VIME and Subjective Image Quality Tests

Michele A. Saad and Philip Corriveau

Work is underway within the VIME (Video/Image Models for consumer content Evaluation) workgroup at VQEG to develop tools for no-reference image quality evaluation that cater to consumers' needs. This means that the tools developed are meant to predict quality for images that are representative of real consumer scenarios. Subjective image quality research is at the forefront of this work and serves a number of purposes: 1) to understand the psychophysical response of subjects to images of disparate quality; 2) to understand the dependence of this response on the subjective study design (in other words, to understand how different study designs elicit different responses from subjects); 3) to generate data for use in objective no-reference image quality evaluation model development and testing.

> "A subjective study design is described where no images included in the test contain simulated distortions. … The images in the test were scenes captured by a multitude of devices and the quality range in the test is entirely determined by the image quality delivered by the devices used to capture the scenes."

The CCRIQ Database

The recent work in [1] describes a subjective study design that is in line with the consumer-oriented image quality evaluation objective. The outcome of the study is an analysis of a new subjective design approach and a database of images with associated subjective scores which has been dubbed the Consumer Content Resolution and Image Quality (CCRIQ) database.

A subjective study design is described where no images included in the test contained simulated distortions. The images in the test however, were scenes captured by a multitude of devices and the quality range in the test was entirely determined by the image quality delivered by the devices used to capture the scenes. The devices used to capture the scenes included four device categories: two tablets, eleven phones, six compact cameras, and four digital single-lens reflex (DSLR) cameras. The devices were chosen to provide images ranging in resolution from 1 megapixel (MP) to 20 MPs. The devices were also chosen to span a wide range of optics characteristics (lens properties, sensor sizes) and post processing capabilities.

"The significance of an equivalent image set is that no two images in the same set are necessarily exactly the same in content."

Eighteen scenes were captured by each of the 23 devices used in the test. The set of photos pertaining to one image scene was defined as an *equivalent image set*. The significance of an equivalent image set is that no two images in the set are necessarily exactly the same in content, even though they are photo captures of the same scene (i.e., they are the same scene captured by multiple cameras). This is due to inherent differences in the cameras used to capture the photos (such as focal length and aspect ratio) as well as the differences in photo capture angle and fluctuations in the image scene (such as moving clouds) as the photographer changes cameras to reshoot the scene. Fig. 1 shows example scenes from two equivalent image sets. Notice how the scene content within one equivalent image set is not identical to the next. The sample images in Fig.1 also show a variety of camera responses. This approach to building the database of images for subjective testing is in contrast to the more traditional approach of getting a number of high quality reference images and simulating distorted versions of them by introducing artifacts such as blur, noise, or compression and transmission errors. In the more traditional case, the content of each distorted image exactly matches the content of its corresponding reference image.

