IMPACT OF CAMERA PIXEL COUNT AND MONITOR RESOLUTION PERCEPTUAL IMAGE QUALITY

Michele A. Saad[1], Margaret H. Pinson[2], David G. Nicholas[3], Niels Van Kets[4], Glenn Van Wallendael[4], Ralston Da Silva[3], Ramesh V. Jaladi[3], and Philip J. Corriveau[5]

[1] Intel Corp., Austin, TX, USA
[2] NTIA's Institute for Telecommunication Sciences (NTIA/ITS), Boulder, CO, USA
[3] Intel Corp., Santa Clara, CA, USA
[4] Ghent University–iMinds, Ghent, Belgium
[5] Intel Corp., Hillsboro, OR, USA

***Abstract*—Traditional 35mm film cameras are no longer the main devices today's consumers use to capture images. Though the dominant technology has shifted to digital cameras and displays that differ widely in pixel count and resolution, our understanding of the quality impact of these variables lags. This paper explores the quality impact of resolution within this new paradigm. Images were collected from 23 cameras, ranging from a 1 megapixel (MP) mobile phone to a 20 MP digital single-lens reflex camera (DSLR). Subjective ratings from three labs were used to explore the relationship between the camera's pixel count, the display resolution, and the overall perceived quality. This dataset and subjective ratings will be made available on the Consumer Digital Video Library (CDVL, www.cdvl.org) when this paper is published. These images can be used royalty free for research and development purposes.**

Keywords—4K, CCRIQ, camera quality, dataset, image quality, resolution, pixel count, subjective test, subjective quality

I. INTRODUCTION

The first digital image was taken by Steven Sasson in December of 1975, using a device built around a 100 × 100 pixel sensor and recorded on a cassette tape in a process that took 23 seconds. That event revolutionized the way visual information is recorded and exchanged. Breakthroughs in digital photography have changed the way people take, view, and exchange photographs. The first digital camera, which was invented only 40 years ago, was a 0.01 megapixel (MP) camera that weighed eight pounds. Today digital cameras have become ubiquitous and range from high-end digital single-lens reflex cameras (DSLR) to cameras embedded in mobile devices. Along with sleek and light form factors, mobile phones increasingly offer better optical and digital capabilities, such as higher image quality and more sophisticated imaging features.

As digital photography continues to advance and as the field becomes more competitive, we need to understand the effect on perceived picture quality of different camera features and different viewing modes (how a digital photograph is viewed). Not only are the optics improving and the number of digital image pixels increasing, but the digital displays that constitute the most popular image viewing paradigm are evolving and increasing in their resolution capabilities (like UltraHD). As digital camera and display technologies evolve, consumers demand better quality of experience.

In this paper, we explore the joint impact of digital image pixel count and display resolution on perceived image quality. The main goal of this work is to characterize this relationship for a variety of digital cameras. This work has two secondary goals as well. One is to identify the important variables that impact perceptual image quality. The other is to demonstrate a subjective experiment design that can be used to compare cameras.

Towards this end, we created the Consumer Content Resolution and Image Quality Dataset (CCRIQ). This dataset contains images from 23 digital cameras, approximately equally distributed between phones, tablets, point-and-shoots, and higher end DSLRs. The cameras were chosen to sweep the resolution range from 1 to 20 MP. A variety of scenes were captured within each resolution category in the range. A subjective study was conducted to collect quality ratings for each image when displayed on two monitors of equal size but of different resolutions (an HD 1080 monitor and an UltraHD 4K monitor). The details of the study and the analysis of the subjective results are presented in the following sections.

CCRIQ contains impairments inherently introduced by commercially available cameras and no simulated distortions. This is unlike most subjective image quality datasets, which typically contain simulated distortions (such as simulated blur, noise, compression, and transmission artifacts). The types of artifacts present in CCRIQ are hence completely realistic and typical of consumer image-capture scenarios. Furthermore, the range of image qualities is realistic and wholly dictated by the quality delivered by the cameras Unlike previous work, CCRIQ contains a wide range of image resolutions, and a well-balanced distribution among phones, tablets, point-and-shoots, and DSLRs is maintained for each image resolution category.

