccurate. Another factor is the synergy between the edge acutance and texture acutance; for the night conditions, the edge acutance is lower in part due to the increased noise. This edge acutance level, in turn, reduces the texture acutance in a synergistic manner.

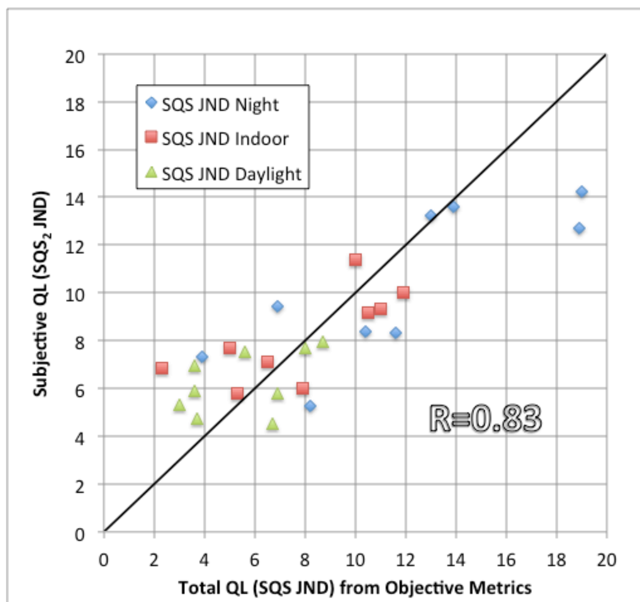

Figure 7. Correlating objective and subjective results for all scene categories and all phones.

Conclusions

The aim of this paper was to describe a validation study of the IEEE P1858 CPIQ Standard, which involved capturing both analytical charts in the lab and real-world images. The results from this study revealed that objective measurements defined in the CPIQ standard correlated highly with perceived image quality for the given set of nine smartphones. In addition, the experimental data showed that the subjective results using a softcopy quality ruler method correlated highly to the results using a paired comparison method.

Future CPIQ work is currently focused on completing the round robin study with a total of 7 labs testing the cameras. The results will be compared to determine repeatability and accuracy of the metrics to be released in the IEEE P1858 CPIQ Standard. In addition, revisions are in process for existing metrics as well as addition of metrics such as AWB and AE performance. As more metrics are developed by the CPIQ working group, they will be incorporated into the total QL calculations and should provide even higher correlation to the real-world subjective data presented in this paper.

Acknowledgement

The authors thank Scott Anthony and Miles Nielsen for their test engineering support at Google for the objective metric testing. The authors would also like to thank Katherine Carpenter and Morteza Maali Amiri at Rochester Institute of Technology for their help in developing and performing the subjective evaluation studies.

References

[1] F. Cao, F. Guichard and H. Hornung, ”Measuring texture sharpness of a digital camera”, Proc. SPIE 7250, 72500H, 2009.

[2] Jonathan B. Phillips, Stephen M. Coppola, Elaine W. Jin, Ying Chen, James H. Clark, Timothy A. Mauer, “Correlating objective and subjective evaluation of texture appearance with applications to camera phone imaging,” Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 7242, 724207, 2009.

[3] F. Cao, F. Guichard and H. Hornung, ”Dead leaves model for measuring texture quality of a digital camera”, Proc. SPIE 7537, 75370E, 2010.

[4] J. S. McElvain et al., ”Texture-based measurement of spatial frequency response using the dead leaves target: extensions, and applications to real camera systems”, Proc. SPIE 7537, 75370D, 2010.

[5] Jonathan Phillips, “Toward a Camera Phone Image Quality Rating Scale: A Multivariate Approach,” 6Sight Future of Imaging Summit, 2011.

[6] Brian W. Keelan, Robin B. Jenkin, and Elaine W. Jin, “Quality versus Color Saturation and Noise”, 2012.

[7] Donald J. Baxter and Andrew Murray, “Calibration and Adaptation of ISO Visual Noise for I3A’s Camera Phone Image Quality Initiative”, Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 8293, 829303, 2012.

[8] Donald Baxter, Jonathan Phillips, and Hugh Denman, “The Subjective Importance of Noise Spectral Content”, Proc. SPIE, 2013.

[9] B. W. Keelan, “Predicting Multivariate Image Quality from Individual Perceptual Attributes”, Proc. IS&T's PICS 2002 Conference, Portland, Oregon, Society for Imaging Science and Technology, Springfield, Virginia, pp. 82–87, 2002.

[10] ISO 20462-3:2012 Photography -- Psychophysical experimental methods for estimating image quality -- Part 3: Quality ruler method, 2012.

[11] Donald Baxter, Frédéric Cao, Henrik Eliasson, and Jonathan Phillips, “Development of the I3A CPIQ spatial metrics”, Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 8293, 829302, 2012.

[12] Jin, E. W., Keelan, B. W., Chen, J., Phillips, J. B., and Chen, Y., “Softcopy quality ruler method: Implementation and validation,” Proc. SPIE 7242, 724206, 2009.

[13] Jin, E. W., & Keelan, B. W., Slider-adjusted softcopy ruler for calibrated image quality assessment. *Journal of Electronic Imaging*, *19*(1), 011009-011009, 2010.

[14] P1858™/D2 Draft Standard for Camera Phone Image Quality, v24, 2016.

[15] Torgerson, Warren S., Theory and methods of scaling, J. Wiley & Sons, 1958.

[16] Engeldrum, P. G., Psychometric Scaling: A Toolkit for Imaging Systems, Imcotek Press, Massachusetts, 2000.