Softcopy quality ruler method: Implementation and validation

Elaine W. Jin*[a], Brian W. Keelan[a], Junqing Chen[a], Jonathan B. Phillips[b], Ying Chen[c]
[a]Aptina Imaging, LLC, 3080 North First Street, San Jose, CA, USA 95134;
[b]Eastman Kodak Company, 343 State Street, Rochester, NY, USA 14650;
[c]Vista Point Technology, 2241 Lundy Ave, San Jose, CA, USA 95131

ABSTRACT

A softcopy quality ruler method was implemented for the International Imaging Industry Association (I3A) Camera Phone Image Quality (CPIQ) Initiative. This work extends ISO 20462 Part 3 by virtue of creating reference digital images of known subjective image quality, complimenting the hardcopy Standard Reference Stimuli (SRS). The softcopy ruler method was developed using images from a Canon EOS 1Ds Mark II D-SLR digital still camera (DSC) and a point-and-shoot Kodak EasyShare P880 zoom digital camera. Images were viewed on an Apple Cinema HD Display (30" flat panel) at a viewing distance of 34 inches. Ruler images were made for 16 scenes. Thirty ruler images were generated for each scene, representing ISO 20462 Standard Quality Scale (SQS) values of approximately 2 to 31 at an increment of one just noticeable difference (JND) by adjusting the system modulation transfer function (MTF). A Matlab GUI was developed to display the ruler and test images side-by-side with a user-adjustable ruler level controlled by a slider. Two validation studies were performed at Eastman Kodak Company, Vista Point Technology, and Aptina Imaging in which all three companies set up a similar viewing lab to run the softcopy ruler method. The results show that the three sets of data are in reasonable agreement with each other, with the differences within the range expected from observer variability. Compared to previous implementations of the quality ruler, the slider-based user interface allows approximately 2x faster assessments with 20% better precision.

Keywords: Softcopy quality ruler, standard reference stimuli, SRS, standard quality scale, SQS, just noticeable difference, JND, ISO 20462

1. INTRODUCTION

The International Imaging Industry Association (I3A) started the Camera Phone Image Quality (CPIQ) initiative in June 2006 [1]. The goal of this initiative was to develop an Image Quality Testing and Performance Rating System that could enable manufacturers and carriers to make comparisons between capture devices, and to communicate a substantiated product quality rating to consumers. CPIQ Phase 2 spanned the period of July 2007 – December 2008. The specific goals for CPIQ Phase 2 were to define objective and subjective test methods for measuring camera phone image quality attributes, and to provide specific tools and validated test methods to facilitate standard-based communication and comparison regarding camera phone image quality. The development of objective metrics called for the implementation of a robust and fast psychophysical method. The softcopy quality ruler method was implemented to meet this need.

The softcopy quality ruler described herein is based on ISO 20462 Part 3 [2-3]. This standard describes a real-time technique optimized for obtaining assessments over a wide range of image quality. The work reported in this paper extends the standard by virtue of creating reference digital images of known subjective image quality, complimenting the hardcopy Standard Reference Stimuli (SRS). When ISO 20462 was first developed, the quality ruler images were made using silver halide prints. This is because at the time most displays used cathode-ray tube technology, in which focus drifted over time and modulation transfer function (MTF) varied from center to corner of the display. With newer flat panel display technologies, MTF is quite consistent over time and spatial position, allowing digital images to be displayed in a well-defined manner. This capability enables the production of calibrated digital standard images for use in the softcopy quality ruler. Compared to the hardcopy ruler method, a softcopy ruler method has several advantages. First, ruler images can be made for many new scenes, making it possible that the test scenes match the ruler scenes. Second, data collection is easier with the softcopy ruler method. Third, the digital reference images can be easily distributed and used in a software application.

*ejin@aptina.com; phone 1 408 660-2581; fax 1 408 660-2343; aptina.com

Image Quality and System Performance VI, edited by Susan P. Farnand, Frans Gaykema,
Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 7242, 724206 · © 2009 SPIE-IS&T
CCC code: 0277-786X/09/$18 · doi: 10.1117/12.805922

In this paper we report the development of a practical softcopy quality ruler, including: a set of ruler images developed to meet the requirements set by ISO 20462; a software tool displaying the ruler and test images and collecting observer responses; and a specified viewing environment. We will also report the results of two validation studies performed by three CPIQ member companies, Eastman Kodak Company, Vista Point Technology, and Aptina Imaging, LLC.

2. ISO 20462 AND THE QUALITY RULER METHOD

ISO 20462 is a three-part standard entitled "Psychophysical experimental methods to estimate image quality" [2-3]. Part 3 of this standard, "Quality ruler method", describes a psychophysical method that involves quality or attribute assessment of a test stimulus against a series of ordered, univariate reference stimuli that differ by known numbers of just noticeable differences (JNDs). The standard defines a JND to be a stimulus difference that yields a 75%:25% proportion in a forced-choice paired comparison. Furthermore, the standard distinguishes two types of JND units, attribute and quality JNDs [4(a)]. An attribute JND is a measure of detectability of appearance differences. In contrast, a quality JND is a measure of the significance or importance of perceived differences on overall image quality. Assessment of quality involves not only detection of appearance differences but also value judgment and personal preference. The quality ruler method provides a robust and convenient way to measure images in the quality JND units.