2015 Colour and Visual Computing Symposium (CVCS)
978-1-4799-1765-5/15/$31.00 ©2015 IEEE

II. BACKGROUND

Image datasets with corresponding subjective quality ratings are extremely useful for quality assessment research and for image perception understanding. There are a number of published and popular datasets in the literature, such as:

- *LIVE Image Quality Assessment Database* [1], which contains JPEG and JPEG2000 compression, Gaussian blur, additive white Gaussian noise, and Rayleigh fast fading channel distortions
- *Tampere Image Database (TID)* [2], which contains a wide variety of simulated distortions including several types of noise, blur, transmission errors, compression errors, local image distortions, and luminance and contrast changes
- *Categorical Subjective Image Quality Database (CSIQ)* [3], which contains compression artifacts, blur, pink noise, and global contrast changes
- *IRCCyN/IVC Image Quality Dataset* [4], which contains compression and blur artifacts
- *MICT Toyama Image Quality Evaluation Database* [5], which contains compression artifacts

Ref. [6] provides a good survey of existing image databases.

We will refer to these as *simulated impairment databases*. The associated experimental design takes a set of source images (usually pristine) and creates multiple impaired versions of each. The impairments are mostly singly occurring simulated distortions. Simulated impairment databases are valuable for understanding of the effect of singly occurring distortions such as blur, noise, or compression artifacts.

Our problem is that simulated impairments are not representative of the type or the level of distortion encountered when people take photos with a variety of consumer devices. Many of the distortion types or levels (the degree of distortion) in the datasets mentioned above do not occur in real consumer scenarios. While it is important to understand the effects of singly occurring distortions, these distortions rarely occur in isolation from other complex image artifacts.

A further limitation of the simulated impairment datasets above is that the resolution of their photos is much lower than state-of-the-art consumer devices. Most of the photos in the LIVE IQA database and those in the MICT database are 768×512 pixels, while those in the CSIQ and the IRCCyN/IVC image databases are 512×512 pixels, and the TID photos are 512×384 pixels. These databases use a pixel count less than 1 MP, while photographs obtained from consumer devices are typically 5 to 20 MP. More details on the limitations of these datasets can be found in [10].

Virtanen et al. [7] start to address the *simulated/artificial distortion* issue in current image quality databases. The authors conducted a study on a dataset of photographs collected from a large number of cameras. This dataset, the *Camera Image Database (CID2013)* is freely available to researchers. CID2013 uses a non-standard rating method, named the dynamic-reference absolute category rating (DR-ACR) study. In addition to collecting subjective ratings for overall quality, subjects were asked to rate four individual image characteristics: sharpness, graininess, lightness, and saturation. The CID2013 contains 480 images from eight scenes. For half of the images in the study, the subjects were instructed to give a score of a 100 for the highest quality photo and a score of 0 for the lowest quality photo within an image set. For the other half of the images, this anchoring of scores was not required of the subjects. CID2013 addresses many of the limitations of prior datasets and is an important step forward in understanding and characterizing real images with subtle and complex realistic distortions.

CID2013 has a number of limitations. The first is related to the scene/content diversity. The images in the dataset are derived from eight scenes, six of which are photographs of people. Most of the scenes are either indoor or outdoor daytime shots; night scenes are lacking. A second limitation is the resolution of the published images. The images in this database have been scaled to 1600×1200 pixels; the original images (with higher resolution) are not available. Third, no information is provided on the cameras. This raises questions about whether this dataset represents typical cameras and photographs and obscures the relationship between device dependent variables and image quality. These relationships play an important role in the development of objective image quality models.

Simulated impairment databases have played an important role in designing objective image quality models, especially for the full reference (FR)[1] quality assessment paradigm. These databases have important limitations for no reference (NR)[2] algorithm design. NR models depend on learning relationships between features and perceived quality. The main challenge is to identify features that generalize to real-world consumer scenarios. If the types of images and associated distortions in the training datasets are not representative of real-world consumer scenarios, then the resulting NR model cannot be expected to generalize. The work in [7]-[9] discusses the generalizability of NR algorithms to consumer usage models.

NR models themselves are critical, because FR models cannot be applied to images captured directly from a camera. A different type of subjective image quality dataset is needed to train NR models. The dataset we present in this paper, CCRIQ, addresses these limitations and lets us explore the relationship between image pixel count, display resolution, and overall image quality.