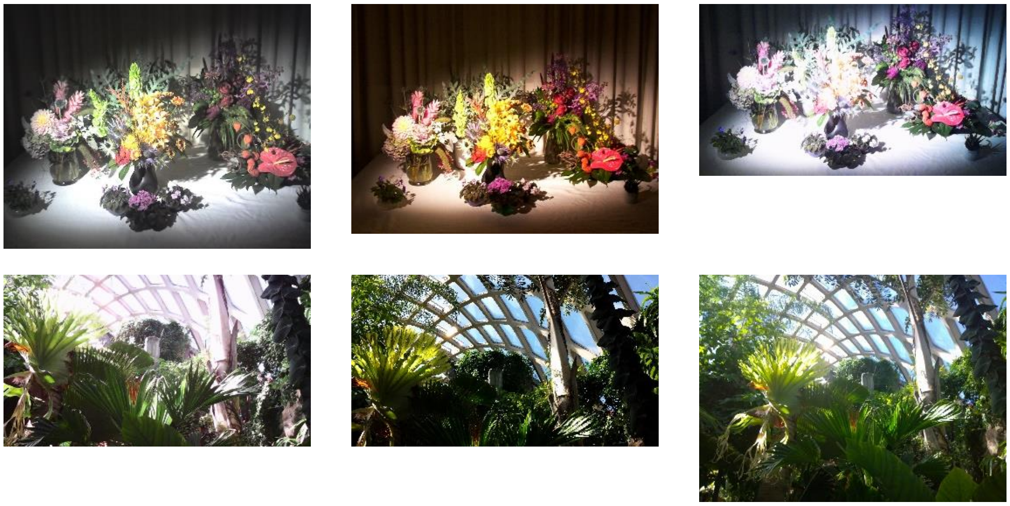

Fig. 1. These sample images show the large variety of camera responses within an equivalent image set. The first row of images beling to one equivalent image set. The second row of images belong to another equivalent image set.

The Subjective Test

The subjective test was performed across three laboratories: 1) NTIA's Institute for Telecommunication Sciences (NTIA/ITS) in Boulder, CO, USA; 2) Ghent University - iMinds, in Ghent, Belgium; and 3) Intel in Santa Clara, CA, USA.

A total of 392 images were rated on two 28'' monitors. One monitor was configured to HD (1920 × 1080) and the other was configured to 4K (3840 × 2160) resolution. The order of image presentation on one monitor or the other was completely randomized. Each scene was rated by either 26 or 27 subjects on each monitor. Table 1 shows the Pearson correlations between the MOS scores obtained from the three labs. The high correlation is an indication of the stability of the experiment design.

	NTIA/ITS	Ghent University	Intel
NTIA/ITS	1	0.952	0.941
Ghent University		1	0.915
Intel			1

Table 1: Pearson correlation between MOS scores of the different labs.

The MOS scores from the HD and 4K monitors were found to be highly linearly correlated. No statistically significant difference was found between the HD and 4K MOS scores on the lower quality images (images that scored a MOS less than 3 on the HD monitor). There was, however, a statistically significant slight difference between the HD and 4K MOS scores on the higher quality images, with the 4K MOSs being on average 0.2 MOS points higher than the HD MOSs.

Several factors impact the final image quality produced by a camera, including the optics and the post processing. Sensor size was found to impact ≈27% to 42% of camera quality on the CCRIQ data set. Among other observations in the study was that the overall quality difference between DSLR cameras and mobile cameras was found to be 0.67 MOS on the CCRIQ data set. Fig. 2 shows the MOS distributions between DSLRs, compact cameras, and mobile devices (phones and tablets). A high degree of overlap in the MOS histograms is observed pointing to the wide range of image quality produced by the different device types.

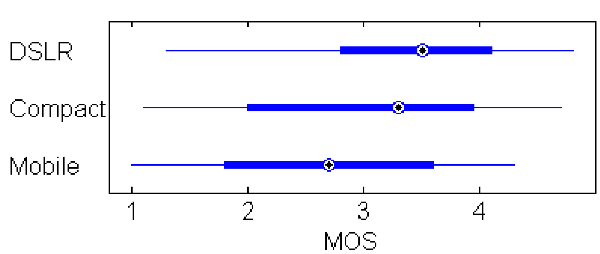

Figure 5: MOS distribution across three device categories.

We refer the reader to [1] for a more complete analysis. The CCRIQ dataset is available to the research community on the Consumer Digital Video Library (CDVL) at www.cdvl.org.

Michele A. Saad *is a Senior Engineer and Researcher in perceptual image and video quality assessment at Intel. She received her Ph.D. degree in electrical and computer engineering from the University of Texas at Austin in 2013. Her research interests include statistical modeling of images and videos, motion perception, design of perceptual image and video quality assessment algorithms, and statistical data analysis and mining and machine learning.*

Philip J. Corriveau *is a Principal Engineer in Experience Development and Assessment in SMG at Intel. He directs a team of human factors engineers conducting user experience research across Intel technologies, platforms and product lines.*

References

[1] M.A. Saad, M.H. Pinson, D.G. Nicholas, N. Van Kets, G. Van Wallendael, R. Da Silva, R.V. Jaladi, and P.J. Corriveau, "Impact of Camera Pixel Count and Monitor Resolution on Perceptual Image Quality", *IEEE Color and Visual Computing Symposium (CVCS)*, August 2015.