A quality ruler is a series of reference images depicting the same scene, but varying widely in a single attribute, with known quality differences (in JNDs) between the samples. Because the JND values of the reference stimuli are known, identification of the point of equality on the ruler allows a numerical value to be associated with the test stimulus immediately upon assessment by the observer. In ISO 20462 Part 3, the ruler images are generated by varying sharpness. Sharpness is a good reference attribute because it: (1) is readily varied by image processing; (2) is correlated with MTF, which can be quantified by measurements from standard targets; (3) exhibits relatively low variability between different observers and scenes; and (4) has a strong affect on image quality in many practical imaging systems. The standard also defines a numerical scale and a set of reference images to be used in the quality ruler method. The standard quality scale (SQS) is a fixed numerical scale of quality having the following properties: (1) it is anchored against physical standards; (2) it has one unit corresponding to one JND; and (3) it has a zero point corresponding to an image having little identifiable information content. The Standard Reference Stimuli (SRS) are a set of reflection prints used in the hardcopy quality ruler, which vary in sharpness and are calibrated against SQS. Three sets of SRS prints were made available through I3A when the standard was published.

ISO 20462 Part 3 provides a method for users to generate their own ruler images with known relative JND values, and if desired, calibrate them absolutely against the SRS. This method includes four steps. First, an aim system MTF is defined to mimic the response of a monochromatic, on-axis, diffraction-limited lens (Eq.1). In Eq.1 *k* is regarded as a reciprocal measure of the system bandwidth. Second, a spatial kernel is created to modify the real system MTF (including capture, image processing, and display), to approximate the aim system MTF of Eq. 1. Third, if the shape of the real system MTF conforms sufficiently to that of Eq. 1 after spatial filtering, the value of *k* fitting the aim system MTF is used to compute the relative JNDs of quality (Eq. 2). This JND value represents the SQS value of an image having average scene content, excellent color and tone reproduction, and no evident artifacts or noise. Fourth, a ruler image is created by applying the spatial kernel to an image captured by the specific camera. A series of ruler images can be created for a series of *k* values (and hence SQS values).

$$\begin{aligned} m(v) &= \frac{2}{\pi}\left(\cos^{-1}(kv) - kv\sqrt{1-(kv)^2}\right) & kv \le 1 \\ m(v) &= 0 & kv > 1 \end{aligned} \qquad (1)$$

where:

m is the modulation transfer of the complete imaging system ν is spatial frequency in cycles per degree (CPD) at the eye of the observer;

k is a reciprocal measure of the system bandwidth in CPD.

$$JNDs = \frac{17249 + 203792k - 114950k^2 - 3571075k^3}{578 - 1304k + 357372k^2} \qquad (1 \le 100k \le 26) \qquad (2)$$

The SQS and SRS defined by the standard represent the first calibrated physical standards and numerical scale of absolute pictorial image quality. They allow the direct comparison of experimental results from different labs performed at different times. For reference, the widths of the categories “excellent”, “very good”, “good”, “fair”, “poor”, and “not worth keeping” average approximately 6 JNDs [4(b)]. A scale value of 31 falls within the “excellent” region while a scale value of 2 falls within the “not worth keeping” region as assigned by consumer photographers.

It should be understood that this section is a very brief summary of parts of ISO 20462 Part 3. Information from Refs. [2-3] is utilized extensively throughout the text even though we may not point to them with individual citations.

3. CREATION OF THE RULER IMAGES

As indicated in the section above, the measurement of the system MTF is an important step in creating the quality ruler images. The system MTF is a cascade of the capture, image processing, and display MTFs.

3.1 Camera image processing pipeline

Two digital still cameras (DSCs) were used in the study. One is a Canon EOS 1Ds Mark II D-SLR camera, which has a resolution of 16 MP (3328 x 4992). The other is a point-and-shoot Kodak EasyShare P880 zoom digital camera, which has a resolution of 8 MP (2448 x 3264). The two cameras were chosen because both provide high quality images and exhibit consistent MTFs regardless of field height and other capture conditions.

The Canon EOS 1Ds Mark II camera (a.k.a. the Canon camera in the following text) generates both a JPEG image and a raw image. For this study only the raw images were used, and a simple image processing pipeline was developed to post-process these images. We chose to use an external pipeline instead of using the camera pipeline for a number of reasons: (1) The Canon pipeline may use edge-aware adaptive algorithms for demosaicing, denoising, sharpening, and other spatial operations. These edge-aware adaptive algorithms can selectively process edges vs. non-edge regions for improved image quality. For our purpose, however, the adaptive algorithms present a challenge to the measurement of the system MTF and hence should be avoided. (2) An external pipeline provides the flexibility we may need in the future work. For example, when creating the ruler images the spatial filtering should be performed in a linear space rather than the gamma corrected space. By controlling the pipeline we can intercept the processed image at different stages of the imaging chain and perform additional processing. (3) By controlling the pipeline we can use different color and tone positions for different scenes so that each scene can be individually rendered in a pleasing fashion.

The pipeline for Canon raw images consists of seven steps that process images between five stages (see Fig. 1). The input to the pipeline is a full-resolution Bayer image captured by the Canon EOS 1Ds Mark II camera. A bilinear demosaicing algorithm transforms the Bayer image to a linear camera RGB space. White balance and exposure are applied manually to individual scenes to achieve a pleasing look. A color correction matrix (CCM) converts the image from camera RGB to a linear sRGB space. A tone mapping curve is applied to render a gamma-corrected image ready for display. Step 5 in the imaging chain, the downsampling step, will be discussed later in Section 3.4. Step 6 in the imaging chain, the spatial filtering step, is related to creating individual ruler images, as described in Section 2. We found that after the individual adjustment to exposure and white balance, a fixed CCM and a single tone curve can render the selected set of Canon images satisfactorily.