III. SUBJECTIVE TEST

CCRIQ analyzes the quality of images that are captured by digital cameras and displayed to a digital monitor. The only intermediate processing was scaling the image to the monitor. We focus on two parameters:

- The pixel count of the image produced by a camera.
- The resolution of the display on which the image is viewed.

Our main goal is to characterize the relationship of an image's pixel count and a display's resolution on image quality. This relationship is impacted by numerous factors, such as sensor size, lighting conditions, focal distance, and subject matter. An experiment including all relevant factors would be infeasible due to the large number of stimuli.

Instead, we establish a compromise between controlled variables, uncontrolled variables, and variable limits. To limit the scope of our problem, CCRIQ is constrained to:

---

[1] FR models perform a comparison between the desired image (i.e., the reference) and the delivered image.
[2] NR models examine the delivered image only.

- Images in landscape orientation
- Automatic camera settings
- Maximum pixel count for each camera
- Full screen display of images
- A subset of all image content
- Large, high resolution displays
- No tripods

While tripods are typically used for camera evaluations (e.g., DxOMark©), CCRIQ matches prevalent consumer behaviors for the mobile phone cameras that dominate today's market. This implies handheld cameras, automatic settings, and a mixture of artificial and natural lighting.

This section summarizes CCRIQ. See [11] for the full subjective test plan.

A. Experiment Design, Terms & Definitions

Traditionally, subjective image and video quality tests take a set of pristine stimuli and create multiple impaired versions of each. The typical experiment design contains one full-factorial matrix of source images (SRC) and impairments (HRC). The (SRC × HRC) design cannot be used here, since the camera itself is the source of impairment. The challenge is that we want to retain the ability to compare cameras based on a single source stimulus.

To do this, we need a new design and consequently new terms. Let us define a *scene* to be one subject matter with similar lighting, framing, distance, and viewing angle. For example, one scene might be your dinner at a restaurant, as seen from your chair. Let us define *equivalent images* as pictures that depict the same scene. An average person is expected to be able to obtain equivalent images by picking up two or more cameras in succession, without the use of a tripod or measuring tool. All pictures associated with a single scene will be referred to as an *equivalent image set*.

We can now replace the (SRC × HRC) matrix with a full-factorial (scene × HRC) matrix, where each HRC is a different camera. This test design reflects the real world situation where a consumer compares several different cameras by taking pictures of the same scene (e.g., a selfie). The (scene × HRC) design allows direct comparisons between cameras.

Our analyses will be confounded by uncontrolled variables surrounding the taking of the pictures. There will be inherent differences within an equivalent image set due to camera differences (e.g., aspect ratio, focal length) and slight changes in the scene (e.g., moving clouds, person in a portrait shifts position, fluctuations in natural light). We consider these uncontrolled variables to be part of the overall influence of the camera or the scene. Work underway within the Video Quality Experts Group (VQEG) is expected to provide some understanding of the impact of this design choice.

B. Cameras

CCRIQ includes 23 cameras (see Table 1). Many factors impact a camera's image quality, not just the image's pixel count and the display's resolution. Therefore, a wide variety of cameras were chosen, based on two characteristics. The first is pixel count in megapixels (MP), which was quantized into five bins: 1, 5, 8, 11 and 18 MP. These bins reflect the resolutions of popular consumer devices available on the market. The second is camera technology: mobile phone cameras, tablet cameras, compact cameras, and DSLRs. Camera availability depends upon the combination of pixel count and technology. For example, 1 MP cameras are an obsolete technology. Mobile phone cameras dominate our experiment, as this is the most popular camera type today. Note that compact camera "C" was eliminated due to an intermittent focusing problem.

Vendor information is intentionally omitted. This study does not provide a fair product comparison and should not be used for that purpose.

C. Scenes

Intel and NTIA took photos of 69 scenes. The scenes were limited to five topic categories: flat surfaces, landmarks at night, landscapes with good lighting, portraits, and still lifes. Cameras were operated in fully automatic mode. For each scene, one of two techniques was used to reduce the impact of random variables like handshake. Either multiple pictures were taken with each camera and one image selected, or an experienced photographer with steady hands took a single picture.