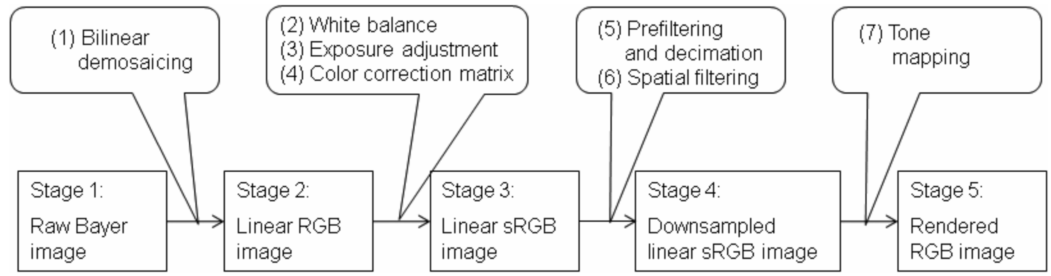

Fig. 1. Image processing pipeline for Canon raw images, which consists of seven image processing steps that process images between five stages.

The point-and-shoot Kodak EasyShare P880 zoom digital camera (a.k.a. the Kodak camera in the following text) only generates JPEG images. These images were mapped back to linear sRGB space through the inverse of the sRGB standard tone scale, and then Steps (5) – (7) from Fig. 1 were applied.

3.2 Camera MTF measurement

The ISO12233 resolution test chart and the Imatest software package [5] were used in the camera MTF measurement. The test chart was uniformly illuminated by cool white fluorescent lights from two SpectralLight III light boxes. The capture conditions of the test chart were similar to those used to capture the reference images. For the Canon camera, the conditions were: a fixed focus 50mm lens; ISO speed at 125; f-number at f/11; and white balance setting at "fluorescent". Auto-exposure was used during the capture, allowing the exposure time to change with the light level. Four captures were taken on the same chart. Each time the shutter button was pressed halfway to enable auto focus before a capture was taken. The output images were in Canon raw format, and they were subsequently processed using the pipeline described above. The input to Imatest was a 16-bit linear sRGB image (Stage 3 in Fig. 1).

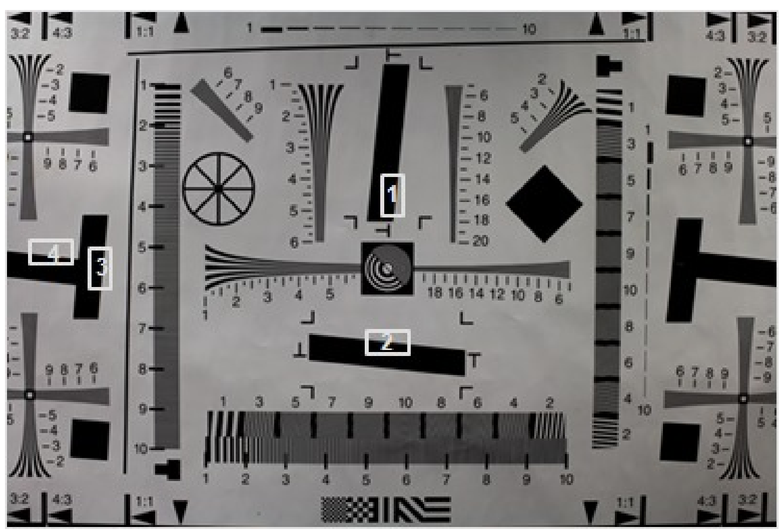

Fig. 2. Test chart used in camera MTF calculations. The white rectangles marked 1-4 show the regions used in the Imatest MTF calculations.

Fig. 2 shows one captured image of the ISO12233 test chart with the regions used in the MTF calculation marked. The four regions allow calculation of horizontal and vertical edge MTFs fairly close to the optical axis and center of the image, and also in the periphery of the frame. In Imatest the four regions were selected manually, and calculation was performed separately for each region. Pixel pitch was specified to be 7.2 μm for the Canon camera, and because the input images were in linear sRGB space, gamma was set to 1 in Imatest.

A comparison of the MTF results calculated from the four captured images showed little change between captures. Therefore, the four MTFs from individual captures were averaged, and the resulting MTFs are shown in Fig. 3. The horizontal axis stands for the spatial frequency in units of cycle/mm in the sensor plane, and the vertical axis is the fractional modulation transfer. It can be observed that there is little variation in MTF with regard to edge orientation or field location. Therefore, the mean MTF averaged over the four curves was used to represent the final MTF of the Canon camera. The smooth fall-off from zero frequency confirms that there is no sharpening in the imaging pipeline.

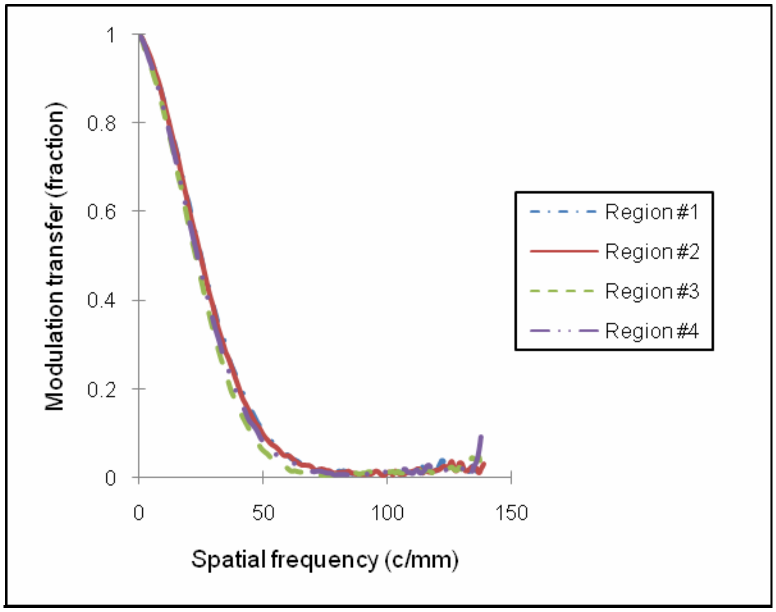

Fig. 3. The MTF measurements of the Canon EOD 1Ds Mark II D-SLR camera. The MTF variations among the four regions (shown in Fig. 2) are minimal.

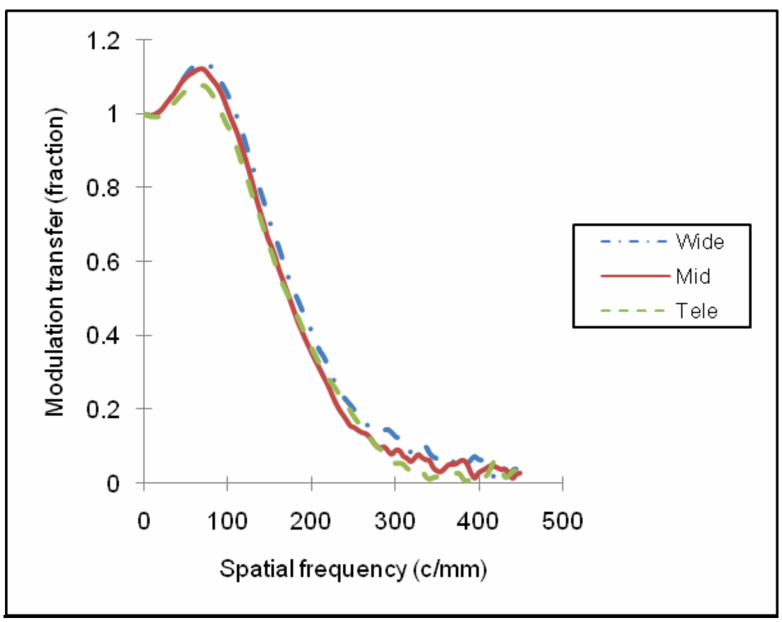

Fig. 4. MTFs for three different zoom positions (‘Wide’, ‘Mid’, and ‘Tele’) at the near center, horizontal edge for the point-and-shoot Kodak EasyShare P880 zoom digital camera. The variations across the zoom positions are fairly minor.

Different zoom positions were used while capturing the reference scenes with the Kodak camera. Because the camera MTF might change for different zoom positions, the ISO12233 test chart was captured at three zoom positions labeled "tele" (17.7 mm focal length), "mid" (12.5 mm), and "wide" (7.9 mm). A captured JPEG image of the ISO12233 test chart was used as input to Imatest. A gamma value of 0.45 was used in Imatest to reflect the nature of the gamma correction applied. The pixel pitch of the Kodak camera is 2.225 µm. Fig. 4 shows the MTFs for the near center, horizontal edge at three zoom positions. As can be seen from the figure, the MTF of the camera improves slightly as the focal length decreases. Because the MTF differences are small, the MTFs for the three zoom positions were averaged, and the results are shown in Fig. 5. Again, the differences between field positions and orientations are not too large, so an average MTF over all three zoom positions and all four regions was used to represent the final MTF for the Kodak camera. It is obvious that the Kodak camera has a significant sharpening operation applied in the imaging pipeline, reflected in the boost of the MTF curves over the frequency range of 50-100 c/mm. Although the resolving power of the Kodak camera is much higher, in part because of its smaller pixels, and our use of a bilinear demosaicing algorithm in the Canon camera pipeline, the difference is mostly compensated by the roughly 22x larger format area of the Canon camera, in terms of the final system MTF at the retina, as shown later.

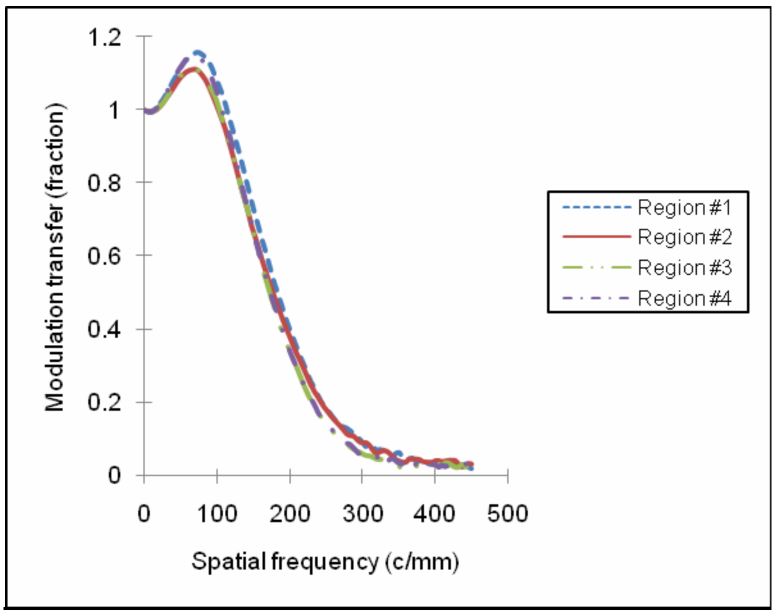

Fig. 5. MTFs for the four regions (shown in Fig. 2) averaged across the three zoom positions for the Kodak camera. The four MTFs are in general quite similar. The bump in the MTF curves is a result of digital sharpening.

3.3 Display MTF measurement

A large panel with sufficiently high pixel count is needed for implementing the softcopy ruler. This is because the ruler method requires that a ruler image and a test image be presented side by side for comparison on the same display. Using two displays is a more difficult option because it can be challenging to match the two displays exactly in color and tone. In this study an Apple Cinema HD Display (30" flat panel, a.k.a. the Apple display in the following text) was used for displaying the softcopy quality ruler. This display uses a thin-film transistor active-matrix liquid-crystal display technology. It has a resolution of 4 MP (1600 x 2560), a pixel pitch of 0.250 mm, a peak brightness of 400 cd/m^2, and a contrast ratio of 700:1.