From the 69 scenes, 18 were selected for inclusion into CCRIQ. Figs. 1 and 2 display one image from each scene. Fig 3

TABLE 1. CAMERA RESOLUTION & TYPE

Camera Type	**1 MP**	**5 MP**	**8 MP**	**11 MP**	**18 MP**
Mobile phone	A	G H I N	L M	Q R S	V
Tablet	D	J			
Compact camera	B	E	O	P T	W
DSLR		F	K	U	X
MP Range	0.9–1.3	4.0–5.3	8.0	9.4–13	16–20

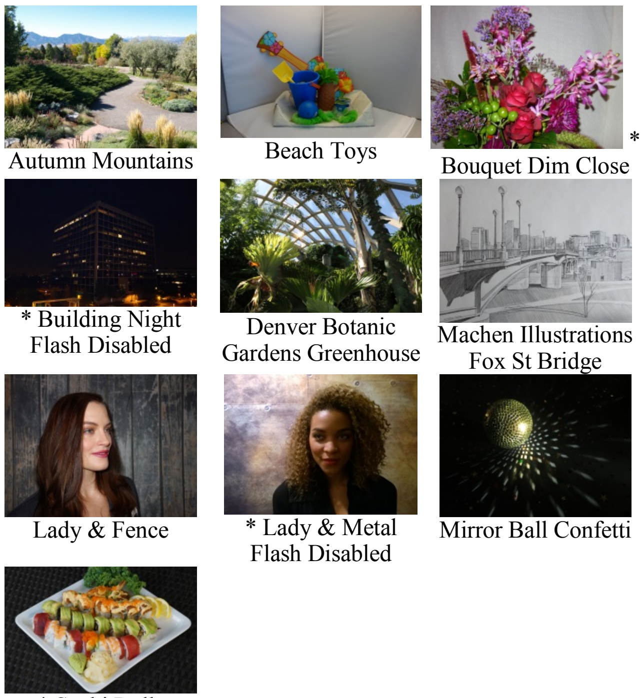

Fig. 1. Representative image of each scene in the red pool. Sets with less than 23 images are marked with an asterisk (*).

Fig. 3. Representative image of each scene in the blue pool. Sets with less than 23 images are marked with an asterisk (*).

shows a few equivalent images, to show the large differences between cameras. At least three scenes were chosen from each topic category, to showcase different camera responses within that category. For example, "Evacuation Plan" shows a large flat surface at a medium distance in good artificial light, while "Machen Illustrations Fox St Bridge" shows a small flat surface at a short distance with dim natural light.

The CCRIQ scenes in Figs 1 and 2 contain a variety of lighting conditions. Illumination ranged from 25,000 lux ("Winter Peaks") to less than 20 lux. "Mirror Ball Confetti" depicts a blue table lit by a yellow spot light. "Denver Botanic Gardens Greenhouse" points the lens near the midday sun.

By default, all compositions used auto-flash, and so most dim lighting compositions contain a mixture of flash and no-flash pictures. "Flash Disabled" indicates that the flash was manually turned off. "Lady & Fence Flash Disabled" and "Lady & Metal Flash Disabled" are partial sets, containing only cameras where auto-flash was triggered for "Lady & Fence" and "Lady & Metal" respectively. Other than this, each equivalent image set was supposed to contain photos from all 23 cameras. Some scenes depict short-lived events, so user error occasionally resulted in one or two images being missing.

D. *Subjective Method and Preliminary Test*

CCRIQ uses ITU-T Rec. P.913 [12] and the absolute category rating (ACR) scale. The scope of P.913 does not include image quality, however we believe this is an appropriate extension. Generally speaking, image quality is simply a special case of still video.

Subjective testing was performed by three laboratories: Intel (Santa Clara, CA, USA), NTIA/ITS (Boulder, CO, USA), and the Ghent University–iMinds (Ghent, Belgium). Each environment was a quiet room devoted to the task, with a mix of natural and artificial light (see Fig. 4). The illumination and viewing distance conditions were in compliance with ITU-T Rec. P.913. Also in compliance with Rec. P.913, visual acuity and color deficiency tests were performed but the subjects were not screened on this basis.