A calibrated camera with a known MTF was needed to measure the display MTF. In this study we used an Aptina demo camera with a Navitar 614 lens, an Aptina monochrome sensor, and an internal software tool to drive the camera. The monochrome sensor has a pixel pitch of 2.2 µm and a sensor array size of 1944 x 2592. The raw linear image was analyzed in Imatest using a gamma of unity. The results of the measurement are shown in Fig. 6. For this demo camera the vertical edge MTF is very slightly better than the horizontal edge MTF.

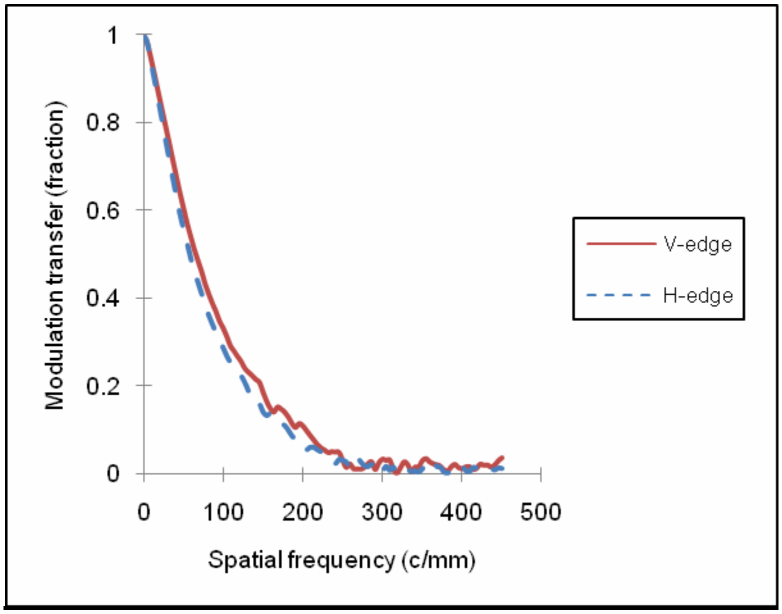

Fig. 6. Horizontal and vertical edge MTFs of the Aptina monochrome demo camera.

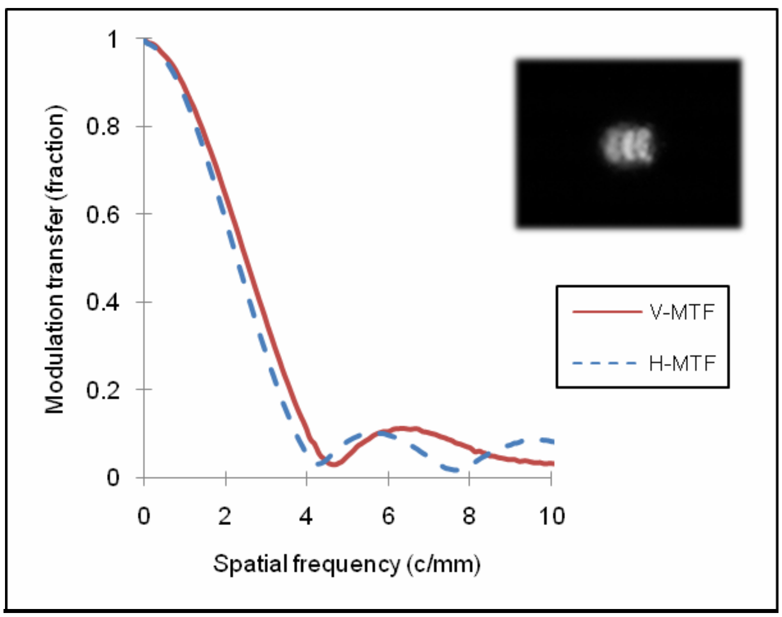

Fig. 7. MTFs of the Apple Cinema HD Display (30” flat panel). The vertical MTF is slightly better than the horizontal MTF for this display. The inset shows the captured image of a single white pixel in an otherwise dark surround.

The display MTF was measured using a captured point source image. A single display pixel was lit at [255, 255, 255] in the center of the display with a black background. The calibrated camera was used to capture a macro image of the bright pixel (see Fig. 7 inset); the magnification was such that about 870 sensor pixels imaged the single display pixel, giving excellent resolution of structure. White and black uniform flat fields were also captured to calibrate the capture.

Fourier transformation was performed on the point spread function to obtain horizontal and vertical MTFs. Finally, the display MTFs were derived by taking the ratio of the measured MTF to the camera MTF. Fig. 7 shows the display MTFs obtained from the experiment. Here the horizontal axis is the spatial frequency in units of cycle/mm at the display surface. A single display MTF was derived by averaging the horizontal and vertical MTFs.

3.4 MTF associated with image downsampling

In order to display both the ruler and the test images side by side on the screen, the displayed image size can only be around 1 MP (i.e. long side < 2560/2 = 1280). This indicates the necessity to downsample the images captured by both cameras (as shown in Fig. 1, Step 5). For the 16MP Canon images, downsampling consists of two repetitions of prefiltering and 2x decimation along both the x and y axes. The prefiltering step was performed to prevent aliasing artifacts that might be produced when an image is decimated. The filter used in this step had a kernel size of 11 x 11. For the 8 MP Kodak images downsampling was performed by prefiltering and 3x decimation along both the x and y axes. A 17 x 17 kernel was used as the prefilter for the Kodak images. After decimation the displayed image sizes became 834 x 1253 for the Canon images, and 816 x 1088 for the Kodak images. Because the prefiltering affects the spatial characteristics of the ruler images, it was included in the calculation of the system MTF. Equation 3 shows the cascade of camera MTF, downsampling MTF, and display MTF to obtain the system MTF. The system MTF is defined in units of cycle per degree (CPD) at the eye of the observer, and it is associated with a specific viewing distance. In the current study the viewing distance is set at 34 inches, close to the shortest distance at which the two images can be compared just by changing direction of glance, without having to shift head position. Fig. 8 shows the MTFs of the two systems used in the current study: the Canon EOS 1Ds Mark II D-SLR camera twice downsampled 2x and shown on the Apple Cinema HD Display (30" flat panel); and the point-and-shoot Kodak EasyShare P880 zoom digital camera downsampled 3x and displayed similarly.

$$\text{System MTF} = \text{camera MTF} * \text{prefilter MTF(s)} * \text{display MTF} \quad (3)$$

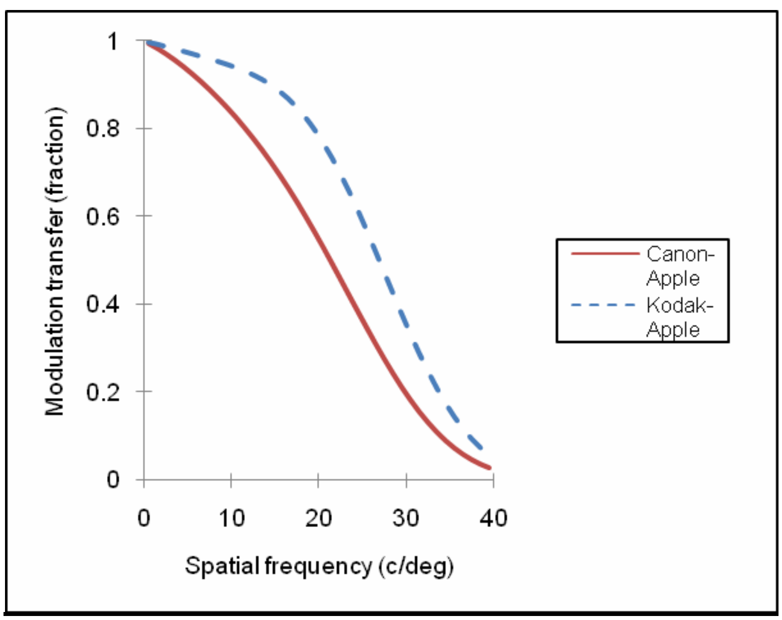

Fig. 8. The MTFs of the two systems used in the current study: the Canon EOS 1Ds Mark II D-SLR camera twice downsampled 2x and shown on the Apple Cinema HD Display (30" flat panel); and the point-and-shoot Kodak EasyShare P880 zoom digital camera downsampled 3x and displayed similarly.

3.5 Creation of the ruler images

As described in Section 2, a set of blur kernels can be created to shape the system MTF to achieve specific JND values. Two different cameras were used in the study, resulting in two different system MTFs, as shown in Fig. 8. Consequently, two different sets of blur kernels were created for the Canon and the Kodak images. Thirty Ruler images were created for each scene by convolving the blur kernels with the downsampled images (as shown in Fig. 1, Step 6).

Fig. 9. Sixteen scenes for which ruler images were created. Seven scenes were provided by Aptina Imaging, captured using a Canon EOS 1Ds Mark II D-SLR camera. Nine scenes were provided by Eastman Kodak Company, captured using a point-and-shoot Kodak EasyShare P880 zoom digital camera.

ISO 20462 Part 3 defines a conversion (reproduced as Eq. 2 in this paper) between system reciprocal bandwidth and relative JND values. Although not stated in the standard, the relative JND values from Eq. 2 exactly equal the SQS values of an average scene if the following requirements are met: (1) imaging artifacts and noise are negligible at the calibrated viewing distance; and (2) color and tone rendition are excellent. We deemed these conditions to be met and so used Eq. 2 to compute SQS values for the ruler images that should be accurate on average, but may err slightly for individual scenes because of scene-specific compositional, lighting, and other factors.

The ruler images for each scene span the SQS range from 2.0 to 31.0 in 1 JND increments, when viewed at the design distance of 34 inches for a 100 pixel per inch display (for other flat panel resolutions, the viewing distance should be 3400x the pixel pitch). We believe that this range is large enough to cover nearly all practical uses in the imaging industry. For reference, the hardcopy Standard Reference Stimuli, released with the publication of ISO 20462 in 2005, span an SQS range of 14 to 32 in 3 JND increments.

The current set of ruler scenes is made of seven Canon camera captures provided by Aptina Imaging, and nine Kodak camera captures provided by Eastman Kodak Company. All sixteen scenes were taken by professional photographers, and they depict a variety of subject matter including people and landscape scenes, indoor and outdoor scenes, and they are rich in memory colors such as skin tone, foliage, and sky (Fig. 9).

4. GRAPHICAL USER INTERFACE AND THE VIEWING ENVIRONMENT

A Matlab tool was developed to deploy the softcopy quality ruler. The graphics user interface (GUI) for the softcopy ruler method is shown in Fig. 10. At a pixel count of 1600 x 2560, the Apple display can handle two images side by side in either landscape or portrait orientation, and for both the Canon camera (834 x 1253) and the Kodak camera (816 x 1088). In the interface two images are presented on the display. The image on the left is the ruler image and the image on the right is the test image. The quality of the ruler image can be varied by moving the bottom slider bar from left (very sharp) to right (blurred). The test images may have different levels of image degradation. The observer is asked to match the overall quality of the ruler image to that of the test image.