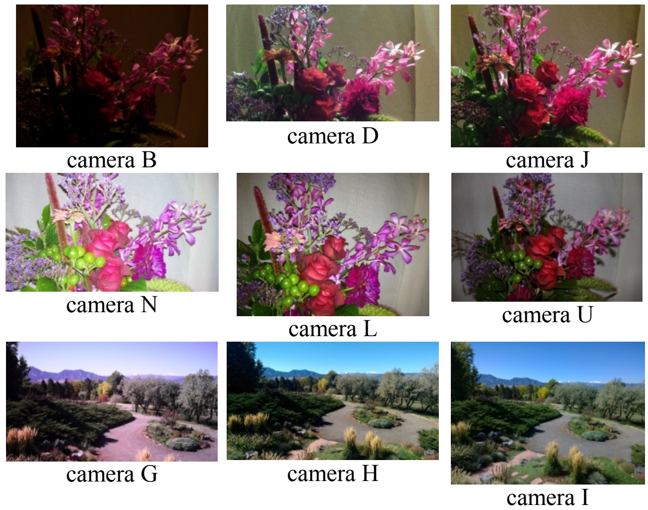

Fig. 2. These sample images show the large variety of camera responses within an equivalent image set. Cameras G, H and I are all 5 MP phone cameras (bottom row).

The subjective test ran automatically, using specialized software developed by Intel for the purpose of the study. Each test computer was configured with two 4K monitors (Samsung U28D590, a 28" 4K display).[3] One was configured for HD resolution (1920×1080) and the other configured for 4K resolution (3840×2160). Images were presented randomly to one or the other monitor throughout the test. Images were left in their original file (as saved by the camera) and scaled to the monitor using the High Quality Fant bitmap scaling algorithm implemented by the Image Class in the Microsoft .Net framework.

A preliminary test was run at a single lab (Intel) using seven subjects who rated all images associated with 15 scenes. This resulted in a set of 367 images, which were rated at both monitor resolutions for a total of $7 \times 367 \times 2$ ratings. This data indicated an undesirable lack of scenes with a high average mean opinion score (MOS). This triggered the collection of two new scenes with good lighting and aesthetics.

The preliminary test was also found to last too long for subject comfort. Therefore, the stimuli and subjects were divided into two pools: the red (Fig. 1) and the blue pool (Fig. 3). Each pool contains approximately half of the scenes. Two scenes are in both pools, to provide rating stability (Mirror

[3] Certain commercial equipment, materials, and/or programs are mentioned in this report to specify adequately the experimental procedure. In no case does such identification imply recommendation or endorsement by the authors, nor does it imply that the program or equipment identified is necessarily the best available for this application.

Ball Confetti and Sushi Rolls). Each subject rated all images in one pool on both monitors (HD and 4K). Each lab provided 8 or 9 subjects for each pool. Thus, each image was rated by 26 or 27 subjects for each monitor resolution and the overlapping images were rated by 53 subjects for each monitor resolution.

TABLE 2. IMPACT OF CAMERA FACTORS

Factor	Pearson Correlation	R^2	Spearman Correlation
All Cameras (2000–2006 & 2011–2014)			
Camera Type	0.365	0.133	0.404
Pixel Count	0.730	0.533	0.682
Sensor Size	0.385	0.148	0.413
Old Cameras (2000–2006)			
Camera Type	0.606	0.367	0.463
Pixel Count	0.988	0.976	0.928
Sensor Size	0.305	0.093	0.205
Modern Cameras (2011–2014)			
Camera Type	0.554	0.306	0.626
Pixel Count	0.670	0.448	0.585
Sensor Size	0.644	0.415	0.676

Fig. 4. Test environment for Ghent University (left), ITS (middle) and Intel (right). All environments had natural light.

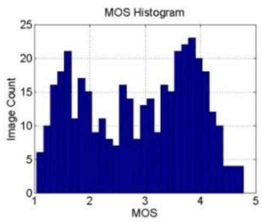

Fig. 5. Overall MOS score histogram for all images in the dataset.

TABLE 3. PEARSON CORRELATION BETWEEN LAB MOS SCORES

	Ghent	ITS	Intel
Ghent	1	0.952	0.915
ITS		1	0.941
Intel			1

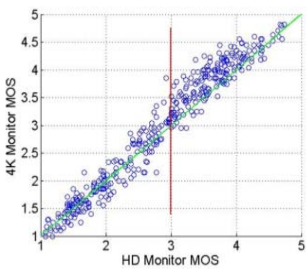

Fig. 6. Scatter plot of MOS scores from the 4K monitor versus MOS scores from the HD monitor.