Fig. 10. Graphics user interface of the softcopy quality ruler. The ruler image is on the left and the test image is on the right. The user can adjust the quality and sharpness level of the ruler image by moving the slider bar below.

This Matlab GUI enables a self-administered study. In the training session a number of training images can be displayed. The subject practices using the slider bar to vary the sharpness and quality of the ruler image displayed. The subject also is familiarized with the image attribute being studied. Each image in the training session is displayed together with some comments regarding the traits exemplified by the image. The test session starts after the subject completes the training session. In the test session all levels of one scene are displayed in one block, so that the observer can establish consistent criteria and judge efficiently. The order of levels within the scene is randomized. For different observers, the order of scenes is also randomized. In addition, the initial ruler position is randomized for each test image. After the subject makes a match, the final ruler position is recorded. A calibration file is referenced to translate this ruler position to an SQS value. The output data file records the SQS values, together with the time duration used for each trial. Each session of the ruler study typically takes one hour or less.

To use the softcopy ruler, observers should have normal visual acuity at the controlled viewing distance (3400x the display pixel pitch). Observers with bifocals or progressive lenses may have difficulty simultaneously seeing the entirety of the two images sharply. Researchers with corrected vision, who expect to use this technique frequently, might consider obtaining a single-vision prescription for the ruler viewing distance.

The viewing environment includes a high-quality LCD monitor, a surround lighting setup, and a device to control the viewing distance, in this work a headrest made at the University of Houston, College of Optometry [6]. The viewing room was painted neutral gray. Fig. 11(a) shows a side view of the lab setup. Fig. 11(b) shows a front view from the observer's eye position. The wall behind the display was moderately uniformly illuminated by D65 fluorescent tubes so that the luminance was similar to that of an average pictorial scene rendered on the display. The illuminance from the fluorescent lights was controlled using a combination of variable density tube guards [7] and a variable power supply; the latter affected color temperature as well. Because there is no light source directly illuminating the monitor, the dynamic range of the displayed image is high. The extended gray surround helps to reduce the fatigue of observers compared to the case of dark-surround viewing.

(a) (b)

Fig. 11. Viewing room setup for the softcopy quality ruler method: (a) side view of the set up; (b) view of the display from the observer's eye position.

5. RULER VALIDATION

After the softcopy quality ruler was developed, a validation study was performed by three CPIQ member companies: Eastman Kodak Company, Vista Point Technology, and Aptina Imaging, LLC. One goal of the validation study was to examine whether the ruler method could be used reliably in different industrial labs if the display and the viewing distance were specified. Although the test images and the software would be the same, there still could be many differences among the labs. For example, the observers could be different in terms of experience in image processing or in subjective evaluation. A second goal was to test how well the slider-based GUI worked for controlling the quality ruler image displayed, compared to the approach described in ISO 20462 Part 3, involving repeated paired comparisons selected by a binary sort routine. In particular, would assessment time be reduced from the 30 seconds reported for the latter method?

All three companies acquired the same Apple Cinema HD Display (30” flat panel) and the same headrest from University of Houston School of Optometry. All three labs were set up to view the display in a similar viewing environment. The method of illuminating the viewing room was slightly different between the labs, including the light source and the geometry of the illumination. Seventeen observers participated in the validation study, including five from Company 1, four from Company 2, and eight from Company 3.

The validation study used ten of the CPIQ scenes (see Fig. 9: numbers 3 and 4 from the left in Row 1, 1 to 3 in Row 2, 1, 3 and 4 in Row 3, and 2 and 4 in Row 4). These scenes included texture elements such as faces, landscapes, architecture and foliage. Eight levels of noise cleaning were applied to those scenes, resulting in different levels of image quality. A companion paper [8] will provide further details on the test image manipulation and the experimental procedures. It should be noted that the validation study used a set of ruler images created for a viewing distance of 43in instead of 34in. The rational for the recent change in viewing distance from 43in to 34in is documented in the Appendix.

Fig. 12 shows the SQS values obtained from the validation study. The results from three companies are shown separately. Each of the dashed or dotted lines shows the results averaged across all test scenes and all subjects at each site. The solid line shows the mean results over data from all three sites. The error bars represent the standard error of means. The results show that the maximum deviation of the individual result from the mean result is 2.3 JND, and the mean deviation is less than 1 JND.

The duration of time spent on each trial was also tracked. At Company 1 the median trial time was 14.3 seconds; at Company 2, 15.4 seconds; and at Company 3, 16.5 seconds. According to Ref. 2, a typical trial in earlier quality ruler implementations would take 30 seconds. This indicates that the slider-based softcopy ruler method can effectively reduce the experimental time by half, an exciting result.

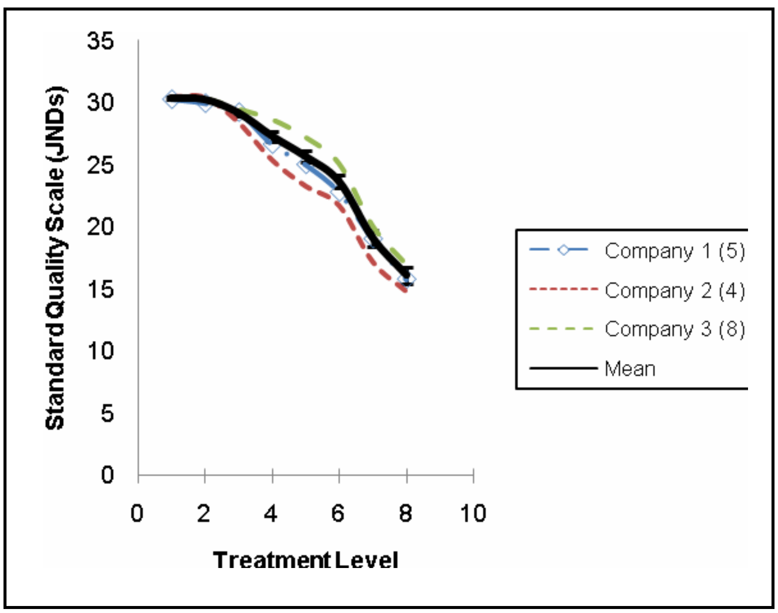

Fig. 12. Results of the validation study. The numbers in the parenthesis are the number of observers participated in each company. The results show that there is a reasonable agreement among the results from the three companies.

In a separate validation study two ruler images per scene were added to the test image set to test the precision and accuracy of the slider bar ruler method. One ruler image was the highest quality level (SRS = 31.06), the purpose of which was to test for the small bias observed with other implementations of the quality ruler. The second ruler image had a SQS value of 17.05, in the middle of the quality range of the ruler. The images were assessed by five observers for six scenes. The mean response to the SRS = 31 level was 30.21, consistent with previous reports of a slight bias caused by observers' slight counter-inclination to gave ratings higher than the best ruler position. The mean response to the SQS = 17 level was 17.08, in close agreement with the true ruler value. The standard deviation of ratings of this level was 1.96 JNDs, which is better than the 2.5 JND value reported for earlier implementations of the quality ruler. The combination of the reduced time per assessment and improved precision suggest that the slider interface allows a fundamentally superior comparison to be made by the observers.

6. CONCLUSIONS

This study describes the creation of a softcopy quality ruler for calibrated image quality studies. Compared to previous implementations of the quality ruler, as described in ISO 20462, Part 3, the slider-based user interface allows approximately 2x faster assessments with 20% better precision. A set of images at 30 levels of quality, spaced by 1 JND increment at a specified viewing distance, will be made available through I3A in connection with ISO 20462. At least 16 different scenes will be represented. A Matlab graphical user interface program for ruler display and data collection will be included with the images.

7. APPENDIX: UPDATES PROPOSED FOR ISO 20462 PART 3

In ISO 20462 Part 3, a method was described to verify whether the shape of a system MTF conforms sufficiently closely to the shape defined by Equation (1). The standard states that the system MTF shall be considered to be within conformance and valid for use if the mean fractional modulation transfer of the system and aim MTFs over each of the frequency bands 0 to 5, 5 to 10, … and 35 to 40 cycles per degree (CPD) agree to within 0.05. Because of the low resolution of the softcopy display (typically 100 pixels per inch) compared to the hardcopy prints (250 pixels per inch or higher), this requirement can be difficult to meet in the highest frequency bands, at the highest MTF levels, for shorter viewing distances. In the present case, a minimum viewing distance of about 43 inches was required to meet the criteria at the highest SQS positions. At this viewing distance, the subtense of the test and ruler images is somewhat smaller than might be desirable. Relaxing the maximum frequency at which the system MTF must be matched, from 40 CPD to 30 CPD permits a viewing distance of about 34 inches, which is about as close as the side-by-side images can be comfortably viewed just by changing the angle of glance, without shifting head position. The potential inaccuracy in prediction of quality with the 30 CPD maximum is quite small; 30 CPD is the frequency of the "E" symbols in the 20/20 line of a Snellen eye chart and so is close to the resolving power of the human visual system with well-corrected vision.

In ISO 20462 Part 3, SQS values can only be reported if the rulers have been directly calibrated against the hardcopy Standard Reference Stimuli. The softcopy rulers described in this work, when displayed on suitable equipment, should provide stimuli for which the SQS values are quite accurately known, but without additional hardcopy-softcopy experimentation, they will not meet the criterion of the standard. Recognizing the utility of these slightly less rigorous SQS values, they are being provided with the ruler images, even though they do not quite meet the current standard.

One further limitation of the current standard is that alternative approaches to the user interface in softcopy quality rulers were not anticipated. For example, the real-time ruler image update via a slider, as employed here, is not specifically acknowledged; only a binary sort via paired comparison was described in detail in the standard.

Given the considerations above, one of the authors (BWK) has submitted a formal proposal for minor revision of ISO 20462 Part 3, to support a wider range of softcopy ruler approaches; to relax the frequency requirement, and to define a second level of accuracy for which SQS values can be reported. If this minor revision is successful, the softcopy ruler approach described herein will be fully compliant with the revised standard.

REFERENCES

[1] CPIQ Initiative Group, "Camera Phone Image quality – Phase 1 – Fundamentals and review of considered test methods," I3A CPIQ Phase 1 document repository (2007). http://www.i3a.org/wp-content/uploads/2008/03/cpiq_white_paper.pdf.

[2] ISO 20462-3, "Photography – Psychophysical experimental methods for estimating image quality – Part 3: Quality ruler method", Technical Committee 42, International Organization for Standardization, Geneva (2005).

[3] Keelan, B. and Urabe, H. "ISO 20462, A psychophysical image quality measurement standard," Proc. SPIE-IS&T Electronic Imaging 5294, 181-189 (2004).

[4] Keelan, B. W., [Handbook of Image Quality: Characterization and Prediction], Marcel Dekker, Inc., New York. (a) 35-37; (b) 81-83 (2002).

[5] Imatest software application, http://www.imatest.com

[6] Headrest, http://www.opt.uh.edu/uhcotech/headspot/index.cfm

[7] General part 76100, F32T8 4' MATRIX TUBE GUARD (http://www.servicelighting.com/catalog_product.cfm?source=Froogle.com&prod=LE76100).

[8] Phillips, J., Jin, E. W., Chen, Y., and Clark, J. "Correlating objective and subjective evaluation of texture appearance with applications to cell phone imaging," Proc. SPIE-IS&T Electronic Imaging (to be published in 2009).