Justification for this split and for combining the data directly (without rescaling) can be found in [13]. No subjects were discarded in post analyses.

IV. ANALYSIS

Fig. 5 shows the overall MOS distribution over the entire data set as a histogram of the scores. Notice that it spans the MOS range from 1 to 4.78. Table 3 shows the lab to lab MOS comparison. The Pearson correlation is computed between the MOSs of the three labs. The high correlations show that the MOS scores correlate well among the three labs.

The data show high linear correlation between MOS scores obtained from the HD monitor and the MOS scores from the 4K monitor. The Pearson correlation, R^2, and Spearman rank order correlation statistics between HD MOS and 4K MOS are 0.979, 0.958, and 0.976 respectively. The scatter plot in Fig. 6 illustrates this highly linear and monotonic trend.

A two-sided Student's *t*-test at the 95% significance level rejects the null-hypothesis that the HD MOSs and the 4K MOSs are statistically equivalent. Let us then partition the data points into MOSs lower than three and those greater than or equal to three, as obtained from the HD monitor (i.e., along the vertical bar in Fig. 6.). Now, a two-sided *t*-test concludes that there is a statistically significant difference between HD and 4K MOSs in the higher quality range (with the 4K MOSs being on average 0.2 MOS points higher than the HD MOSs), but no statistical difference in the lower quality range. Indeed a closer look at the scatter plot in Fig. 6 shows that the data points in the higher quality range mostly fall above the 45 degree line; whereas in the lower quality range they are approximately symmetrically distributed around it.

Figs. 7 and 8 show rating trends via boxplots of image MOSs (i.e., averaging over all subjects, labs, and monitors). The blue box spans the 25th to 75th percentile, the white dot in the center marks the mean MOS, the red vertical line marks the median, and all MOS values are rounded to one decimal place. Fig. 7 shows a boxplot for each scene; note the strong scene dependent quality response. Fig. 8 shows camera and scene factors. Phones and tablets are combined into a single category "mobile," due to the small number of tablets.

Several factors impact the quality of an image produced by a camera. Table 3 shows statistics comparing per camera MOSs (i.e., averaged over all subjects, labs, monitors, and scenes) with specific factors. The factor analysis of old cameras (2000-2006) is suspect for two reasons. First, there are only six old cameras in the study, three of which are obsolete 1 MP cameras with bad quality (see Fig. 8). Second, camera technology has changed significantly from 2000–2006 to 2010–2014, particularly in the post processing on the chip. Additionally, the pixel count and sensor size of modern cameras are not independent variables (R^2 = 0.35).

V. CONCLUSION

Focusing primarily on modern cameras, we reach the following conclusions:

- Pixel count impacts ≈30% to 45% of camera quality.
- Sensor size impacts ≈27% to 42% of camera quality.
- The overall quality difference between DSLR cameras and mobile cameras is 0.67 MOS.

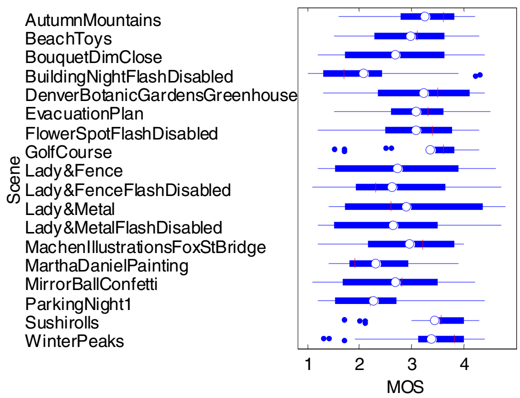

Fig. 7. Boxplot shows range of quality spanned by each scene.

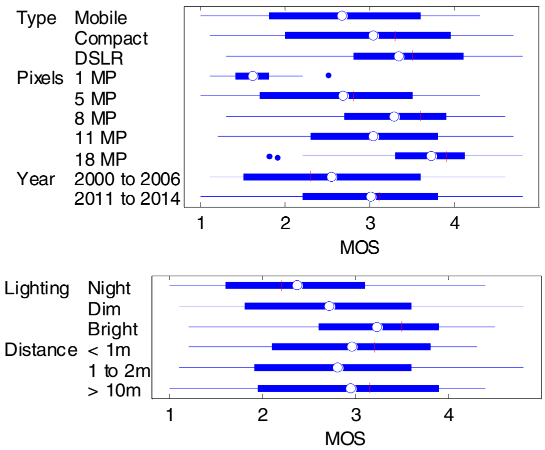

Fig. 8. Boxplots show rating trends by camera (top) and scene (bottom).

- The overall quality difference between a 18 MP camera and a 5 MP camera is 1.0 MOS.

These trends may be influenced by confounding factors. CCRIQ contains balanced experiment design (see Table 1) with the intent of reducing the influence of secondary factors.

The entire CCRIQ dataset is available for research and development purposes on the Consumer Digital Video Library (CDVL, www.cdvl.org). These image files retain the camera's format, compression, and metadata (i.e., except for vendor information). This is a valuable resource for the community. The CCRIQ dataset proposes a method for subjective testing of cameras with different resolutions, and statistics that relate pixel count and resolution to image quality. This information is expected to apply to video quality subjective testing.

Acknowledgment

Margaret Pinson would like to thank the following companies for kindly allowing photographs of their artwork to be distributed: Sturtz & Copeland, the Denver Botanic Gardens, and Machen Illustrations.

"Martha Daniel Painting" depicts an oil painting portrait of Martha Daniel by Aurelius O. Revenaugh in an ornate, gilt frame, circa 1900. Mr. Revenaugh was a self-taught, American painter (1840-1908). Martha Daniel was Margaret Pinson's great grandmother.

References

[1] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "LIVE image quality assessment database release," 2006, http://live.ece.utexas.edu/ research/quality/subjective.htm.

[2] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, F. Battisti, "TID2008 - A Database for Evaluation of Full-Reference Visual Quality Assessment Metrics," Advances of Modern Radioelectronics, Vol. 10, pp. 30-45, 2009.

[3] E. C. Larson and D. M. Chandler, "Consumer subjective image quality database," 2009, http://vision.okstate.edu/index.php?loc=csiq.

[4] P. Le Callet and F. Autrusseau, "Subjective quality assessment IRCCyN/IVC database," 2005, http://www.irccyn.ec-nantes.fr/ivcdb/.

[5] Z. M. Parvez Sazzad, Y. Kawayoke, and Y. Horita, "MICT image quality evaluation database," 2008, http://mict.eng.u-toyama.ac.jp/mictdb.html.

[6] S. Winkler: "Analysis of public image and video databases for quality assessment." IEEE Journal on Selected Topics in Signal Processing, vol. 6, no. 6, Oct. 2012.

[7] Virtanen, T., Nuutinen, M., Vaahteranoksa, M., Oittinen, P. and Häkkinen, J. "CID2013: a database for evaluating no-reference image quality assessment algorithms," IEEE Transactions on Image Processing, vol. 24, no. 1, pp. 390-402, Jan. 2015.

[8] M.A. Saad, P. Corriveau, and R. Jaladi, "Objective consumer device photo quality evaluation," IEEE Signal Processing Letters, March 2015. (to appear).

[9] M.A. Saad, P. Corriveau, and R. Jaladi, "Consumer content framework for blind photo quality evaluation," In Proc. 9th International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM), February, 2015

[10] M.A. Saad, P. Corriveau, and P. Le Callet, "Blind image quality assessment: unanswered questions and future directions in the light of consumers needs," VQEG eLetter, no. 2, vol. 1, pp. 62-66. Available: http://www.vqeg.org/

[11] M. Pinson, M.A. Saad, D.G. Nicholas, G. Van Wallendae, R.V. Jaladi, P.J. Corriveau, "Consumer Content Resolution and Image Quality Dataset (CCRIQ): An image quality subjective test of pixel count and display resolution," NTIA-TM, forthcoming.

[12] M. Pinson and L. Janowski, "A new subjective audiovisual & video quality testing recommendation," VQEG eLetter, no. 2, vol. 1, 2014. Available: http://www.vqeg.org/

[13] M. Pinson and A. Webster, "T1A1 validation test database," VQEG eLetter, no. 2, vol. 1, 2014. Available: http://www.vqeg.org/